The Use of Stochastic Methods to Explore the Thermal Equilibrium Distribution and Define Entropy Production out of Equilibrium

A Dissertation Submitted in Partial Fulfilment of the Requirements for the Degree of Doctor of Philosophy

Department of Physics and Astronomy
Faculty of Mathematical and Physical Sciences
UCL

Author:
Richard Edward Spinney

Supervisors:
Prof. Ian Ford
Prof. Mike Gillan
Dr. Dave Bowler

August 2012

Declaration

I, Richard Spinney, confirm that the work presented in this thesis is my own. Where information has been derived from other sources, I confirm that this has been indicated in the thesis.

Abstract

This thesis contains two separate bodies of research, both in terms of the period of time in which the work was done and their content, and as such is presented in two parts each of which are summarised below.

The first part concerns work on entropy production in stochastic systems and describes the breakage of time reversal symmetry that arises in irreversible stochastic processes that one can associate with an entropy production contribution for a single realisation. The paradigm utilised is that of Markovian dynamics expressed using master equations and stochastic differential equations. By generalising some previously reported concepts so as to explicitly concern odd variables, some recent advances in non-equilibrium thermodynamics are refined which are then illustrated with several examples. The place of such results within the existing literature, particularly the extensive literature on fluctuation theorems, is emphasised allowing us to simultaneously demonstrate some of the widely celebrated symmetry relations to emerge from the field in recent years.

The second part concerns the construction and implementation of a new Markov chain sampling algorithm called spatially local parallel tempering which improves the scaling of computational effort with system size of the well known thermal equilibrium sampling algorithm, parallel tempering. Parallel tempering accelerates thermal equilibrium sampling by performing regular sampling techniques on a composite system of replicas, each possessing a different temperature, and introducing configurational exchanges between those replicas so as to acquire configurations that would otherwise take a long time to reach. However, as the system size increases, the number of replicas required, and therefore computational effort, increases faster than linearly. To avoid this we propose local variations where this is not the case. We demonstrate these claims on several simple one dimensional models and show that the algorithms can reproduce thermodynamic accuracy in one and two dimensions.

Acknowledgements

As with all things in life, no work is performed in isolation, either professionally or personally, and so I take this opportunity to acknowledge those who have contributed, in their own ways, towards this completed manuscript.

First and foremost I must extend my deep thanks to my supervisor, Prof. Ian Ford, for his patience, encouragement and support which has made this thesis possible. In particular, I am immensely grateful for a genuinely collaborative working relationship which, without question, I deem responsible for a welcoming, productive and rewarding environment within which to work. In addition I thank Prof. Mike Gillan and Dr. Dave Bowler for their guidance and considerable understanding, particularly in my early postgraduate years. It would also be remiss of me not to mention the efforts of some of my fellow PhD students who have been there to provide anything from useful discussions, an ear to bend, through to well needed distraction and everything in between.

Finally I thank my family for all their support, financial and otherwise, throughout my entire academic journey from childhood inspiration to the challenges of a 'writing up' PhD student. None of this would have happened without you. This thesis is dedicated to my mother.

Publications

- R. E. Spinney, D. R. Bowler and M. J. Gillan "Spatially local parallel tempering for thermal-equilibrium sampling", J. Chem. Phys. **132**, 104106 (2010)
- Richard E. Spinney and Ian J. Ford "Nonequilibrium thermodynamics of stochastic systems with odd and even variables", Phys. Rev. Lett. **108**, 170603 (2012)
- Richard E. Spinney and Ian J. Ford "Entropy production in full phase space for continuous stochastic dynamics", Phys. Rev. E **85**, 051113 (2012)
- Ian J. Ford and Richard E. Spinney "Entropy production from stochastic dynamics in discrete full phase space", Phys. Rev. E **86**, 021127 (2012)
- Richard E. Spinney and Ian J. Ford "Fluctuation Relations: A pedagogical overview". Chapter contributed to: R.Klages, W.Just, C.Jarzynski (Eds.), Nonequilibrium Statistical Physics of Small Systems: Fluctuation Relations and Beyond (Wiley-VCH, Weinheim, 2012; ISBN 978-3-527-41094-1)

Contents

List of Figures

List of Tables

Part I

Entropy Production in Stochastic Systems

Foreword regarding the structure of work on entropy production in stochastic systems

The following work describes and illustrates a novel division of the entropy production for stochastic systems with a particular emphasis on such a formalism's place within the literature on fluctuation theorems. After an introduction, the requisite theory on stochastic processes, particularly that which we find crucial in order to define central quantities in later chapters such as the definition of path probability functionals and the subtleties surrounding stochastic integration, shall be addressed in chapter 2. In chapter 3 a literature review is given by means of general derivation of some key identities along with applications to important results in the literature and leads into a motivation for the main body of the thesis. In chapters 4 and 5 the main results are presented and specifically concern the novel division of entropy production; chapter 4 focuses on discrete stochastic processes whilst chapter 5 focuses on continuous processes. Most properties are shared between such approaches, however some points of discussion are delayed until chapter 5, where proceeding is slightly more intuitive and allows a closer alignment with the literature. Several illustrative examples are then presented in chapter 6 including under-damped Langevin particles under the influence of a non-conservative force, where we consider various models for the nature of the damping and in the relativistic limit, and in a spatially varying temperature field along with some discrete state space models. A brief discussion is then given in chapter 7.

Chapter 1

Introduction

Long after the classical works on equilibrium statistical mechanics, much of the nature of matter out of equilibrium remains elusive. Whilst classical thermodynamics qualitatively states that an irreversible process must increase the entropy of the universe, the result of course being the second law of thermodynamics, surprisingly little else can be said about such a procedure despite non-equilibrium phenomena being ubiquitous throughout the physical world. Of course the relative difficulty in describing such processes could quite easily be interpreted as a monument to the success that equilibrium statistical mechanics has enjoyed in describing systems of such a large magnitude. The classical theory's predictive power is made possible by its ability to reduce extraordinary numbers of degrees of freedom into several simple functions of state; we can describe the behaviour of a gas without concerning ourselves with the motion of its constituent atoms. When dealing with a system driven out of equilibrium however, no such simplification exists and so we resign ourselves to a consideration of how the system evolves rendering a description far more challenging [1].

The gap in our knowledge between our understanding of the underlying microscopic processes that allow us to describe physical phenomena and the macroscopic concepts of irreversibility and entropy (and indeed of time) is well known and can be summarised by the reversibility paradox usually attributed to Loschmidt. In response to Boltzmann's H-Theorem it was postulated that no measure of irreversibility can be obtained when the underlying dynamics are reversible since for every entropy producing path there must be an equally valid entropy destroying path. This is of course at odds not only with classical theory, but with our experience of the world. For example, one would be equally alarmed to observe a cloud of smoke spontaneously reform from dispersed particles in the air as if one were to observe a clock running backwards. Within the last twenty years however, several disarmingly simple relations have been derived that concern the properties of recognisable thermodynamic quantities which we can associate with a measure of irreversibility by explicitly exploiting the time reversibility of the underlying equations of motion [2–12]. The origins of these relations can be traced back to a result due to Evans et al. who described, originally for sheared two dimensional fluids in the steady state in the long time limit, a symmetry relation explicitly detailing the probability of an increase in a generalised entropy in relation to the probability of the same amount of entropy being destroyed [2] of the form

$$p(-A) = p(A) \exp(-A). \tag{1.1}$$

This description was the first quantitative account of how irreversibility can arise from reversible dynamics and has been claimed to solve Loschmidt's paradox [13], though perhaps most would still, strictly,

remain unconvinced. Furthermore it confirmed, in a quantitative sense, that for small enough systems one should expect to observe an entropy decrease with a non-vanishing probability; for small enough systems we should expect, occasionally, for them to run in reverse. Of equal note is that, compared to most results in modern statistical mechanics, no assumption of proximity to equilibrium is required in its derivation as it relies only on the underlying dynamics. The existence of the fluctuations described in such a relation, of which the negative ones are sometimes called 'second law violations' [14], require us not to think of thermodynamic quantities such as work, heat and entropy as sharp singular values as is so familiar in the usual thermodynamics, but rather consider them to be described by distributions where the exact values they take must depend on how the system has evolved as a result of some non-equilibrium influence. This notion, whilst conceptually straightforward for mechanical quantities such as the work done, is less straightforward when considering the usually ensemble quantity of entropy, but has now become commonplace, encouraging discussion of thermodynamics explicitly in terms of fluctuations. Such treatments have paved the way for a vast body of work, from which the most notable results are in the spirit of the symmetry in observed second law violations and include asymptotic and finite time relations for generalised entropy productions, particle currents and some well known non-equilibrium work relations [8, 10]. All however, are underpinned by the exploitation of the same property of the underlying dynamics, namely that of micro-reversibility, which in such a context is interpreted as the fact that both the underlying dynamics and their time reversed counterpart permit the same solutions. The similarity that exists between the members of this family of relations has led to some broad generalisations and unifying descriptions which has allowed, as something of an umbrella term, such results to be collectively known as fluctuation theorems or fluctuation relations.

As arguably one of the only real developments in thermodynamics for some 50 years it would be fair to say there has been something of an explosion of interest in such descriptions. Indeed, a quick online search of the phrase 'fluctuation theorem' will return a number of articles running into the thousands and certainly one could not reasonably hope to account for every development. However, we make note of many excellent review articles which give well rounded accounts from particular perspectives, be that theoretical, with a view towards stochastic or deterministic behaviour, or experimental [6, 13, 15–25]. Broadly speaking however, the progress in the field has branched into several illuminating parallel descriptions dependent on the particular model utilised to describe the underlying dynamics with fluctuation theorems having been developed for Hamiltonian dynamics [3, 4, 26], several classes of stochastic dynamics [25, 27, 28] and some excursions into quantum descriptions as well [29, 30] along with relevant examples and applications. Perhaps an avenue of enquiry which has been most productive in terms of novel results is the application to stochastic systems where theoretical treatment is more tractable and where such a description is most readily applicable to the length and timescales typically available in experiment. As alluded to, treatment in such models can be discussed because the property that led to the original fluctuation theorems, that of micro-reversibility, can be defined in a quite general sense and so allows progress in the context of models without explicitly time reversible equations of motion. As such the goals of such stochastic approaches are not to identify irreversibility per se, but rather to describe its properties in, potentially experimental, systems where practically one is forced to view the dynamics as uncertain and dissipative. Historically, the path taken to arrive at the descriptions we shall utilise is somewhat complicated. Following the original results of Evans et al., an asymptotic relation concerning the phase space contraction, identifiable as a dynamical entropy production, but differing from the argument of the original fluctuation theorem by boundary terms for chaotic, deterministically thermostatted systems was derived by Gallavotti and Cohen [31, 32]. This asymptotic description was then applied to commonly used Markovian models by Kurchan [27] and then for a more general glass of

Markovian systems by Lebowitz and Spohn [28]. More recently, an argument for the inclusion of such boundary terms, whose absence resulted in the asymptotic results, was offered [12]. Such an approach has then allowed a general thermodynamic description of such stochastic systems which has come to be known as stochastic thermodynamics [12, 33]. Drawing strongly on the concepts used in defining fluctuation theorems for stochastic dynamics, and by utilising ideas from a body of work first performed by Sekimoto [34] which assigned dissipative and work quantities to terms in a Langevin equation, a central concept in such a unifying relationship is that of an individual, path dependent, fluctuating value of entropy production. More recently still, by making connections with an axiomatically proposed division of heat dissipation [35], such that one can meaningfully discuss the thermodynamics of non-equilibrium steady states and the transitions between them, several distinct results have been brought together into a further unifying thermodynamics, investigating the precise origins of entropy production, couched firmly in the language of fluctuation theorems [36].

The following work amounts to a continuation of these ideas for stochastic systems by examining the stochastic entropy production a little more closely and exploring some of its consequences. After a brief review of the stochastic processes we shall utilise, we present an overview of some of the key results that emerge from stochastic thermodynamics along with relevant developments in the literature. Such an overview is, as much as is possible, designed to make the work self contained rather than to be an exhaustive review, with the latter goal simply being impractical given the volume of inter-related work. Reflecting upon some of the recent developments we revisit the connection between entropy and time reversal and then seek to generalise the quantities and results which have used a more narrow definition and so introduce a new formalism for the division of entropy production in stochastic systems. We then consider some example systems in our proposed operational thermodynamics which we find necessary to fully describe certain models of heat conduction and distinguish between variants of Brownian motion, such as dynamics that include dry friction.

Chapter 2

Theory on Stochastic Processes

2.1 Stochastic Processes and Brownian Dynamics

2.1.1 From Determinism to Chaos & Stochasticity

In dealing with non-equilibrium phenomena no longer can we consider a system in terms of its functions of state. Instead we are obliged to consider the microscopic dynamics of our system. At or near room temperature (which includes the circumstances of a vast array of processes one may wish to consider) these dynamics can quite readily be supposed to be classical such that we consider them to be governed by Newton's equations. Such a system, for example, could be the archetypal Brownian motion of a single pollen grain suspended on the surface of water. The behaviour of the pollen grain is theoretically uniquely determined by constructing Hamilton's equations for the water molecules and of the grain. In reality however, understanding the dynamics in this way would be hopeless. Not only are the mathematics intractable, but also non-linear and thus chaotic and as such incredibly sensitive to the complete initial conditions of the pollen grain and the water molecules. In practice the construction or even specification of these initial conditions would be impossible to achieve. In order, then, to make any progress at all, we cease to think of the system as a set of equations between all the water molecules and the pollen grain, but as the pollen grain under the influence of an environment of which we are uncertain such that we do not concern ourselves with its precise dynamics. Couched in this way, if we measured the trajectory of the pollen grain at a specified time after releasing it from a known initial position we would record a different path with each repeated measurement. We would observe an apparent randomness in its behaviour. Accordingly we abandon any sense of a deterministic solution, but instead consider the evolution in terms of probabilities and expected values.

2.1.2 The Chapman-Kolmogorov Equation

Specifically we now consider properties of any such particle (for example position and/or velocity) to be stochastic variables. The evolution of these variables is now considered to be a stochastic process. Such a process can be very generally defined (for any $t_2 > t_1 > t_0$) by the use of joint probabilities of any path $(x_0 \to x_1 \to x_2)$ as the product of conditional probabilities used as transition probabilities and simple probability distributions as initial conditions, where we note that in the approach used here strictly all are probability *densities*, such that

$$p(x_2, t_2; x_1, t_1; x_0, t_0) = p(x_2, t_2 | x_1, t_1; x_0, t_0) p(x_1, t_1 | x_0, t_0) p(x_0, t_0) \tag{2.1}$$

which describes the joint probability density of observing a state x_0 at time t_0, x_1 at t_1 and x_2 at t_2. One can then find the marginal probability distribution by integrating the left hand side with respect to x_1 giving

$$\int dx_1\, p(x_2, t_2; x_1, t_1; x_0, t_0) = p(x_2, t_2; x_0, t_0). \tag{2.2}$$

Similarly this marginal distribution can then be expressed as a joint distribution using a conditional probability and a one time distribution such that

$$p(x_2, t_2; x_0, t_0) = p(x_2, t_2|x_0, t_0)p(x_0, t_0). \tag{2.3}$$

Considering now the right hand side we integrate to find

$$\int dx_1\, p(x_2, t_2|x_1, t_1; x_0, t_0)p(x_1, t_1|x_0, t_0)p(x_0, t_0) = \\ p(x_0, t_0) \int dx_1\, p(x_2, t_2|x_1, t_1; x_0, t_0)p(x_1, t_1|x_0, t_0). \tag{2.4}$$

Comparing the integral of both the left and right hand sides leads to the Chapman-Kolmogorov equation which is generally given by

$$p(x_2, t_2|x_0, t_0) = \int dx_1\, p(x_2, t_2|x_1, t_1; x_0, t_0)p(x_1, t_1|x_0, t_0). \tag{2.5}$$

If we now allow ourselves to consider the stochastic process as independent of the stochastic variable's history we can consider the process to be Markov. For a Brownian particle this would be equivalent to considering the medium so complex that the force incident on the particle is not correlated on any time scale allowing us to think of the stochastic process as memory-less. To describe a general Markov process we start with the Markov property

$$p(x_n, t_n|x_{n-1}, t_{n-1}; \ldots x_1, t_1; x_0, t_0) = p(x_n, t_n|x_{n-1}, t_{n-1}). \tag{2.6}$$

Noting its importance in relation to the Chapman-Kolmogorov equation we identify

$$p(x_2, t_2|x_1, t_1; x_0, t_0) = p(x_2, t_2|x_1, t_1). \tag{2.7}$$

Accordingly the Chapman-Kolmogorov equation for a Markov process reduces to

$$p(x_2, t_2|x_0, t_0) = \int dx_1\, p(x_2, t_2|x_1, t_1)p(x_1, t_1|x_0, t_0) \tag{2.8}$$

which is a fundamental identity of a Markov process along with

$$p(x_1, t_1) = \int dx_0\, p(x_1, t_1|x_0, t_0)p(x_0, t_0) \tag{2.9}$$

which is found by integrating Eq. (2.3) again to find the marginal probability and relabelling.

2.1.3 The Master Equation

Taking the Chapman-Kolmogorov equation (Eq. (2.8)) we Taylor expand $p(x_2, t_2|x_1, t_1)$ for small $dt = t_2 - t_1$ and so approximate

$$p(x_2, t_1 + dt|x_1, t_1) = \delta(x_1 - x_2)(1 - a(x_1)dt) + \mathcal{T}(x_2|x_1, t_1)dt + O(dt^2) \tag{2.10}$$

where we define $\mathcal{T}(x_2|x_1, t_1)$ to be the transition probability density per unit time, or simply the transition rate density, at time t_1 and

$$a(x_1) = \int dx_2 \mathcal{T}(x_2|x_1, t_1) \tag{2.11}$$

ensuring normalisation and which corresponds to total mean escape rate from state x_1. Inserting this into Eq. (2.8) yields

$$p(x_2, t_1+dt|x_0, t_0) = p(x_2, t_1|x_0, t_0)+dt \int dx_1 \mathcal{T}(x_2|x_1, t_1)p(x_1, t_1|x_0, t_0)-dta(x_2)p(x_2, t_1|x_0, t_0). \tag{2.12}$$

Since we may write

$$a(x_2) = \int dx_1 \mathcal{T}(x_1|x_2, t_1) \tag{2.13}$$

we have, in the limit $dt \to 0$,

$$\frac{\partial p(x_2, t_1|x_0, t_0)}{\partial t} = \int dx_1 \left[\mathcal{T}(x_2|x_1, t_1)p(x_1, t_1|x_0, t_0) - \mathcal{T}(x_1|x_2, t_1)p(x_2, t_1|x_0, t_0)\right]. \tag{2.14}$$

Given an initial distribution we may employ Eq. (2.9), relabel $t_1 \to t$, $x_2 \to x$ and $x_1 \to x'$ and write the master equation

$$\frac{\partial p(x, t)}{\partial t} = \int dx' \left[\mathcal{T}(x|x', t)p(x', t) - \mathcal{T}(x'|x, t)p(x, t)\right] \tag{2.15}$$

with the two terms corresponding to transitions into and out of x respectively. Such a description allows a very general account of stochastic phenomena and allows, in principle, discussion of any Markovian process given that one can identify the transition rate densities $\mathcal{T}(x|x', t)$. More often, a master equation is used to model a system of discrete states, for which there is an equivalent form, simply rewritten using transition rates, T, as

$$\frac{dP(x, t)}{dt} = \sum_{x' \neq x} \left[T(x|x', t)P(x', t) - T(x'|x, t)P(x, t)\right]. \tag{2.16}$$

If we then utilise the notation

$$T(x|x', t) = \begin{cases} T(x|x', t) & x \neq x' \\ -\sum_{x'' \neq x} T(x''|x, t) & x = x', \end{cases} \tag{2.17}$$

thus defining a mean escape rate $T(x|x, t)$, we can write it in the equivalent matrix form

$$\frac{dP(x, t)}{dt} = \sum_{x'} T(x|x', t)P(x', t), \tag{2.18}$$

such that $\dot{P}(t) = \boldsymbol{T}(t)P(t)$ where $T_{x,x'} = T(x|x', t)$ which is the form of the master equation we shall use to model stochastic motion on sets of discrete states throughout.

2.1.4 Kramers-Moyal Expansion of the Chapman-Kolmogorov Equation

As an alternative treatment to the master equation which, in principle, describes the entire probabilistic behaviour, we may seek a differential form based on not only a small increment in time, but also in the stochastic variables. Taking the Chapman-Kolmogorov identity (Eq. (2.8)) we do so by taking the time interval $t_2 - t_1 = dt \to 0$ as before and proceed by taking this limit and integrating over an arbitrary function $f(x_2)$ which vanishes at $\pm\infty$ so that

$$\int dx_2\, f(x_2)p(x_2, t_1 + dt|x_0, t_0) = \int dx_2 \int dx_1\, f(x_2)p(x_2, t_1 + dt|x_1, t_1)p(x_1, t_1|x_0, t_0). \tag{2.19}$$

Since we are taking the limit $dt \to 0$ we can expand the left hand side up to first order in dt such that we have

$$\int dx_2\, f(x_2)p(x_2, t_1 + dt|x_0, t_0) \simeq \int dx_2\, f(x_2) \left[p(x_2, t_1|x_0, t_0) + \frac{\partial}{\partial t_1} p(x_2, t_1|x_0, t_0)dt\right]. \tag{2.20}$$

Now crucially, we demand that x is a suitably chosen stochastic variable such that when dt is small we can expect dx also to be small so that we can simultaneously expand the arbitrary function $f(x_2)$ around a nearby point x_1 so that

$$f(x_2) \simeq \sum_{n=0}^{\infty} \frac{1}{n!} \frac{d^n f(x_1)}{dx_1^n} (x_2 - x_1)^n. \tag{2.21}$$

At this point we define a set of coefficients $M_n(x, t)$ known as the Kramers-Moyal coefficients or jump moments as

$$M_n(x_1, t_1) = \lim_{dt\to 0} \frac{1}{dt} \int dx_2\, (x_2 - x_1)^n p(x_2, t_1 + dt|x_1, t_1) \tag{2.22}$$

with their physical meaning becoming more transparent when written as

$$M_n(x, t) = \lim_{dt\to 0} \frac{\langle [x(t + dt) - x(t)]^n \rangle}{dt}. \tag{2.23}$$

Using this definition we now express the right hand side as

$$\begin{aligned}
&\int dx_2 \int f(x_2)dx_1\, p(x_2, t_1 + dt|x_1, t_1)p(x_1, t_1|x_0, t_0) \\
&\qquad \simeq \int dx_1 \left[\sum_{n=0}^{\infty} \frac{M_n(x_1, t_1)}{n!} \frac{d^n f(x_1)}{dx_1^n} p(x_1, t_1|x_0, t_0)dt\right] \\
&\qquad \simeq \int dx_1 \left[f(x_1)p(x_1, t_1|x_0, t_0) + \sum_{n=1}^{\infty} \frac{M_n(x, t)}{n!} \frac{d^n f(x_1)}{dx_1^n} p(x_1, t_1|x_0, t_0)dt\right].
\end{aligned} \tag{2.24}$$

We can then equate this to the right hand side from Eq. (2.20) with the label x_2 changed for convenience to x_1 so that

$$\int dx_1 f(x_1) \frac{\partial}{\partial t_1} p(x_1, t_1|x_0, t_0)dt = \int dx_1 \sum_{n=1}^{\infty} \frac{M_n(x_1, t_1)p(x_1, t_1|x_0, t_0)}{n!} \frac{d^n f(x_1)}{dx_1^n} dt. \tag{2.25}$$

Evaluating the integral on the right hand side using integration by parts n times we obtain

$$\int dx_1 \sum_{n=1}^{\infty} \frac{M_n(x,t)p(x_1,t_1|x_0,t_0)}{n!} \frac{d^n f(x_1)}{dx_1^n} dt$$
$$= \sum_{n=1}^{\infty} \left[\sum_{i=0}^{n-1} (-1)^i \frac{\partial^i}{\partial x_1^i} \left(\frac{M_n(x_1,t_1)p(x_1,t_1|x_0,t_0)}{n!} \right) \frac{d^{n-1-i} f(x_1)}{dx_1^{n-1-i}} \right]_{-\infty}^{+\infty}$$
$$+ \sum_{n=1}^{\infty} \int dx_1 f(x_1) \frac{(-1)^n}{n!} \frac{\partial^n}{\partial x_1^n} M_n(x_1,t_1)p(x_1,t_1|x_0,t_0)dt. \tag{2.26}$$

Assuming compact form in both $f(x_1)$ and $p(x_1,t_1|x_0,t_0)$ such that their values and derivatives vanish at infinity we find

$$0 = \int dx_1 f(x_1) \left[\frac{\partial}{\partial t_1} - \sum_{n=1}^{\infty} \frac{(-1)^n}{n!} \frac{\partial^n}{\partial x_1^n} M_n(x_1,t_1) \right] p(x_1,t_1|x_0,t_0). \tag{2.27}$$

Since $f(x_1)$ is arbitrary we can then identify

$$\frac{\partial}{\partial t_1} p(x_1,t_1|x_0,t_0) = \sum_{n=1}^{\infty} \frac{(-1)^n}{n!} \frac{\partial^n}{\partial x_1^n} M_n(x_1,t_1)p(x_1,t_1|x_0,t_0) \tag{2.28}$$

which is an infinite order partial differential equation which describes the evolution of the conditional probability called the Kramers-Moyal expansion.

2.1.5 A Fokker-Planck Equation

Eq. (2.28) exactly models the stochastic process laid out above although, however it is an infinite order partial differential equation and so a question arises as to how many terms one should utilise in practice. A theorem by Pawula [37] showed that either the sequence terminates after the first or second term or not at all and that if any even terms vanish all $n > 2$ terms vanish. Consequently the first two terms are usually retained as either an exact representation of the process or as an approximation to a process whose Kramers-Moyal expansion does not terminate and is usually referred to as a Fokker-Planck equation or more generally a forward Kolmogorov equation. Additionally it can be shown that the Kramers-Moyal equation reduces to a Fokker-Planck equation when the random variable on which the process is based has a continuous sampling path [38]. Consequently the Fokker-Planck equation is written as

$$\frac{\partial p(x_1,t_1|x_0,t_0)}{\partial t_1} = -\frac{\partial}{\partial x_1} \left[A(x_1,t_1)p(x_1,t_1|x_0,t_0) \right] + \frac{\partial^2}{\partial x_1^2} \left[D(x_1,t_1)p(x_1,t_1|x_0,t_0) \right] \tag{2.29}$$

where $A(x,t)$ and $D(x,t)$ are the first and second Kramers-Moyal coefficients respectively. Although Eq. (2.29) describes conditional probabilities, the Fokker-Planck equation can equally describe one time probabilities $p(x,t)$ using the relation

$$p(x_1,t_1) = \int dx_0 p(x_1,t_1|x_0,t_0)p(x_0,t_0) \tag{2.30}$$

as long as an initial condition is satisfied

$$p(x,t)\big|_{t=t_0} = p(x,t_0). \tag{2.31}$$

Consequently we can write down the Fokker-Planck equation in a more usual form

$$\frac{\partial p(x,t)}{\partial t} = -\frac{\partial}{\partial x}\left[A(x,t)p(x,t)\right] + \frac{\partial^2}{\partial x^2}\left[D(x,t)p(x,t)\right]. \tag{2.32}$$

In this form, we see that the equation describes the time evolution of a probability density function as the sum of two contributions: a drift term dependent on $A(x,t)$ and a diffusive term dependent on $D(x,t)$. We can loosely consider the two terms as contributions from the deterministic and diffusive elements of the dynamics respectively since $D(x,t)$ controls a parabolic term which regularly arises in diffusive behaviour such as the heat equation and by recognising that in the case of $D(x,t) = 0$ the equation resembles the Liouville equation of Hamiltonian mechanics.

2.1.6 A Theory of Brownian Motion

It is observed that the motion of a Brownian particle whilst stochastic is continuous and, as a reasonable approximation, can readily be assumed to be Markovian. As such one should expect to be able to describe it with a Fokker-Planck equation. We simply need to determine the relevant coefficients

$$A(x,t) = \lim_{dt\to 0} \frac{\langle x(t+dt) - x(t)\rangle}{dt} \tag{2.33}$$

and

$$2D(x,t) = \lim_{dt\to 0} \frac{\left\langle [x(t+dt) - x(t)]^2 \right\rangle}{dt}. \tag{2.34}$$

We understand that the first coefficient corresponds to the particle's deterministic behaviour for which we turn to the particle's equation of motion. This is the motion we should observe when fluctuations are unimportant such as relatively massive particles and is given by

$$m\ddot{x} + m\gamma\dot{x} = \mathcal{F}(x), \tag{2.35}$$

where m is the mass of the particle, γ is the damping or friction coefficient due to the medium and $\mathcal{F}$ is the macroscopic force experienced by the particle. The Fokker-Planck equation however is a function of only one stochastic variable x, but not v and as such we take a limit in which we can ignore the velocity. This is the over-damped limit and corresponds to $\gamma \to \infty$ and allows us to consider that there is instantaneous acceleration acting upon the particle leaving us with

$$\dot{x} = \frac{\mathcal{F}(x)}{m\gamma}. \tag{2.36}$$

Considering the deterministic contribution or equivalently the motion of a large particle we can now associate the evolution of x with the mean of the distribution in Eq. (2.32). To obtain this we multiply by x and integrate over all space and assume a well behaved density function such that $p(x,t)$ and its gradient vanish at $\pm\infty$. Additionally, for simplicity, we consider a spatially uniform coefficient for $B(t)$ corresponding to a homogeneous medium giving

$$\int_{-\infty}^{+\infty} dx\, \frac{\partial}{\partial t} x p(x,t) = -\int_{-\infty}^{+\infty} dx\, x \frac{\partial}{\partial x}\left[A(x,t)p(x,t)\right] + \int_{-\infty}^{+\infty} dx\, xD(t)\frac{\partial^2}{\partial x^2} p(x,t). \tag{2.37}$$

Evaluating all integrals whilst taking the time derivative outside of the first then gives

$$\frac{d}{dt}\langle x\rangle = -\left[xA(x,t)p(x,t)\right]_{-\infty}^{+\infty} + \int_{-\infty}^{+\infty} dx A(x,t)p(x,t) + \left[xD(t)\frac{\partial}{\partial x}p(x,t)\right]_{-\infty}^{+\infty} - \int_{-\infty}^{+\infty} dx D(t)\frac{\partial}{\partial x}p(x,t). \quad (2.38)$$

Finally, by discarding surface terms we can write

$$\frac{d}{dt}\langle x\rangle = \langle A(x,t)\rangle \quad (2.39)$$

which allows us to identify $A(x,t)$ by comparison with Eq. (2.36) giving us

$$\frac{\partial p(x,t)}{\partial t} = -\frac{\partial}{\partial x}\left[\frac{\mathcal{F}(x)}{m\gamma}p(x,t)\right] + D(t)\frac{\partial^2}{\partial x^2}p(x,t). \quad (2.40)$$

Now imagining such a system confined within a potential which is not pathological, in as much as a stationary solution exists, we should expect the Fokker-Planck equation to be consistent with the usual Boltzmann formula from statistical mechanics. We would generally require that when stationary we have

$$0 = -\frac{\mathcal{F}(x)}{m\gamma}p(x,t) + D(t)\frac{\partial}{\partial x}p(x,t) \quad (2.41)$$

and therefore that

$$p(x,t) \propto \exp\left[\frac{\int^x dx'\mathcal{F}(x')}{k_B T}\right]. \quad (2.42)$$

If we substitute this expression for the density function into the result for the stationary state we obtain

$$D(t) = \frac{k_B T}{m\gamma} \quad (2.43)$$

allowing us to determine the expected diffusivity. The expression is often further prescribed by the use of Stokes' Law such that if the particle is spherical with radius r and the fluid has viscosity η we can describe the friction coefficient as $m\gamma = 6\pi r\eta$. This relation is often referred to as the Einstein relation or Stokes-Einstein relation when used in conjunction with Stokes' Law. Its application now allows us to write down the Fokker-Planck equation and thus describe the Brownian motion

$$\frac{\partial p(x,t)}{\partial t} = -\frac{\partial}{\partial x}\left[\frac{\mathcal{F}(x)}{m\gamma}p(x,t)\right] + \frac{k_B T}{m\gamma}\frac{\partial^2}{\partial x^2}p(x,t). \quad (2.44)$$

An important interpretation of the result in Eq. (2.43) is that it provides a connection between the size of the fluctuations *at equilibrium* due to the medium and the particles deterministic behaviour, namely the 'dissipative' damping that particles experience within it. As such this relation is an example of a so called *fluctuation-dissipation* relation and is a powerful tool allowing us to describe certain *non-equilibrium* behaviour with a simple equilibrium result. We are able to make this extension because the system cannot distinguish between perturbations away from equilibrium due to fluctuations that arise whilst in thermal equilibrium and perturbations arising externally meaning the system will relax back to equilibrium in the same way. Such results strictly only hold in the linear regime, but provide a good starting point for model behaviour.

2.1.7 The Langevin Equation

Whilst Eq. (2.44) gives a full description of the Brownian motion outlined so far it itself is a deterministic equation which describes the probability distribution of the particle's position. An alternative description first used by Langevin [39] considered a microscopic description of the particle's motion. The construction is phenomenological in nature and starts with the deterministic behaviour of the particle seen in Eq. (2.35), but then includes a force term ($\Gamma(t)$), controlled by a constant parameter b', which describes the effect of the medium upon the particle such that

$$m\ddot{x} + m\gamma\dot{x} = \mathcal{F}(x) + b'\Gamma(t). \tag{2.45}$$

The force $\Gamma(t)$ that would accurately describe the medium ultimately is unknown and so we must proceed largely with intuition. As the force term (usually called *noise*) is modelling the environment, which we do not intend to consider deterministically, it inherits a stochastic treatment and as such we describe it by specifying its statistical properties. The Fokker-Planck treatment revealed that for an homogeneous medium the mean of the distribution is governed by the particle's deterministic behaviour so we should expect no further contribution on average due to the fluctuations. Consequently we can say that the mean of the noise must be zero

$$\langle\Gamma(t)\rangle = 0. \tag{2.46}$$

We next further describe the noise term using its auto-correlation function which in turn gives a measure of its variance. At this point we make an approximation analogous to that of the Markov property used in Eq. (2.6). We consider the collisions from the surrounding molecules to be occurring so rapidly that the force is entirely uncorrelated with itself on any non-zero macroscopic timescale. This can be specified by allowing the auto-correlation to be a delta function

$$\langle\Gamma(t)\Gamma(t')\rangle = \delta(t - t'). \tag{2.47}$$

This specification is called *white noise* and is the idealisation of a fluctuating force that varies on an infinitely small timescale. The white noise property is of course impossible to realise in any real system and indeed the real dynamics of the environment must have some finite correlation, however the idealisation is in line with viewing the process as Markov and allows us to make contact with an equivalent Fokker-Planck equation (Eq. (2.44)) by suitably choosing the parameter b' and calculating the Kramers-Moyal coefficients of the resulting process [40].

These two statistical constraints despite being the most cited properties of the stochastic term are not enough to fully determine $\Gamma(t)$ and therefore $x(t)$ as there is no mention of higher moments. Consequently the form of the distribution of $\Gamma(t)$ is usually taken to be Gaussian. This has the advantage that the distribution is uniquely described by only the mean and variance allowing all the statistical properties of the noise to be determined by b'. Further justification is based on the postulation that the result of a very large number of independent collisions (or more generally interactions with the environment) will produce a normal distribution because of the central limit theorem.

To establish the so far unfixed parameter b' we once again take the over-damped limit meaning our Langevin equation is given by

$$\dot{x} = \frac{\mathcal{F}(x)}{m\gamma} + b\Gamma(t), \tag{2.48}$$

where $b = b'/m\gamma$. We see that to calculate the Kramers-Moyal coefficients given by Eq. (2.23) we require

an integration of Eq. (2.48) of the form

$$x(t+dt) - x(t) = \int_t^{t+dt} dt' \, \frac{1}{m\gamma}\mathcal{F}(x(t')) + \int_t^{t+dt} dt' \, b\Gamma(t') \tag{2.49}$$

where $\mathcal{F}(x)$ is the force due to the potential. Expanding about $\mathcal{F}(x(t))$ and averaging gives

$$\langle x(t+dt) - x(t)\rangle = \frac{1}{m\gamma}\mathcal{F}(x(t))dt + O(dt)^2 \tag{2.50}$$

in agreement with $A(x,t)$ from Eq. (2.44). The second coefficient in the expansion is given by

$$(x(t+dt) - x(t))^2 = \left[\int_t^{t+dt} dt' \, \frac{1}{m\gamma}\mathcal{F}(x(t')) + \int_t^{t+dt} dt' \, b\Gamma(t')\right]^2. \tag{2.51}$$

Upon averaging this yields

$$\begin{aligned}\langle (x(t+dt) - x(t))^2\rangle = &\left[\frac{1}{m\gamma}\mathcal{F}(x(t))dt + O(dt)^2\right]^2 \\ &+ 2b\int_t^{t+dt} dt' \int_t^{t+dt} dt'' \, \left\langle \frac{1}{m\gamma}\mathcal{F}(x(t'))\Gamma(t'')\right\rangle + b^2\int_t^{t+dt} dt' \int_t^{t+dt} dt'' \, \langle \Gamma(t')\Gamma(t'')\rangle.\end{aligned} \tag{2.52}$$

The first term is immediately identifiable as $O(dt)^2$ and the second term can be expanded such that

$$\begin{aligned}2b\int_t^{t+dt} dt' \int_t^{t+dt} dt'' \, \left\langle \frac{1}{m\gamma}\mathcal{F}(x(t'))\Gamma(t'')\right\rangle = &\frac{2b}{m\gamma}\mathcal{F}(x(t))dt\int_t^{t+dt} dt' \, \langle\Gamma(t')\rangle \\ &+ \frac{2b}{m\gamma}\frac{d\mathcal{F}(x(t))}{dx}\int_t^{t+dt} dt' \int_t^{t+dt} dt'' \, \langle (x(t') - x(t))\,\Gamma(t'')\rangle\end{aligned} \tag{2.53}$$

of which the terms either vanish or are $O(dt)^2$ leaving

$$\begin{aligned}\langle (x(t+dt) - x(t))^2\rangle &= b^2\int_t^{t+dt} dt' \int_t^{t+dt} dt'' \, \langle\Gamma(t')\Gamma(t'')\rangle \\ &= b^2\int_t^{t+dt} dt' \int_t^{t+dt} dt'' \, \delta(t'-t'') \\ &= b^2 dt.\end{aligned} \tag{2.54}$$

A similar approach then shows that all higher moments are at least order $(dt)^2$ and therefore vanish in the limit $dt \to 0$ meaning the probability distribution for dynamics of the form in Eq. (2.48) corresponds to the Fokker-Planck description of Eq. (2.44) if we choose the parameter b by comparing Eq. (2.54) to Eq. (2.43) such that

$$b = \sqrt{\frac{2k_BT}{m\gamma}}. \tag{2.55}$$

The Langevin equation that is represented by the relevant Fokker-Planck equation is therefore given by

$$\dot{x} = \frac{\mathcal{F}(x)}{m\gamma} + \sqrt{\frac{2k_BT}{m\gamma}}\Gamma(t). \tag{2.56}$$

This equation now allows us to describe single realisations of the dynamics of the Brownian particle that collectively produce probability distributions described by the Fokker-Planck representation of

Eq. (2.44). In this sense, despite the differing nature of the approaches, the two descriptions are considered equivalent. The key difference between them is that the Langevin formulation considers the stochastic behaviour as being completely wrapped up into a new idealised noise term which effects a normal deterministic equation of motion. As such this noise term must then have some rather uncomfortable properties owing to the white noise approximation in order to ensure the process is both Markov and continuous. Using this approach however, allows us to produce single trajectories by using a single realisation of the noise term which allows a further understanding and a direct consideration of distributions. One should exercise caution however because the Langevin equation or rather a generalised Langevin equation in which an arbitrary deterministic equation is altered by noise controlled by some other general function b is not guaranteed to relate to the observed stochastic behaviour x as it is not derived from a master equation such as Eq. (2.15). Instead care must be taken to ensure that the Kramers-Moyal coefficients given by the Langevin description faithfully represent the expected values from the macroscopic stochastic process. Specifically this can cause ambiguity when dealing with a noise strength dependent on the stochastic variable. This issue is further explored in Sect. 2.2.1.

2.2 Basic Properties of Stochastic Differential Equations

2.2.1 The Wiener Process and Stochastic Differential Equations

One of the most important processes in the study of stochastic systems allowing us to mathematically formalise equations such as Eq. (2.56) is also one of the most simple. Known as the Wiener process, it is described as a continuous random process with no drift such that all Kramers-Moyal coefficients are equal to zero except the second being equal to one. Consequently it obeys the forward equation

$$\frac{\partial p(x_1,t_1|x_0,t_0)}{\partial t_1} = \frac{1}{2}\frac{\partial^2 p(x_1,t_1,x_0,t_0)}{\partial x_1^2}. \tag{2.57}$$

We can solve this equation by defining the Fourier transform of $p(x_1,t_1|x_0,t_0)$ as

$$\tilde{p}(k,t_1) = \int_{-\infty}^{+\infty} dx_1\, p(x_1,t_1|x_0,t_0)e^{ikx_1} \tag{2.58}$$

allowing us to transform Eq. (2.57) to find

$$\frac{\partial \tilde{p}(k,t_1)}{\partial t_1} = -\frac{k^2}{2}\tilde{p}(k,t_1). \tag{2.59}$$

This then gives the solution in reciprocal space

$$\tilde{p}(k,t_1) = \tilde{p}(k,t_0)e^{-\frac{k^2}{2}(t_1-t_0)}. \tag{2.60}$$

Applying the appropriate initial condition, $p(x_1,t_0|x_0,t_0) = \delta(x_1-x_0)$ we have

$$\tilde{p}(k,t_0) = e^{ikx_0}. \tag{2.61}$$

The solution is then given as an inverse Fourier transform

$$\begin{aligned} p(x_1,t_1|x_0,t_0) &= \frac{1}{2\pi}\int_{-\infty}^{+\infty} e^{-ik(x_1-x_0)-\frac{k^2}{2}(t_1-t_0)}dk \\ &= \frac{1}{\sqrt{2\pi(t_1-t_0)}}\exp\left[-\frac{(x_1-x_0)^2}{2(t_1-t_0)}\right]. \end{aligned} \tag{2.62}$$

This then demonstrates the fundamental properties of the Wiener process. The probability density function is Gaussian with zero mean and variance proportional to the elapsed time. Relabelled such that for the Wiener process $x(t) = W(t)$ from an initial condition $W(t_0) = W_0$ this means we have

$$\langle W(t)\rangle = W_0 \tag{2.63}$$

and

$$\langle (W(t)-W_0)^2\rangle = t - t_0. \tag{2.64}$$

Further we shall find use for the variance of $(W(t)-W_0)^2$ which is straightforwardly

$$\begin{aligned} \langle ((W(t)-W_0)^2)^2 - (t-t_0)^2\rangle &= \langle ((W(t)-W_0)^2)^2\rangle - (t-t_0)^2 \\ &= \langle (W(t)-W_0)^4\rangle - (t-t_0)^2 \\ &= \int_{-\infty}^{+\infty} d(W(t)-W_0)\,(W(t)-W_0)^4 \\ &\qquad \times\sqrt{\frac{1}{2\pi(t-t_0)}}\exp\left[-\frac{(W(t)-W_0)^2}{2(t-t_0)}\right] - (t-t_0)^2 \\ &= 3(t-t_0)^2 - (t-t_0)^2 \\ &= 2(t-t_0)^2. \end{aligned} \tag{2.65}$$

Another, somewhat uncomfortable, property of the Wiener process is that it is non-differentiable. This property is the result of an unbounded variation in the sampling path of $W(t)$ (that is to say total accumulated motion of its projection on the $W(t)$ axis with time) arising from a self similar, almost fractal, behaviour consistent with it being both continuous and Markovian whilst having the same mean and variance characteristics on any timescale. As a result, various limits, when taken in the form of a gradient, diverge and do so depending on precisely how that limit is taken [41]. If we were now to consider the Wiener process in terms of its corresponding Langevin equation we would have

$$\dot{W} = \Gamma(t). \tag{2.66}$$

This now contains $\Gamma(t)$ as the derivative of the supposedly non-differentiable function with such a contradiction reflected in the white noise idealisation and corresponding infinite variance. To proceed without this explicit contradiction we must form integral equations from the corresponding Langevin equation such that Eq. (2.66) is interpreted as

$$W(t) = \int_0^t \Gamma(t')dt'. \tag{2.67}$$

This avoids the derivative of the Wiener process by using an integral which is, in principle, well defined. Similarly the Langevin equation

$$\dot{x} = a(x,t) + b(x,t)\Gamma(t) \tag{2.68}$$

should also be interpreted as an integral equation, a fact first pointed out by Doob [42],

$$x(t) - x(0) = \int_0^t a(x, t')dt' + \int_0^t b(x, t')\Gamma(t')dt'. \tag{2.69}$$

Following from the interpretation of Eq. (2.67) we identify a small change in the Wiener process as

$$dW = W(t + dt) - W(t) = \Gamma(t)dt. \tag{2.70}$$

Doing so then allows us to consider the evolution of Eq. (2.69) in terms of the Wiener process such that

$$x(t) - x(0) = \int_0^t a(x, t')dt' + \int_0^t b(x, t')dW(t'). \tag{2.71}$$

This is now known as a stochastic integral equation. Whilst Eq. (2.67) avoided the use of the differential of the Wiener process this ambiguity is still manifest in the second integral of Eq. (2.71). In the above form the interpretation of the first integral is unambiguous and independent of any limiting summation procedure since dt behaves smoothly and with bounded variation, reflecting the usual deterministic behaviour. The second integral, however, requires extra care. Since dW is non-differentiable it has unbounded variation and so is not a smooth function *on any timescale*. As such there is an inherited ambiguity as to where to evaluate the function $b(x, t)$ when it too depends on the non-differentiable Wiener process through x. Considering a Riemann-Stieltjes integral representation of the second term in Eq. (2.69), by writing it as the limit of a discrete sum of the partition $t = t_0, t_1, t_2 \dots t_n = t + dt$, we can write

$$\int_t^{t+dt} b(x, t')dW(t') = \lim_{n\to\infty} \sum_{i=0}^{n-1} \left[(1 - \alpha)b(x(t_i), t_i) + \alpha b(x(t_{i+1}), t_{i+1})\right] (W(t_{i+1}) - W(t_i)). \tag{2.72}$$

Crucially, the result is not independent of the choice of α. This is because neither $b(x)$ nor dW behaves smoothly as $dt \to 0$ and $b(x(t_{i+1}), t_{i+1})$ is correlated with the corresponding increment of the Wiener process whilst $b(x(t_i), t_i)$ is not. Consequently to make Eq. (2.71) well defined we much choose a set of integration rules with the most common being that of Itō and that of Stratonovich corresponding to $\alpha = 0$ and $\alpha = 0.5$ respectively, but also more recently that of Hänggi and Klimontovich corresponding to $\alpha = 1.0$. We now understand that to properly specify processes described by Langevin equations such as Eq. (2.56) we must consider their integrated form for which we then choose an integration scheme. Eq. (2.68) is therefore described properly by Eq. (2.71) supplemented by a choice usually between Itō, Stratonovich and Hänggi-Klimontovich. We write this choice in short hand as

$$dx = a(x, t)dt + b(x, t)dW \tag{2.73}$$

for the Itō interpretation,

$$dx = a(x, t)dt + b(x, t) \circ dW \tag{2.74}$$

for the Stratonovich interpretation and

$$dx = a(x, t)dt + b(x, t) \bullet dW \tag{2.75}$$

for the Hänggi-Klimontovich interpretation with each producing a different path for x owing to the degree of correlation of the integrand with the integrator under the chosen scheme. Considering a small time step dt we now consider them as stochastic differential equations (SDEs) and allows us to unambiguously

define continuous stochastic processes mathematically. Interpreting a Langevin equation in this sense allows us to consider the microscopic behaviour of the Brownian particle much more robustly. We now consider individual trajectories as solutions to a stochastic differential equation which are dependent upon the increment of a well understood process, namely the Wiener process. Interpreting Eq. (2.56) as an Itō stochastic differential equation we would formally have

$$dx = \frac{\mathcal{F}(x)}{m\gamma}dt + \sqrt{\frac{2k_BT}{m\gamma}}dW. \tag{2.76}$$

2.2.2 Properties of Stochastic Differential Equations Under Different Interpretations

The Itō Integral and Itō Calculus

The Itō interpretation of the stochastic integral can be thought of the limit of a sum where the random force or Wiener increment occurs at the *beginning* of the time step and can be represented as

$$\int_t^{t+\tau} f(x)dW = \lim_{n\to\infty}\sum_{i=0}^{n-1} f(x_i)(W_{i+1} - W_i) \tag{2.77}$$

for the partition $t = t_0, t_1, \ldots, t_n = t+\tau$ where $x_i = x(t_i)$ and $W_i = W(t_i)$ used throughout. Examining its form we notice that if the function $f(x)$ is *non-anticipating*, that is if it does not depend on any information following time t_i, the increment in the Wiener process is uncorrelated with the value of the function $f(x)$ at each time step. Consequently an advantage often associated with the Itō interpretation is the intuitive result

$$\left\langle \int_t^{t+\tau} f(x)dW \right\rangle = \lim_{n\to\infty}\sum_{i=0}^{n-1}\langle f(x_i)(W_{i+1} - W_i)\rangle = \lim_{n\to\infty}\sum_{i=0}^{n-1}\langle f(x_i)\rangle\langle (W_{i+1} - W_i)\rangle = 0 \tag{2.78}$$

which follows simply from the statistical properties of the Wiener process. This property can be exploited not only when considering the mean of a stochastic process, but by naturally forming the basis of a simple numerical integration scheme where the integrals are approximated for a short time dt as

$$\int_t^{t+dt} b(x,t')dW(t') \simeq b(x,t)\int_t^{t+dt} dW(t') \simeq b(x,t)(W(t+dt) - W(t)) \tag{2.79}$$

and

$$\int_t^{t+dt} a(x,t')dt' \simeq a(x,t)dt. \tag{2.80}$$

A numerical solution can then be constructed by allowing us to imagine the process as the limit of a forward Euler integration scheme consisting of both a deterministic and a stochastic component. The details of such a numerical approach are given in appendix A.

Now considering a physical process f that depends on the stochastic variable dx given by a stochastic differential equation one would expect that the physical process too is non-deterministic and will obey its own stochastic dynamics. Ordinarily one could describe the evolution of this resultant process simply using the chain rule of normal calculus, however the evolution of x over some time dt depends upon the Wiener process which is the result of an infinite number of fluctuations and as shown above obeys the scaling shown in Eq. (2.64). In some sense we can therefore expect a small change in the Wiener

process to be proportional to $dt^{1/2}$. The specific case of Itō interpretations of the stochastic integral illustrates how we can treat this dependence upon the Wiener process most clearly since the integrand is uncorrelated with the update in the Wiener process and as such depends the most trivially upon it. Using the Itō interpretation yields an important result called the Itō isometry [41] which we shall not formally derive here, but is stated as

$$\left\langle \left(\int x dW \right)^2 \right\rangle = \left\langle \int x^2 dt \right\rangle \tag{2.81}$$

where we can think of the result arising from Eq. (2.64) or rather restated in the limit $dt \to 0$, $\langle (dW)^2 \rangle = dt$. This result can crucially be extended further to the properties of a *specific realisation* of the Wiener increment given an infinitesimal time step as opposed to equality just in expectation. We can heuristically demonstrate this by understanding that any increment in the Wiener process, however small, is still the result of an arbitrarily large number of independent contributions, and examining the statistical properties of the square of an increment in the Wiener process of arbitrary size, $(\Delta W)^2$. Since $\langle \Delta W \rangle = 0$, the mean of $(\Delta W)^2$ is simply equal to the variance Δt, whilst the variance of $(\Delta W)^2$ is given by Eq. (2.65). Consequently as we reduce the time step $\Delta t \to dt$ the variance of the corresponding increment in the Wiener process vanishes faster than the mean such that for an infinitesimal step dt we can consider the variance to vanish so we have $(dW)^2 = dt$ valid for all dW. If we now consider the process $f(x)$ to be uncorrelated with an increment dx when considering multiplications of the form $f(x)dx$ such that the Itō interpretation is followed this has the consequence of introducing extra terms into the usual chain rule producing a result known as Itō's Lemma. It can be informally derived by Taylor expanding the function $f(x,t)$ but retaining higher order terms until application of the statistical properties of the Wiener process. Starting from the Taylor expansion about $f(x,t)$ for small time dt we have

$$df = \frac{\partial f}{\partial x}dx + \frac{1}{2}\frac{\partial^2 f}{\partial x^2}(dx)^2 + \frac{\partial f}{\partial t}dt + \frac{1}{2}\frac{\partial^2 f}{\partial t^2}(dt)^2 + \frac{\partial^2 f}{\partial x \partial t}dxdt. \tag{2.82}$$

Now inserting an Itō SDE of form Eq. (2.73) for the small increment $(dx)^2$ and cross term $dxdt$ we have

$$df = \frac{\partial f}{\partial x}dx + \frac{\partial f}{\partial t}dt + \frac{1}{2}\frac{\partial^2 f}{\partial t^2}(dt)^2 + \frac{1}{2}\frac{\partial^2 f}{\partial x^2}(a^2(dt)^2 + 2abdtdW + b^2(dW)^2) + \frac{\partial^2 f}{\partial x \partial t}(adt + bdW)dt. \tag{2.83}$$

Only now can we ignore terms higher than order one in dt by heuristically considering $(dW)^2 = dt$ argued above and thus $dW \sim dt^{1/2}$ and $dtdW \sim dt^{3/2}$. We note more rigorous proofs for these results exist by explicitly considering summation limits of the integral forms which dx represents [38]. Retaining terms that are then only first order in dt or lower we find

$$df = \left(\frac{\partial f}{\partial t} + \frac{1}{2}b^2 \frac{\partial^2 f}{\partial x^2} \right) dt + \frac{\partial f}{\partial x}dx \tag{2.84}$$

which acts as a chain rule for Itō SDEs and is known as Itō's Lemma. The application of the heuristic rule $dW \sim dt^{1/2}$ can be extended to deal with other situations encountered within normal calculus. A further consequence of keeping higher terms is the retention of a cross term in the Itō product rule. This states that for two functions X and Y that depend on the same stochastic variable the difference of their product is given by

$$d(XY) = XdY + YdX + dXdY \tag{2.85}$$

assuming the stochastic process for both functions follows the same Wiener process.

Stratonovich Integrals

Application of the above rules is generally referred to as Itō calculus since the results are readily applied under Itō integration because of the associated lack of correlation in its formulation. Integrals formulated under Stratonovich do not follow the Itō calculus as they do not possess this property. However, since the Stratonovich integral is just an altered summation over the same increments in the Wiener process we can relate the two by constructing the value of the integrand under Stratonovich using Itō calculus rules [38]. We then use such a procedure as an example for how one can relate an arbitrary stochastic integral, that is any choice $\alpha \in [0:1]$, using the same procedure. As a first case we consider the Stratonovich integral found in an SDE written in the shorthand $dx = adt + b \circ dW$ such that we explicitly have

$$\begin{aligned}
\int_t^{t+\tau} b(x,t) \circ dW &= \lim_{n\to\infty} \sum_{i=0}^{n-1} \left[\frac{b(x_i) + b(x_{i+1})}{2} \right] dW_i \\
&= \lim_{n\to\infty} \sum_{i=0}^{n-1} \left[b(x_i) + \frac{db(x_i)}{2} \right] dW_i \\
&= \lim_{n\to\infty} \sum_{i=0}^{n-1} \left[b(x_i) + \frac{1}{2} \left(\frac{\partial b}{\partial x} dx_i + \frac{1}{2} b^2 \frac{\partial^2 b}{\partial x^2} dt_i + \frac{\partial b}{\partial t} dt_i \right) \right] dW_i \\
&= \lim_{n\to\infty} \sum_{i=0}^{n-1} \left[b(x_i) + \frac{1}{2} \left(\frac{\partial b}{\partial x} (a dt_i + b dW_i) + \frac{1}{2} b^2 \frac{\partial^2 b}{\partial x^2} dt_i + \frac{\partial b}{\partial t} dt_i \right) \right] dW_i
\end{aligned} \tag{2.86}$$

where $dW_i = W_{i+1} - W_i$, $dx_i = x(t_{i+1}) - x(t_i)$ and $dt_i = t_{i+1} - t_i$. By ignoring all terms higher than dt by considering $dW \sim dt^{1/2}$ we find

$$\begin{aligned}
\int_t^{t+\tau} b(x,t) \circ dW &= \lim_{n\to\infty} \sum_{i=0}^{n-1} b(x_i) dW_i + \lim_{n\to\infty} \sum_{i=0}^{n-1} \frac{1}{2} b(x_i, t_i) \frac{\partial b(x_i, t_i)}{\partial x_i} dt_i \\
&= \int_t^{t+\tau} b(x,t) dW + \int_t^{t+\tau} \frac{1}{2} b(x,t) \frac{\partial b(x,t)}{\partial x} dt.
\end{aligned} \tag{2.87}$$

Consequently, after returning to the shorthand notation, we can arrive at the important identity

$$dx = a(x,t)dt + b(x,t) \circ dW = \left[a(x,t) + \frac{1}{2} b(x,t) \frac{\partial b(x,t)}{\partial x} \right] dt + b(x,t) dW. \tag{2.88}$$

This means that an Itō and Stratonovich SDE will produce the *same solution* $x(t)$ from the same Wiener process dW if related in this way. More importantly it provides a method for us to transform between the two different interpretations. Generalising by considering the arbitrary stochastic increment dx we can ask how the integral over an x dependent function $g(x,t)$ is related to the manner in which we choose to integrate. By introducing the Stratonovich integral

$$\int g(x,t) \circ dx \tag{2.89}$$

we first note we can construct dx as the increment in an *Itō* process since we are only relating the correlation of $f(x)$ and dx which if produced under Stratonovich rules could easily be transformed to an

Itō form anyway. As such we construct

$$\begin{aligned}\int_t^{t+\tau} g(x,t)\circ dx &= \int_t^{t+\tau} g(x)\circ(adt+bdW)\\ &= \lim_{n\to\infty}\sum_{i=0}^{n-1}\left[g(x_i)+\frac{dg(x_i)}{2}\right](adt_i+bdW_i)\\ &= \lim_{n\to\infty}\sum_{i=0}^{n-1}\left[g(x_i)+\frac{1}{2}\left(\frac{\partial g}{\partial x}dx_i+\frac{1}{2}b^2\frac{\partial^2 g}{\partial x^2}dt_i+\frac{\partial g}{\partial t}dt_i\right)\right](adt_i+bdW_i)\\ &= \lim_{n\to\infty}\sum_{i=0}^{n-1}\left[g(x_i)+\frac{1}{2}\left(\frac{\partial g}{\partial x}(adt_i+bdW_i)+\frac{1}{2}b^2\frac{\partial^2 g}{\partial x^2}dt_i+\frac{\partial g}{\partial t}dt_i\right)\right](adt_i+bdW_i).\end{aligned} \tag{2.90}$$

By dropping all terms higher than dt we then obtain the general relation

$$\int g(x,t)\circ dx = \int g(x,t)dx+\frac{1}{2}\int\frac{\partial g(x,t)}{\partial x}b^2dt. \tag{2.91}$$

Once again we point out that this result is independent of the nature of the SDE generating the increment dx a fact clarified by the fact that the correction term doesn't contain the drift (or modified drift) a. Returning once again to the shorthand notation this result has a specific consequence if we choose $g(x,t)=\partial f(x,t)/\partial x$ giving

$$\frac{\partial f}{\partial x}\circ dx = \frac{\partial f}{\partial x}dx+\frac{1}{2}b^2\frac{\partial^2 f}{\partial x^2}dt \tag{2.92}$$

which when substituted into Itō's lemma gives

$$df = \frac{\partial f}{\partial t}dt+\frac{\partial f}{\partial x}\circ dx \tag{2.93}$$

such that if we follow the *Stratonovich* summation conventions the *normal rules of calculus* are obeyed.

Connection with Generalised Fokker-Planck Equations

It was shown before that a simple Langevin equation possesses the same Kramers-Moyal coefficients as the Fokker-Planck equation for the case of an inhomogeneous medium when the forcing term due to the environment was considered as white noise. However the connection between continuous Markov processes and deterministic partial differential equations can be derived much more directly and generally by formulation as stochastic differential equations which we know how to manipulate much more robustly. Taking the general form of an Itō SDE

$$dx = a(x,t)dt+b(x,t)dW \tag{2.94}$$

we can use Itō's lemma to describe the evolution of a arbitrary spatially dependent function $f(x)$

$$df = \frac{\partial f}{\partial x}adt+\frac{1}{2}b^2\frac{\partial^2 f}{\partial x^2}dt+\frac{\partial f}{\partial x}bdW \tag{2.95}$$

and then use the favourable quality of the Itō integral to trivially calculate the rate of change of the mean of $f(x)$ such that

$$\frac{d\langle f(x)\rangle}{dt} = \left\langle\frac{\partial f}{\partial x}a\right\rangle+\left\langle\frac{b^2}{2}\frac{\partial^2 f}{\partial x^2}\right\rangle. \tag{2.96}$$

We can then perform the averaging of the terms on the right hand side by integrating by parts over the probability distribution function $p(x,t)$. Doing so yields

$$\begin{aligned}\frac{d\langle f(x)\rangle}{dt} &= \int_{-\infty}^{+\infty} dx\; p(x,t)a(x,t)\frac{\partial f(x)}{\partial x} + \frac{1}{2}\int_{-\infty}^{+\infty} dx\; p(x,t)b^2(x,t)\frac{\partial^2 f(x)}{\partial x^2} \\ &= \left[p(x,t)a(x,t)f(x)\right]_{-\infty}^{+\infty} - \int_{-\infty}^{+\infty} dx\; f(x)\frac{\partial}{\partial x}\left[a(x,t)p(x,t)\right] \\ &\quad + \frac{1}{2}\left[p(x,t)b^2(x,t)\frac{\partial f(x)}{\partial x}\right]_{-\infty}^{+\infty} - \frac{1}{2}\int_{-\infty}^{+\infty} dx\; \frac{\partial f(x)}{\partial x}\frac{\partial}{\partial x}\left[b^2(x,t)p(x,t)\right] \\ &= \left[p(x,t)a(x,t)f(x)\right]_{-\infty}^{+\infty} - \int_{-\infty}^{+\infty} dx\; f(x)\frac{\partial}{\partial x}\left[a(x,t)p(x,t)\right] \\ &\quad + \frac{1}{2}\left[p(x,t)b^2(x,t)\frac{\partial f(x)}{\partial x}\right]_{-\infty}^{+\infty} - \frac{1}{2}\left[\frac{\partial}{\partial x}\left[p(x,t)b^2(x,t)\right]f(x)\right]_{-\infty}^{+\infty} \\ &\quad + \frac{1}{2}\int_{-\infty}^{+\infty} dx\; f(x)\frac{\partial^2}{\partial x^2}\left[b^2(x,t)p(x,t)\right]. \end{aligned} \tag{2.97}$$

Discarding surface terms we then equate with the simple form of the mean of $f(x)$ given by

$$\frac{d\langle f(x)\rangle}{dt} = \frac{d}{dt}\int_{-\infty}^{+\infty} f(x)p(x,t)dx = \int_{-\infty}^{+\infty} f(x)\frac{\partial p(x,t)}{\partial t}dx \tag{2.98}$$

which gives

$$0 = \int_{-\infty}^{+\infty} dx\; f(x)\left[-\frac{\partial p(x,t)}{\partial t} - \frac{\partial}{\partial x}\left[a(x,t)p(x,t)\right] + \frac{1}{2}\frac{\partial^2}{\partial x^2}\left[b^2(x,t)p(x,t)\right]\right]. \tag{2.99}$$

Consequently we can describe the probability distribution of the random variable x given by an Itō SDE $dx = adt + bdW$ by the partial differential equation

$$\frac{\partial p(x,t)}{\partial t} = -\frac{\partial}{\partial x}\left[a(x,t)p(x,t)\right] + \frac{1}{2}\frac{\partial^2}{\partial x^2}\left[b^2(x,t)p(x,t)\right]. \tag{2.100}$$

The adaptation to a Stratonovich SDE is simple and is performed using the transformation formula of Eq. (2.88) and gives

$$\begin{aligned}\frac{\partial p(x,t)}{\partial t} &= -\frac{\partial}{\partial x}\left[\left(a(x,t) + \frac{1}{2}b(x,t)\frac{\partial b(x,t)}{\partial x}\right)p(x,t)\right] + \frac{1}{2}\frac{\partial^2}{\partial x^2}\left[b^2(x,t)p(x,t)\right] \\ &= -\frac{\partial}{\partial x}\left[a(x,t)p(x,t)\right] + \frac{1}{2}\frac{\partial}{\partial x}\left[b(x,t)\frac{\partial}{\partial x}\left[b(x,t)p(x,t)\right]\right]. \end{aligned} \tag{2.101}$$

Similarly, the equivalent Fokker-Planck equation for a Hänggi-Klimontovich interpretation is given by

$$\begin{aligned}\frac{\partial p(x,t)}{\partial t} &= -\frac{\partial}{\partial x}\left[\left(a(x,t) + b(x,t)\frac{\partial b(x,t)}{\partial x}\right)p(x,t)\right] + \frac{1}{2}\frac{\partial^2}{\partial x^2}\left[b^2(x,t)p(x,t)\right] \\ &= -\frac{\partial}{\partial x}\left[a(x,t)p(x,t)\right] + \frac{1}{2}\frac{\partial}{\partial x}\left[b^2(x,t)\frac{\partial p(x,t)}{\partial x}\right]. \end{aligned} \tag{2.102}$$

Multidimensional Processes and General Stochastic Integrals

One may generalise the above to multidimensional Itō processes which are generally of the form

$$d\boldsymbol{x} = \boldsymbol{A}(\boldsymbol{x})dt + \boldsymbol{B}(\boldsymbol{x})d\boldsymbol{W} \tag{2.103}$$

where all bold quantities are vectors of the form $\boldsymbol{x} = (x_1, x_2 \dots x_n)$ and all W_i are independent uncorrelated Wiener processes, except $\boldsymbol{B}(\boldsymbol{x})$ which is a matrix so that one can write the evolution for one variable as

$$dx_i = A_i(\boldsymbol{x})dt + \sum_j B_{ij}(\boldsymbol{x})dW_j \tag{2.104}$$

noting that dW_j here is taken to mean an increment in the independent Wiener process labelled j rather than the increment in the Wiener process corresponding to the jth increment dt_j of an approximated integral as written elsewhere, implicit in the absence of such a summation structure. The corresponding Fokker-Planck equation then is governed by the diffusion matrix $\boldsymbol{D} = (1/2)\boldsymbol{B}\boldsymbol{B}^{\mathrm{T}}$ so that $D_{ij} = (1/2)\sum_k B_{ik}B_{jk}$. We may then proceed to find the relevant conversion formulae for an arbitrary stochastic integral denoted $*$ characterised by evaluation point $(1-\alpha)x_i + \alpha x_{i+1}$ noting that

$$\int_t^{t+\tau} g(\boldsymbol{x}) * dW = \lim_{n\to\infty} \sum_{i=0}^{n-1} \left[g(\boldsymbol{x}_i) + \alpha dg(\boldsymbol{x}_i)\right] dW_i, \tag{2.105}$$

$dW_i dW_j = \delta_{ij}dt$ and using the same reasoning to find the analogous conversion formulae

$$\begin{aligned} g(\boldsymbol{x}) * dx_i &= g(\boldsymbol{x})dx_i + \alpha \sum_j \frac{\partial g(\boldsymbol{x})}{\partial x_j} \left(\sum_m B_{jm}(\boldsymbol{x})dW_m\right)\left(\sum_n B_{in}(\boldsymbol{x})dW_n\right) \\ &= g(\boldsymbol{x})dx_i + \alpha \sum_j \frac{\partial g(\boldsymbol{x})}{\partial x_j} \left(\sum_k B_{ik}(\boldsymbol{x})B_{jk}(\boldsymbol{x})\right) dt \\ &= g(\boldsymbol{x})dx_i + 2\alpha \sum_j \frac{\partial g(\boldsymbol{x})}{\partial x_j} D_{ij}(\boldsymbol{x})dt \end{aligned} \tag{2.106}$$

and

$$\begin{aligned} \sum_j B_{ij}(\boldsymbol{x}) * dW_j &= \sum_j B_{ij}(\boldsymbol{x})dW_j + \alpha \sum_j \left[\left(\sum_k \frac{\partial B_{ij}(\boldsymbol{x})}{\partial x_k}\left(A_k(\boldsymbol{x})dt + \sum_n B_{kn}(\boldsymbol{x})dW_n\right)\right) dW_j\right] \\ &= \sum_j B_{ij}(\boldsymbol{x})dW_j + \alpha \sum_j \sum_k \frac{\partial B_{ij}(\boldsymbol{x})}{\partial x_k} B_{kj}(\boldsymbol{x})dt \end{aligned} \tag{2.107}$$

leading to a drift term correction

$$A_i(\boldsymbol{x}) \to A_i(\boldsymbol{x}) + \alpha \sum_j \sum_k \frac{\partial B_{ij}(\boldsymbol{x})}{\partial x_k} B_{kj}(\boldsymbol{x}). \tag{2.108}$$

The Itō-Stratonovich Dilemma

Examining the form of Eqs. (2.108), (2.100) and (2.101) we observe that each interpretation of the stochastic integral produces *measurably different* behaviour in the stochastic variable when there is *multiplicative noise* which occurs when b is x dependent even if the form of the SDE is obtained from a seemingly unambiguous single Langevin equation. This situation can arise in the case of diffusion in an inhomogeneous medium for example. The question that arises is how should one interpret this Langevin equation as an SDE? This is the Itō-Stratonovich dilemma and is the decision about which integration scheme is suitable in which situation because of the idealised white noise approximation. Much work has been done on which choice is optimal in which situation with arguments suggesting the Stratonovich interpretation is more suitable for systems where the white noise is an approximation of a noise with finite correlation [43]. This is frequently the case in real physical systems since the Langevin equation is often

a coarse grained form of a process under the influence of a correlated force. On the other hand when the SDE is a continuous approximation to a discrete process or when the noise exists as a series of separated pulses the Itō form may be more appropriate. However there are strong arguments that suggest when the noise is intrinsic to the model, or rather internal, such that it cannot be trivially switched off (as is the case for Brownian motion) no choice is correct as a Langevin equation is not sufficient to describe the behaviour [40]. The argument loosely states that in such a case the Kramers-Moyal coefficients are unspecifiable when it is impossible to conceive of the system with the noise 'switched off' which if possible would allow one to measure the underlying drift and when there is no physical argument to justify the somewhat arbitrary additional drift from a non-Itō interpretation. Consequently one must exercise caution when faced with a system with multiplicative noise.

Of course for the case of a homogeneous medium we have the special case of *additive noise* which exists when b has no x dependence meaning any interpretation of the stochastic integral produces the same result. When this is the case both formulations are equivalent and the choice is essentially arbitrary. A question that does arise in this situation, however, is how one should utilise the random increment dx in conjunction with other x dependent functions, for example multiplications of the form $f(x)dx$. Here despite there being no ambiguity in the generation of the increment dx one must specify an integration choice over other functions. It is in this situation that for the modelling of physical systems where each part of the summation limit of a stochastic integral represents some kind of coarse graining of a finite process with finite correlation that the strictly non-anticipating quality of the function $f(x)$ in the Itō integral is deemed inappropriate [44, 45]. Consequently when these situations arise the convention of Stratonovich is usually followed along with a Stratonovich interpretation of the underlying SDE as a starting point for more complex behaviour. These specific issues are revisited in Sect. 3.3 when such multiplications are attributed to thermodynamic quantities such as work done.

2.3 Constructing Path Probability Functionals

2.3.1 Master Equations

We will find later that we wish to discuss and compare the likelihood of certain sequences or trajectories occurring in time in probabilistic terms. To do so we must write such a quantity in terms of the behaviour of the system. We may describe such a quantity as a functional of a general trajectory which we define as $\vec{\boldsymbol{x}} = \boldsymbol{x}(t)$ for a suitable interval in time, where an instantaneous value $\boldsymbol{x}$ represents a general state label which may specify any number of individual system properties such that $\boldsymbol{x} = (x_0, x_1 \dots x_n)$. Considering a master equation of the form in Eq. (2.18), we must realise that as it is in continuous time, the exact probability of observing any one particular trajectory is strictly zero. Consequently the quantity we consider is necessarily a probability density with dimension proportional to the number of transitions which occur in the path $\vec{\boldsymbol{x}}$. Considering again a master equation of the form in Eq. (2.18) we may describe the trajectory explicitly by the discrete sequence $\boldsymbol{x}_i$ where transitions from $\boldsymbol{x}_{i-1}$ to $\boldsymbol{x}_i$ occur at times t_i. For a sequence of N transitions between $N+1$ configurations between times $t = t_0 = 0$ and $t = t_{N+1} = \tau$ we can construct the probability of observing the sequence $\boldsymbol{x}_i$ with transitions occurring within an infinitesimal time dt_i around each transition time t_i as the product of a single initial probability distribution, N probabilities of jumps occurring within the infinitesimal time dt_i and $N+1$ probabilities

of observing no jumps in the resultant intervals so that

$$P[\vec{\boldsymbol{x}}] = P_{\text{initial}}(\boldsymbol{x}_0, t_0) P_{\text{no transition}}(\boldsymbol{x}_0, t_0 \to t_1) \prod_{i=1}^{N} P_{\text{transition in } dt}(\boldsymbol{x}_i | \boldsymbol{x}_{i-1}, t_i) P_{\text{no transition}}(\boldsymbol{x}_i, t_i \to t_{i+1}). \tag{2.109}$$

The initial condition can be arbitrary, however the probability of a transition, within an infinitesimal time dt_i, must be related to the rate appearing in the master equation such that

$$P_{\text{transition in } dt}(\boldsymbol{x}_i | \boldsymbol{x}_{i-1}, t_i) = T(\boldsymbol{x}_i | \boldsymbol{x}_{i-1}, t_i) dt_i. \tag{2.110}$$

As such we can then write the total probability in terms of the probability density associated with the path

$$\begin{aligned} P[\vec{\boldsymbol{x}}] &= P_{\text{initial}}(\boldsymbol{x}_0, t_0) P_{\text{no transition}}(\boldsymbol{x}_0, t_0 \to t_1) \prod_{i=1}^{N} T(\boldsymbol{x}_i | \boldsymbol{x}_{i-1}, t_i) dt_i P_{\text{no transition}}(\boldsymbol{x}_i, t_i \to t_{i+1}) \\ &= p[\vec{\boldsymbol{x}}] \prod_{i=1}^{N} dt_i. \end{aligned} \tag{2.111}$$

The probability of having no transition between times t_{i-1} and t_i, given no transition by time t_{i-1}, is equal to the integral from t_i to infinity of the exponential probability density function as it is the only distribution which is memory-less so as to align with the Markov property. This can be illustrated simply for an homogeneous process with mean escape rate $\lambda = -T(\boldsymbol{x}_i | \boldsymbol{x}_i)$. In this instance the probability density function is

$$p(t) = \lambda \exp(-\lambda t). \tag{2.112}$$

First we consider the probability of not having made a transition by a time t_0, equal to the integral of the distribution from t_0 to infinity, thus given by

$$P(t > t_0) = \exp(-\lambda t_0). \tag{2.113}$$

In order to be Markovian, the probability of the transition occurring at some time $t > t_0$, both measured from an arbitrary time origin s_0, should be independent of s_0. If we let $s > (s_0 + t_0)$ be the time of the same transition measured from an earlier origin then this is described by the quantity $P(t > t_0 | s > s_0)$. As such we may write

$$\begin{aligned} P(t > t_0 | s > s_0) &= P(s > t_0 + s_0 | s > s_0) \\ &= P(s > t_0 + s_0, s > s_0) / p(s > s_0) \\ &= P(s > t_0 + s_0) / p(s > s_0) \\ &= \exp(-\lambda(t_0 + s_0)) \exp(\lambda s_0) \\ &= \exp(-\lambda t_0) \\ &= P(t > t_0) \end{aligned} \tag{2.114}$$

as required. By considering a spatially and temporally inhomogeneous processes such that we have $T(\boldsymbol{x}_i | \boldsymbol{x}_i, t) = -\lambda(\boldsymbol{x}_i, t)$ again as function of time and current state we can identify the particular exponential form by constructing the probability of having no transition as the limit of a product of infinitesimal probabilities of having no transition each over a short time dt. The probability of a transition in an interval dt by the above methods is $\lambda(\boldsymbol{x}_i, t) dt + O(dt^2)$ meaning that, for division of the finite

interval $t_i = t'_0, t'_1, \ldots, t'_N = t_{i+1}$ with $dt'_j = t'_{j+1} - t'_j$, we have

$$\begin{aligned} P_{\text{no transition}}(\boldsymbol{x}_i, t_i \to t_{i+1}) &= \lim_{N\to\infty} \prod_{j=0}^{N-1} (1 - \lambda(\boldsymbol{x}_i, t'_j) dt'_j) \\ &= \lim_{N\to\infty} \prod_{j=0}^{N-1} (1 + T(\boldsymbol{x}_i|\boldsymbol{x}_i, t'_j) dt'_j) \\ &= \exp\left[\int_{t_i}^{t_{i+1}} dt'\, T(\boldsymbol{x}_i|\boldsymbol{x}_i, t')\right]. \end{aligned} \tag{2.115}$$

Consequently we write the probability density of observing the trajectory $\vec{\boldsymbol{x}}$

$$p[\vec{\boldsymbol{x}}] = P(\boldsymbol{x}(0)) e^{\int_{t_0}^{t_1} dt' T(\boldsymbol{x}_0|\boldsymbol{x}_0, t')} \prod_{i=1}^{N} T(\boldsymbol{x}_i|\boldsymbol{x}_{i-1}, t_i) e^{\int_{t_i}^{t_{i+1}} dt' T(\boldsymbol{x}_i|\boldsymbol{x}_i t')}. \tag{2.116}$$

2.3.2 Stochastic Differential Equations

Similarly, we can construct a similar quantity in terms of the continuous approach utilised in stochastic differential equations. For the SDE given by Eq. (2.76) the probability density function is given by the Fokker-Planck equation of Eq. (2.44) and as such we should be able to describe the statistics of a path using conditional probabilities derived from that formalism. Generalising Eq. (2.44) in operator form we have

$$\frac{\partial p(x,t)}{\partial t} = \hat{L} p(x,t). \tag{2.117}$$

We expect from our formulation of the stochastic process outlined in Sect. 2.1.2, specifically Eq. (2.9), that $p(x,t)$ can be written as a path integral solution using the transition probability $p(x,t|x',t')$ which yields

$$p(x,t) = \int dx'\, p(x',t') p(x,t|x',t'). \tag{2.118}$$

We recognise this as the conditional probability which acts as the Green's function of Eq. (2.117). A system in which the conditional probability can be readily derived is that of the harmonic oscillator in the over-damped limit where $V(x) = \frac{1}{2}\kappa x^2$. The probability density function describing the evolution of such a system therefore obeys the Fokker-Planck equation

$$\frac{\partial p(x,t)}{\partial t} = \frac{\partial}{\partial x}\left[\frac{\kappa x}{m\gamma} p(x,t)\right] + \frac{k_B T}{m\gamma} \frac{\partial^2}{\partial x^2} p(x,t) \tag{2.119}$$

identified as being of the same form as the Ornstein-Uhlenbeck process. We can obtain the Green's function by Fourier transforming Eq. (2.119) and then using the method of characteristics to find a solution given by

$$p_{\text{har}}(x,t|x',t') = \sqrt{\frac{\kappa}{2\pi k_B T (1 - e^{-\frac{2\kappa}{m\gamma}(t-t')})}} \exp\left[-\frac{\kappa\left(x - x' e^{-\frac{\kappa}{m\gamma}(t-t')}\right)^2}{2k_B T (1 - e^{-\frac{2\kappa}{m\gamma}(t-t')})}\right]. \tag{2.120}$$

One can then construct a path probability utilising the Markov property by multiplying many of these solutions together. In general however, it is not possible to find a general conditional probability and so we necessarily rely on so-called short time propagators of Eq. (2.44) to construct the path probability and then formally reduce the time step in order to use the continuous limit where they hold exactly. We

consider the Fokker-Planck Eq. (2.117) in the conditional form seen in Eq. (2.29) so that we have

$$\frac{\partial p(x,t|x',t')}{\partial t} = \hat{L}(x,t)p(x,t|x',t') \tag{2.121}$$

which has general solution

$$p(x,t|x',t') = C + \int_{t'}^{t} \hat{L}(x,t_1)p(x,t_1|x',t')dt_1. \tag{2.122}$$

Considering a path probability the initial condition for $t - t' = 0$ must be a delta function therefore giving

$$p(x,t|x',t') = \delta(x-x') + \int_{t'}^{t} \hat{L}(x,t_1)p(x,t_1|x',t')dt_1. \tag{2.123}$$

This is first approximated by substituting the form of $p(x,t|x',t')$ in for the integrand producing

$$p(x,t|x',t') \simeq \delta(x-x') + \int_{t'}^{t} \hat{L}(x,t_1)\left[\delta(x-x') + \int_{t'}^{t_1} \hat{L}(x,t_2)p(x,t_2|x',t')dt_2\right]dt_1. \tag{2.124}$$

This procedure is then iterated to obtain the formal solution known as a Dyson series [46]

$$\begin{aligned} p(x,t|x',t') &= \delta(x-x') + \int_{t'}^{t} dt_1\, \hat{L}(x,t_1)\delta(x-x') + \int_{t'}^{t} dt_1 \int_{t'}^{t_1} dt_2\, \hat{L}(x,t_1)\hat{L}(x,t_2)\delta(x-x') + \dots \\ &= \left[1 + \sum_{n=1}^{\infty} \int_{t'}^{t} dt_1 \dots \int_{t'}^{t_{n-1}} dt_n\, \hat{L}(x,t_1)\dots\hat{L}(x,t_n)\right]\delta(x-x'). \end{aligned} \tag{2.125}$$

For small times $\Delta t = t - t' \ll 1$ this can be approximated as

$$p(x,t'+\Delta t|x',t') \simeq (1 + \hat{L}(x,t')\Delta t + O(\Delta t^2))\delta(x-x') \simeq e^{\Delta t \hat{L}(x,t)}\delta(x-x'). \tag{2.126}$$

By utilising a Fourier representation for the delta function

$$\delta(x-x') = \frac{1}{2\pi}\int_{-\infty}^{\infty} d\lambda\, e^{i\lambda(x-x')}, \tag{2.127}$$

we are then able to write

$$p(x,t'+\Delta t|x',t') = \frac{1}{2\pi}\int_{-\infty}^{\infty} d\lambda\, (1 + \Delta t\hat{L}(x,t'))e^{i\lambda(x-x')}. \tag{2.128}$$

We then proceed by inserting the Fokker-Planck operator

$$\hat{L}(x,t') = -\frac{\partial}{\partial x}A(x,t') + \frac{\partial^2}{\partial x^2}D(x,t'). \tag{2.129}$$

However, there is a well documented freedom associated with the evaluation point utilised in the functions A and D. Risken [46], for example, demonstrates that since they are immediately followed by the delta function one may exploit the identity

$$f(x)\delta(x-x') = f(x')\delta(x-x') \tag{2.130}$$

and thus arrive at two equivalent results depending on the evaluation point used for A and D. We generalise, in the spirit of previous sections and more transparently than in [47], by writing

$$\hat{L}(x,t') = -\frac{\partial}{\partial x}A(r(x,x'),t') + \frac{\partial^2}{\partial x^2}D(r(x,x'),t'), \tag{2.131}$$

where we define $r \equiv r(x,x') = ax + (1-a)x'$ and where $a \in [0:1]$. Substituting in with this form we find we then have to first order in Δt

$$\begin{aligned} p(x,t'+\Delta t|x',t') = \frac{1}{2\pi}\int_{-\infty}^{\infty} d\lambda\, e^{i\lambda(x-x')} \\ \times\left[1 - \frac{\partial A(r,t')}{\partial x}\Delta t - i\lambda A(r,t')\Delta t + \frac{\partial^2 D(r,t')}{\partial x^2}\Delta t + 2i\lambda\frac{\partial D(r,t')}{\partial x}\Delta t - D(r,t')\lambda^2\Delta t\right] \end{aligned} \tag{2.132}$$

which again to first order in Δt can be represented by an exponential function

$$\begin{aligned} p(x,t+\Delta t|x',t') \\ = \frac{1}{2\pi}\int_{-\infty}^{\infty} d\lambda e^{i\lambda(x-x')}e^{-\frac{\partial A(r,t')}{\partial x}\Delta t - i\lambda A(r,t')\Delta t + \frac{\partial^2 D(r,t')}{\partial x^2}\Delta t + 2i\lambda\frac{\partial D(r,t')}{\partial x}\Delta t - D(r,t')\lambda^2\Delta t}. \end{aligned} \tag{2.133}$$

This can now be readily solved by standard methods to give

$$\begin{aligned} p(x,t'+\Delta t|x',t') \\ = \sqrt{\frac{1}{4\pi D(r,t')\Delta t}}\exp\left[-\frac{\left[(x-x') - A(r,t')\Delta t + 2\frac{\partial D(r,t')}{\partial x}\Delta t\right]^2}{4D(r,t')\Delta t} - \frac{\partial A(r,t')}{\partial x}\Delta t + \frac{\partial^2 D(r,t')}{\partial x^2}\Delta t\right]. \end{aligned} \tag{2.134}$$

By application of the chain rule we then find

$$\begin{aligned} p(x,t'+\Delta t|x',t') \\ = \sqrt{\frac{1}{4\pi D(r,t')\Delta t}}\exp\left[-\frac{\left[(x-x') - A(r,t')\Delta t + 2a\frac{\partial D(r,t')}{\partial r}\Delta t\right]^2}{4D(r,t')\Delta t} - a\frac{\partial A(r,t')}{\partial r}\Delta t + a^2\frac{\partial^2 D(r,t')}{\partial r^2}\Delta t\right] \end{aligned} \tag{2.135}$$

which acts as a short time propagator for a general Fokker-Planck equation and is a conditional probability density.

One can then consider the continuum limit by repeatedly applying the Chapman-Kolmogorov equation, integrating over a product of such quantities, in order to construct the conditional probability valid for any time τ as

$$p(x(\tau),\tau|x(0),0) = \lim_{\Delta t\to 0,\, n\to\infty}\int dx_1 \ldots \int dx_{n-1}\prod_{i=0}^{n-1}\sqrt{\frac{1}{4\pi D(r_i,t_i)\Delta t}} \tag{2.136}$$

$$\times\exp\left[-\frac{\left[(x_{i+1}-x_i) - A(r_i,t_i)\Delta t + 2a\frac{\partial D(r_i,t_i)}{\partial r_i}\Delta t\right]^2}{4D(r_i,t_i)\Delta t} - a\frac{\partial A(r_i,t_i)}{\partial r_i}\Delta t + a^2\frac{\partial^2 D(r_i,t_i)}{\partial r_i^2}\Delta t\right]. \tag{2.137}$$

This can then in turn be written in terms of a functional integral in a form first used by Onsager and Machlup [48], where one generally assigns a weight to each path, $\exp\left[\int_0^\tau dt\, L(x,\dot{x})\right]$, such that one writes

$$p(x(\tau)|x(0)) = \lim_{\Delta t\to 0,\, n\to\infty} \int dx_1 \ldots \int dx_{n-1} \prod_{i=0}^{n-1} \sqrt{\frac{1}{4\pi D(r_i,t_i)\Delta t}} \tag{2.138}$$

$$\times \exp\left[-\frac{\left[(x_{i+1}-x_i) - A(r_i,t_i)\Delta t + 2a\frac{\partial D(r_i,t_i)}{\partial r_i}\Delta t\right]^2}{4D(r_i,t_i)\Delta t} - a\frac{\partial A(r_i,t_i)}{\partial r_i}\Delta t + a^2\frac{\partial^2 D(r_i,t_i)}{\partial r_i^2}\Delta t \right]$$

$$= \int [\mathcal{D}x]\mathcal{P}[\vec{x}|x(0)]$$

$$= \int [\mathcal{D}x] \exp\left[\int_0^\tau dt\, L(x,\dot{x})\right]. \tag{2.139}$$

As indicated, $\mathcal{P}[\vec{x}|x(0)]$, which we call the path probability functional, can be written explicitly in a functional form by writing it in terms of an exponentiated integral of a quantity, L, known as the Onsager Machlup function, stochastic Lagrangian or rate function. We also mention the notation $[\mathcal{D}x] = \lim \prod_i dx_i (4\pi D(r_i,t_i)\Delta t)^{-1/2}$ which forms a measure for the functional integral and may be thought of as being defined as the limit of a sequence of many regular integrals over the time interval formed in a procedure called 'time-slicing'. Here the limit indicates $\Delta t \to 0$ as the number of integrals over position variables x_i approaches infinity. This approach is used when performing path integral averages, such as when finding expectation values, in addition to integrating over the initial and final positions and including an initial probability density function. These path integrals would then be of the form

$$\begin{aligned}\int dx(\tau) \int dx(0)\, p(x(0))p(x(\tau)|x(0)) &= \int dx(\tau) \int dx(0) \int [\mathcal{D}x]\, p(x(0))\mathcal{P}[\vec{x}|x(0)] \\ &= \int dx(\tau) \int dx(0) \int [\mathcal{D}x]\, \mathcal{P}[\vec{x}] \\ &= \int d\vec{x}\, \mathcal{P}[\vec{x}]\end{aligned} \tag{2.140}$$

where we have introduced the notation of a total path integral $d\vec{x}$ symbolising integration over the end points and path probability functional and where we denote $\mathcal{P}[\vec{x}] = p(x(0))\mathcal{P}[\vec{x}|x(0)]$ the total path probability functional.

We note however, that we shall avoid using this path integral formalism when explicitly considering the individual probabilistic behaviour of certain trajectories for the sake of clarity. This is because the additional terms in a in Eq. (2.135) can appear to lead to spurious terms in the continuous limit because, in this limit, there is no explicit distinction between r and x. As such one strictly needs to include a discretisation prescription not explicit in its form which can be easy to deal with improperly. This ambiguity arises since the functional, when written in terms of the Onsager Machlup function L, can be seen to depend on the undefined $\dot{x}$. As such we consider the path probability functionals as being constructed from the short time propagators. These may then be generalised to multidimensional correlated process in N dimensions, such that we have $\boldsymbol{x} = (x_1, x_2, \ldots x_N)$ and $\boldsymbol{r} = (r_1, r_2, \ldots r_N)$, where the Fokker-Planck operator is of the form

$$\hat{L}(\boldsymbol{x},t') = -\sum_i \frac{\partial}{\partial x_i} A_i(\boldsymbol{x},t') + \sum_{i,j} \frac{\partial^2}{\partial x_i \partial x_j} D_{ij}(\boldsymbol{x},t'). \tag{2.141}$$

Specifying an initial condition $\delta(\boldsymbol{x}-\boldsymbol{x}') = \prod_i \delta(x_i - x_i')$ the short time propagator becomes

$$
\begin{aligned}
p(\boldsymbol{x}, t' + \Delta t | \boldsymbol{x}', t') &= \sqrt{\frac{1}{(4\pi\Delta t)^N \mathrm{Det}\left[\mathbf{D}(\boldsymbol{r}, t')\right]}} \\
&\times \exp\left[\sum_{i,j} -\frac{1}{4\Delta t}\left((x_i - x_i') - A_i(\boldsymbol{r}, t')\Delta t + 2a\sum_m \frac{\partial D_{im}(\boldsymbol{r}, t')}{\partial r_m}\Delta t\right) D_{ij}^{-1}(\boldsymbol{r}, t')\right. \\
&\quad \times\left((x_j - x_j') - A_j(\boldsymbol{r}, t')\Delta t + 2a\sum_n \frac{\partial D_{jn}(\boldsymbol{r}, t')}{\partial r_n}\Delta t\right) \\
&\quad \left. - \sum_i a\frac{\partial A_i(\boldsymbol{r}, t')}{\partial r_i}\Delta t + \sum_{i,j} a^2 \frac{\partial^2 D_{ij}(\boldsymbol{r}, t')}{\partial r_i \partial r_j}\Delta t\right]
\end{aligned}
\tag{2.142}
$$

where D_{ij} are the elements of the matrix $\mathbf{D}$ with inverse $\mathbf{D}^{-1}$ with elements D_{ij}^{-1} satisfying $\mathbf{D}\mathbf{D}^{-1} = \mathbf{I}$.

Chapter 3

Stochastic Thermodynamics and Fluctuation Theorems

We now provide a short review of the relevant literature on fluctuation theorems for stochastic dynamics necessary for appreciation of the original contributions offered in the subsequent chapters and understanding of their place within the literature. Owing to the extensive body of work on fluctuation theorems, rather than an historical account, a general formulation of the common underlying concepts, along with some minor generalisations and appropriate notation, are offered which are then given context by application to the advances we deem pertinent which have occurred over the past 20 years. We start by consolidating a consistent notation and identifying a quantity which possesses the basic symmetry required for the relations in the literature. We then provide a description of how such a quantity can be understood as an entropy production in general terms and how this can be interpreted as a physical quantity for a commonly used paradigmatic model. Application of the general fluctuation symmetry is then used in the context of this model to illustrate some well known results, such as the Crooks relation and Jarzynski equality, before an account of a thermodynamically revealing division of the total entropy production is given which we aim to generalise by considering time reversal more broadly, for which the motivating definition in the literature is also given.

3.1 A General Fluctuation Theorem

Fluctuation theorems that arise in stochastic dynamics are, in short, relations that describe symmetries in the averages and distributions of observed functionals of the stochastic paths or trajectories that the stochastic dynamics themselves generate. These symmetries arise specifically because one can define such a functional based on the probabilistic nature of the dynamics themselves. In a physical setting the existence of such a functional is sometimes referred to as micro-reversibility and is often cited as the explanation for the relations. In such a setting the physical significance of these functionals must of course be justified and naturally depends on how one comes to define them. However, the symmetries themselves are more general and can be defined mathematically without such considerations. As such we derive the key fluctuation symmetries in general terms in order for an appreciation of their common basis which we intend to use as a recipe for the construction of specific relations that one may subsequently consider physically relevant.

The starting point for all the relations we consider is to consider a quantity which is a functional of

the path $\vec{\boldsymbol{x}} \equiv \boldsymbol{x}(t)$ produced by some stochastic dynamics where $0 \leq t \leq \tau$. Proceeding using notation for a continuous probability space, but noting the proofs for discrete spaces are entirely analogous using the relevant path probability densities, this path then has a corresponding total path probability functional written $\mathcal{P}^{\mathrm{F}}[\vec{\boldsymbol{x}}]_{p_0^{\mathrm{F}}}$. Importantly, the superscript provides a label for the nature and time dependence of the dynamics where F here denotes 'forward' taken as a synonym for 'normal' and the subscript denotes a given, arbitrary, starting distribution $p_0^{\mathrm{F}}(\boldsymbol{x}(0))$. We then write this functional as follows

$$\mathcal{A}[\vec{\boldsymbol{x}}] = \ln\left[\frac{\mathcal{P}^{\mathrm{F}}[\vec{\boldsymbol{x}}]_{p_0^{\mathrm{F}}}}{\mathcal{P}^{*}[\vec{\boldsymbol{x}}^{*}]_{p_0^{*}}}\right]. \tag{3.1}$$

In similar notation, here $\mathcal{P}^{*}[\vec{\boldsymbol{x}}^{*}]_{p_0^{*}}$ is the total path probability functional of some path $\vec{\boldsymbol{x}}^{*}$ under some other dynamics with nature and time dependence denoted $*$ and subject to another arbitrary starting distribution $p_0^{*}(\boldsymbol{x}^{*}(0))$. Specifically, we imagine that the path $\vec{\boldsymbol{x}}^{*}$ is obtained by some transformative procedure upon $\vec{\boldsymbol{x}}$. For clarity we point out that we can write the total path probabilities in the form $\mathcal{P}^{\mathrm{F}}[\vec{\boldsymbol{x}}]_{p_0^{\mathrm{F}}} = p^{\mathrm{F}}(\boldsymbol{x}(0))\mathcal{P}^{\mathrm{F}}[\vec{\boldsymbol{x}}|\boldsymbol{x}(0)]$ so that an alternative form for the functional in terms of conditional and initial probability densities is

$$\mathcal{A}[\vec{\boldsymbol{x}}] = \ln\left[\frac{p^{\mathrm{F}}(\boldsymbol{x}(0))}{p^{*}(\boldsymbol{x}^{*}(0))}\right] + \ln\left[\frac{\mathcal{P}^{\mathrm{F}}[\vec{\boldsymbol{x}}|\boldsymbol{x}(0)]}{\mathcal{P}^{*}[\vec{\boldsymbol{x}}^{*}|\boldsymbol{x}^{*}(0)]}\right]. \tag{3.2}$$

As a quantity of possible interest, we proceed by considering the probability density function describing the distribution of observing such a quantity by constructing the path integral,

$$p^{\mathrm{F}}(\mathcal{A}[\vec{\boldsymbol{x}}] = A)_{p_0^{\mathrm{F}}} = \int d\vec{\boldsymbol{x}}\ \mathcal{P}^{\mathrm{F}}[\vec{\boldsymbol{x}}]_{p_0^{\mathrm{F}}}\delta(A - \mathcal{A}[\vec{\boldsymbol{x}}]), \tag{3.3}$$

noting that it is explicitly an average over all possible realisations of $\vec{\boldsymbol{x}}$ under dynamics denoted F conditioned on the specified, but arbitrary initial distribution $p^{\mathrm{F}}(\boldsymbol{x}(0))$. Following Harris et al. [25] we continue by considering a similar functional of the alternative path $\vec{\boldsymbol{x}}^{*}$ chosen specifically to satisfy

$$\hat{\mathcal{A}}[\vec{\boldsymbol{x}}^{*}] = \ln\left[\frac{\mathcal{P}^{*}[\vec{\boldsymbol{x}}^{*}]_{p_0^{*}}}{\mathcal{P}^{\mathrm{F}}[\vec{\boldsymbol{x}}]_{p_0^{\mathrm{F}}}}\right] = -\mathcal{A}[\vec{\boldsymbol{x}}]. \tag{3.4}$$

In a similar manner we can then construct the probability density of observing given values of this functional, under dynamics designated $*$ conditioned on the arbitrary distribution $p^{*}(\boldsymbol{x}^{*}(0))$, by writing

$$p^{*}(\hat{\mathcal{A}}[\vec{\boldsymbol{x}}^{*}] = A)_{p_0^{*}} = \int d\vec{\boldsymbol{x}}^{*}\ \mathcal{P}^{*}[\vec{\boldsymbol{x}}^{*}]_{p_0^{*}}\delta(A - \hat{\mathcal{A}}[\vec{\boldsymbol{x}}^{*}]). \tag{3.5}$$

Now, because of the definitions of the two functionals we are able to relate this distribution to that of the distribution of $\mathcal{A}[\vec{\boldsymbol{x}}]$ under the dynamics which produced $\vec{\boldsymbol{x}}$. To do so we consider the value the distribution takes for $\hat{\mathcal{A}}[\vec{\boldsymbol{x}}^{*}] = -A$. Such a consideration gives

$$p^{*}(\hat{\mathcal{A}}[\vec{\boldsymbol{x}}^{*}] = -A)_{p_0^{*}} = \int d\vec{\boldsymbol{x}}^{*}\ \mathcal{P}^{*}[\vec{\boldsymbol{x}}^{*}]_{p_0^{*}}\delta(A + \hat{\mathcal{A}}[\vec{\boldsymbol{x}}^{*}]). \tag{3.6}$$

Now by writing

$$\mathcal{P}^{*}[\vec{\boldsymbol{x}}^{*}]_{p_0^{*}} = \mathcal{P}^{\mathrm{F}}[\vec{\boldsymbol{x}}]_{p_0^{\mathrm{F}}}e^{-\mathcal{A}[\vec{\boldsymbol{x}}]}, \tag{3.7}$$

identifying $\hat{\mathcal{A}}[\vec{\boldsymbol{x}}^{*}] = -\mathcal{A}[\vec{\boldsymbol{x}}]$ and crucially assuming that the Jacobian between path integral measure $d\vec{\boldsymbol{x}}^{*}$

and $d\vec{\boldsymbol{x}}$ is unity we find

$$\begin{aligned} p^*(\hat{\mathcal{A}}[\vec{\boldsymbol{x}}^*] = -A)_{p_0^*} &= \int d\vec{\boldsymbol{x}}\, \mathcal{P}^{\mathrm{F}}[\vec{\boldsymbol{x}}]_{p_0^{\mathrm{F}}} e^{-\mathcal{A}[\vec{\boldsymbol{x}}]} \delta(A - \mathcal{A}[\vec{\boldsymbol{x}}]) \\ &= e^{-A} \int d\vec{\boldsymbol{x}}\, \mathcal{P}^{\mathrm{F}}[\vec{\boldsymbol{x}}]_{p_0^{\mathrm{F}}} \delta(A - \mathcal{A}[\vec{\boldsymbol{x}}]) \\ &= e^{-A}\, p^{\mathrm{F}}(\mathcal{A}[\vec{\boldsymbol{x}}] = A)_{p_0^{\mathrm{F}}}. \end{aligned} \tag{3.8}$$

Such a procedure then yields [25]

The Transient Fluctuation Theorem:

$$p^*(\hat{\mathcal{A}}[\vec{\boldsymbol{x}}^*] = -A)_{p_0^*} = e^{-A}\, p^{\mathrm{F}}(\mathcal{A}[\vec{\boldsymbol{x}}] = A)_{p_0^{\mathrm{F}}}. \tag{3.9}$$

By integrating over both sides we then obtain an [12, 49]

Integral Fluctuation Theorem (IFT):

$$1 = \langle e^{-\mathcal{A}[\vec{\boldsymbol{x}}]} \rangle^{\mathrm{F}}_{p_0^{\mathrm{F}}} \tag{3.10}$$

where the angled brackets with superscript denote the path average with respect to the forward dynamics, $\mathcal{P}^{\mathrm{F}}$, time dependence and initial distribution $p^{\mathrm{F}}(\boldsymbol{x}(0))$. We take this opportunity to point out that $\mathcal{A}[\vec{\boldsymbol{x}}]$ is a path dependent quantity and so averages and distributions of such quantities are necessarily formed from path integrals which are always given with the appropriate path and initial distribution information, indicated by superscripts and subscripts, and are to be distinguished from one time averages which have the same notation without such additional features. The sole exception to this rule is for the solution of the appropriate forward equation, $p^{\mathrm{F}}(\boldsymbol{x}(t), t)$, for which we deem the initial distribution implicit. It is important to note that in addition to the expected normalisation condition for both $\mathcal{P}^*[\vec{\boldsymbol{x}}]$ and $\mathcal{P}^{\mathrm{F}}[\vec{\boldsymbol{x}}]$ in the derivation a number of other conditions are required in the conversion of the path integral over $\vec{\boldsymbol{x}}^*$ into one over $\vec{\boldsymbol{x}}$. Firstly, such a conversion implicitly assumes that the integral bounds cover the same region of trajectory space. That is to say all paths $\vec{\boldsymbol{x}}$ should be contained within the bounds of the integral constructed over paths $\vec{\boldsymbol{x}}^*$. In general this can be achieved if for all paths $\vec{\boldsymbol{x}}$ where $\mathcal{P}^{\mathrm{F}}[\vec{\boldsymbol{x}}] = 0$, then the corresponding transformed path $\vec{\boldsymbol{x}}^*$ has a total path probability functional under the alternative dynamics $\mathcal{P}^*[\vec{\boldsymbol{x}}^*] = 0$ and vice versa. As such the condition depends on the relationship between paths $\vec{\boldsymbol{x}}$ and $\vec{\boldsymbol{x}}^*$ and the relationship between the dynamics $\mathcal{P}^{\mathrm{F}}$ and $\mathcal{P}^*$ which generate them. Consequently these fluctuation relations can exist between non-common sets of paths $\vec{\boldsymbol{x}}$ and $\vec{\boldsymbol{x}}^*$ as long as the dynamics $\mathcal{P}^*$ is suitably chosen. In other words, paths $\vec{\boldsymbol{x}}^*$ possible under $\mathcal{P}^*$ need not be possible under $\mathcal{P}^{\mathrm{F}}$, which whilst perhaps not assumed in, might be inferred from, the literature which tends to consider one particular choice of $\vec{\boldsymbol{x}}^*$. This can be seen as a generalised stochastic version of the ergodic consistency requirement[1]. If the paths are common to both sets of dynamics it then reduces to the more familiar requirement that the initial distributions $p^{\mathrm{F}}(\boldsymbol{x}(0))$ and $p^*(\boldsymbol{x}^*(0))$ are nowhere zero since we assume any stochastic dynamics we consider to be ergodic. Secondly, we note that the assumption that the transformation of the integral over all paths $\vec{\boldsymbol{x}}^*$ to $\vec{\boldsymbol{x}}$ left its form otherwise unchanged (such that the Jacobian is unity) is crucial for such relations and can be achieved through appropriate choices

[1]This in turn can be viewed, in a measure-theoretic approach, as the condition of equivalence of measure between the forward dynamics or measure $P^{\mathrm{F}}[\vec{\boldsymbol{x}}]$ and the measure of the alternative dynamics $P^*[\vec{\boldsymbol{x}}^*]$ utilising the alternative path. Indeed such a condition also implies that $d\vec{\boldsymbol{x}} = d\vec{\boldsymbol{x}}^*$ allowing the discussion of the functional $\mathcal{A}[\vec{\boldsymbol{x}}]$ as being formed from path probability functionals in this manner.

of $\vec{\boldsymbol{x}}^*$ and can be generally assured by requiring that the transform used to generate it from $\vec{\boldsymbol{x}}$ is involutive.

The two relations derived above are the most common relations found in the literature, but one should realise that a great many extended, but ultimately homologous, symmetries can then be invoked using the form of Eq. (3.9) if one exchanges the delta function in Eqs. (3.3) and (3.8) with an arbitrary function, g, of some other functional $\mathcal{B}[\vec{\boldsymbol{x}}]$ of the same path, with sole additional assumption that one can simultaneously define $\mathcal{C}[\vec{\boldsymbol{x}}^*] = \mathcal{B}[\vec{\boldsymbol{x}}]$ [9, 25]. By means of the same simple substitution

$$\mathcal{P}^*[\vec{\boldsymbol{x}}^*]_{p_0^*} = \mathcal{P}^{\mathrm{F}}[\vec{\boldsymbol{x}}]_{p_0^{\mathrm{F}}} e^{-\mathcal{A}[\vec{\boldsymbol{x}}]}, \tag{3.11}$$

and the definition of $\mathcal{C}[\vec{\boldsymbol{x}}^*]$ one may write

$$\langle g(\mathcal{C}[\vec{\boldsymbol{x}}^*])\rangle^*_{p_0^*} = \langle g(\mathcal{B}[\vec{\boldsymbol{x}}]) e^{-\mathcal{A}[\vec{\boldsymbol{x}}]}\rangle^{\mathrm{F}}_{p_0^{\mathrm{F}}} \tag{3.12}$$

noting that since one also has $\mathcal{P}^*[\vec{\boldsymbol{x}}^*]_{p_0^*} = \mathcal{P}^{\mathrm{F}}[\vec{\boldsymbol{x}}]_{p_0^{\mathrm{F}}} e^{\hat{\mathcal{A}}[\vec{\boldsymbol{x}}^*]}$ this is equivalent to

$$\langle g(\mathcal{C}[\vec{\boldsymbol{x}}^*]) e^{-\hat{\mathcal{A}}[\vec{\boldsymbol{x}}^*]}\rangle^*_{p_0^*} = \langle g(\mathcal{B}[\vec{\boldsymbol{x}}])\rangle^{\mathrm{F}}_{p_0^{\mathrm{F}}} \tag{3.13}$$

relating, in principle, arbitrary functions of related functionals under dynamics matched through $\mathcal{A}[\vec{\boldsymbol{x}}]$. As such, almost all further fluctuation symmetries can be considered to be special cases of such a result by utilising a suitable choice of g and $\mathcal{B}[\vec{\boldsymbol{x}}]$ [9, 25], though we note that mathematically more general and rigorous contributions have been offered [50–56]. As an example, we may choose g to be a simple function, such that we have perhaps $g(\mathcal{B}[\vec{\boldsymbol{x}}]) = g(\boldsymbol{x}(t))$. The corresponding 'functional' for the alternative dynamics is $g(\mathcal{C}[\vec{\boldsymbol{x}}^*]) = g(\mathcal{B}[\vec{\boldsymbol{x}}]) = g(\hat{T}(\boldsymbol{x}^*(t)) = \boldsymbol{x}(t))$ where $\hat{T}$ is the relevant transform of $\boldsymbol{x}^*(t)$ that gives $\boldsymbol{x}(t)$. Doing so gives a relation between one time averages at specific points in the evolution of the relevant dynamics and is something of a more general relation of the type that reduces to a so-called Kawasaki response relation, conditioned upon certain choices of $*$ etc.

$$\langle g(\hat{T}(\boldsymbol{x}^*(t)) = \boldsymbol{x}(t))\rangle^*_{p_0^*} = \langle g(\boldsymbol{x}(t)) e^{-\mathcal{A}[\vec{\boldsymbol{x}}]}\rangle^{\mathrm{F}}_{p_0^{\mathrm{F}}}. \tag{3.14}$$

We note that we have restricted ourselves to leaving the functional $\mathcal{A}[\vec{\boldsymbol{x}}]$ as it is, but point out even more relations can be derived if one subsequently divides up the functional based on creative choices of the boundary distributions [57]. Largely, however, the most pertinent results in the literature arise from the choice $\mathcal{B}[\vec{\boldsymbol{x}}] = \mathcal{A}[\vec{\boldsymbol{x}}]$ with appropriate $\mathcal{C}[\vec{\boldsymbol{x}}^*] = -\hat{\mathcal{A}}[\vec{\boldsymbol{x}}^*]$ which leads to

$$\langle g(-\hat{\mathcal{A}}[\vec{\boldsymbol{x}}^*]) e^{-\hat{\mathcal{A}}[\vec{\boldsymbol{x}}^*]}\rangle^*_{p_0^*} = \langle g(\mathcal{A}[\vec{\boldsymbol{x}}])\rangle^{\mathrm{F}}_{p_0^{\mathrm{F}}}. \tag{3.15}$$

Given such a form, a particularly relevant choice of function g, considering the origins and implications of the fluctuation theorem is the choice $g(A) = e^{-\lambda A}$ which yields the symmetry

$$\langle e^{-(1-\lambda)\hat{\mathcal{A}}[\vec{\boldsymbol{x}}^*]}\rangle^*_{p_0^*} = \langle e^{-\lambda\mathcal{A}[\vec{\boldsymbol{x}}]}\rangle^{\mathrm{F}}_{p_0^{\mathrm{F}}} \tag{3.16}$$

with implications that will be commented on subsequently.

3.2 Time Reversal and the Foundations of Entropy Production

Let us imagine now a path integral average over such a functional $\mathcal{A}[\vec{\boldsymbol{x}}]$ conditioned on p_0^{F}. This will have the form

$$\langle \mathcal{A}[\vec{\boldsymbol{x}}] \rangle_{p_0^{\mathrm{F}}}^{\mathrm{F}} = \int d\vec{\boldsymbol{x}}\, \mathcal{P}^{\mathrm{F}}[\vec{\boldsymbol{x}}]_{p_0^{\mathrm{F}}} \ln \frac{\mathcal{P}^{\mathrm{F}}[\vec{\boldsymbol{x}}]_{p_0^{\mathrm{F}}}}{\mathcal{P}^{*}[\vec{\boldsymbol{x}}^{*}]_{p_0^{*}}}. \tag{3.17}$$

This is recognisable as a Kullback-Leibler divergence or relative entropy, albeit between path probability densities rather than distributions, and so must therefore be positive. Further, let us imagine that the initial distribution $p^{\mathrm{F}}(\boldsymbol{x}(0)) = p_0^{\mathrm{F}}$ is not arbitrary, but in fact chosen so as to be equal to the initial distribution of probability over all states at the start of some physical process such that we consider it the initial condition for the forward equation that describes the process with solution $p^{\mathrm{F}}(\boldsymbol{x}(t), t) \equiv p_t^{\mathrm{F}}$. In these terms, such an 'entropy' has been considered, even for quite some time now [58–60], as the intrinsic positive entropy production, introduced originally for stationary non-equilibrium systems arising from a time homogeneous birth-death master equation, given, importantly, a particular choice of $\mathcal{P}^{*}[\vec{\boldsymbol{x}}^{*}]_{p_0^{*}}$. Taking this 'entropy production' as a starting point, the choice of $\mathcal{P}^{*}[\vec{\boldsymbol{x}}^{*}]_{p_0^{*}}$ will come to define a simple, yet important concept in the construction, generalisation and interpretation of the quantity $\mathcal{A}[\vec{\boldsymbol{x}}]$.

First, however, in order to characterise more general examples of non-equilibrium behaviour, particularly that of driving, we should however, go a step further in the development by the consideration of time inhomogeneous processes. This is generally achieved by allowing the probabilistic behaviour of the system to vary in time and might be realised physically by the alteration of some physical parameter influencing the system such as, for example, the nature of a Hamiltonian or perhaps the temperature or chemical potential of a heat or particle bath. Frequently this time dependence is assumed to occur through a 'switching protocol' or simply a 'protocol' often denoted $\lambda(t)$. This protocol then implicitly becomes part of the definition of the forward path probability functional such that $\mathcal{P}^{\mathrm{F}}[\vec{\boldsymbol{x}}]$ is taken to be the total probability functional for the path $\vec{\boldsymbol{x}}$ under forward dynamics *and* forward protocol which we consequently denote $\lambda^{\mathrm{F}}(t)$.

Given such a time dependence in the dynamics, the crucial choice in the construction of $\mathcal{A}[\vec{\boldsymbol{x}}]$ that determines the path $\vec{\boldsymbol{x}}^{*}$, dynamics $\mathcal{P}^{*}$ and initial condition p_0^{*} is that which constitutes *time reversal*. The concept of time reversal is crucial in the study of fluctuation theorems as it identifies a connection between their subject, entropy production, and the physical manifestation of entropy production, irreversibility. We do this by considering that the irreversibility of a process is characterised by the relative likelihood of observing the original system behaviour compared to the likelihood of observing the system behaviour that would precisely reverse or 'undo' the previous motion. For example, we would consider some driving process reversible if, for all paths, the probability of the forward behaviour under the forward driving was identical to the probability of the reverse behaviour, starting in conditions identical to those at the end of the forward process, under the reverse driving. Such a statement then reveals the choice of $\mathcal{P}^{*}[\vec{\boldsymbol{x}}^{*}]_{p_0^{*}}$ which would allow us to discern this information from $\mathcal{A}[\vec{\boldsymbol{x}}]$. By considering the behaviour of the system to be the path that the system takes, $\vec{\boldsymbol{x}}$, then the reverse behaviour is this same path 'played backwards' or rather in reverse sequence. This allows us to define the choice $\vec{\boldsymbol{x}}^{*}$ as the reversed path $\vec{\boldsymbol{x}}^{*} = \vec{\boldsymbol{x}}^{\mathrm{R}} = \boldsymbol{x}(\tau - t)$. Similarly, if the protocol $\lambda^{\mathrm{F}}(t)$ defines the forward driving then the reverse driving is driving that arises from the reversed sequence of the protocol $\lambda^{\mathrm{R}}(t) = \lambda^{\mathrm{F}}(\tau - t)$. Finally, if we ask that the irreversibility be measured from the system conditions at the end of the forward process then we can define the initial distribution $p^{*}(\boldsymbol{x}(0))$ to be the probability distribution at the end of the forward process such that we write $p_0^{*} = p_\tau^{\mathrm{F}}$. Explicitly, $\mathcal{P}^{*}[\vec{\boldsymbol{x}}^{*}]_{p_0^{*}}$ becomes the total path probability functional

associated with the path $\vec{\boldsymbol{x}}^* = \vec{\boldsymbol{x}}^{\mathrm{R}}$ under the original dynamics subject to time dependence characterised by the *reversed protocol* $\lambda^{\mathrm{R}}(t) = \lambda^{\mathrm{F}}(\tau - t)$ leading us to write $\mathcal{P}^*[\vec{\boldsymbol{x}}^*|\boldsymbol{x}^*(0)] = \mathcal{P}^{\mathrm{R}}[\vec{\boldsymbol{x}}^{\mathrm{R}}|\boldsymbol{x}(\tau)]$ with choice of initial condition $p_0^* = p_\tau^{\mathrm{F}}$. Given these choices, Eq. (3.17) becomes

$$\langle \mathcal{A}[\vec{\boldsymbol{x}}] \rangle_{p_0^{\mathrm{F}}}^{\mathrm{F}} = \int d\vec{\boldsymbol{x}}\, \mathcal{P}^{\mathrm{F}}[\vec{\boldsymbol{x}}]_{p_0^{\mathrm{F}}} \ln \frac{\mathcal{P}^{\mathrm{F}}[\vec{\boldsymbol{x}}]_{p_0^{\mathrm{F}}}}{\mathcal{P}^{\mathrm{R}}[\vec{\boldsymbol{x}}^{\mathrm{R}}]_{p_\tau^{\mathrm{R}}}} \tag{3.18}$$

and we begin to see the physical meaning that it imparts. In this form is becomes an average, by means of a path integral, of the likelihood of paths, generated by evolving forwards in time, compared to that of the specific, related paths which would fully reverse the observed forward behaviour of the system, where averaging occurs over the ensemble of *forward* paths.

Further, when considering path probabilities that arise from a master equation for a set of discrete states one can show [12, 25, 36, 58] that the mean rate of change of $\mathcal{A}[\vec{\boldsymbol{x}}]$, which we state to be the mean rate of non-negative, irreversible or ensemble entropy production of the universe or 'total entropy production' $\dot{S}_{\mathrm{tot}}$ with suitable consideration of units, is of the form

$$k_B \frac{d\langle \mathcal{A}[\vec{\boldsymbol{x}}] \rangle_{P_0^{\mathrm{F}}}^{\mathrm{F}}}{dt} = \dot{S}_{\mathrm{tot}} = k_B \sum_{\boldsymbol{x},\boldsymbol{x}'} P_0^{\mathrm{F}}(\boldsymbol{x}) T(\boldsymbol{x}'|\boldsymbol{x}, \lambda^{\mathrm{F}}(t)) \ln \frac{P_0^{\mathrm{F}}(\boldsymbol{x}) T(\boldsymbol{x}'|\boldsymbol{x}, \lambda^{\mathrm{F}}(t))}{P_0^{\mathrm{F}}(\boldsymbol{x}') T(\boldsymbol{x}|\boldsymbol{x}', \lambda^{\mathrm{F}}(t))} \tag{3.19}$$

where $T(\boldsymbol{x}'|\boldsymbol{x}, \lambda^{\mathrm{F}}(t))$ is the transition rate of transitions from state $\boldsymbol{x}'$ to $\boldsymbol{x}$ at time t characterised by protocol $\lambda^{\mathrm{F}}(t)$. Such a form has provided the starting point for many subsequent modern treatments [12, 25, 61] emphasising a profound connection between such observed irreversibility that arise from the dynamics of a particular system in this way and the entropy production we expect from such behaviour. Indeed, the use of Kullback-Leibler divergences and other more sophisticated functionals of path probabilities borrowed from information theory have been explicitly used in attempts to 'quantify time's arrow' [62–66].

More recently, encouraged by progress in work on thermostatted deterministic systems and perhaps by improved technological techniques to probe and influence dynamics on smaller and smaller scales the concept of viewing usual thermodynamic quantities such as work, heat and indeed entropy as taking distinct, single values has given way to a more general consideration as uncertain, path dependent, quantities described by probability distributions. Unifying descriptions of these quantities, particularly when considering entropy production, again have at their core the principle of time reversal [49, 67] and put most simply one considers the physical interpretation of the specific value of the functional $\mathcal{A}[\vec{\boldsymbol{x}}]$ due to a given realisation $\vec{\boldsymbol{x}}$ rather than the average value over all possible realisations as considered in [58]. This reasoning implies that if the intrinsic ensemble entropy production of the universe, the time integral of $\dot{S}_{\mathrm{tot}}$, is a mean value, by means of a path integral, of the functional $k_B\mathcal{A}[\vec{\boldsymbol{x}}]$ then we can consider a microscopic *path dependent* entropy production which takes the precise value of the functional such that $\Delta\mathcal{S}_{\mathrm{tot}} = k_B\mathcal{A}[\vec{\boldsymbol{x}}]$ which can fluctuate around that mean. One can see that the consideration of such a quantity would then naturally lead to the fluctuation relations detailed above, but one might ask whether such a deconstruction of an ensemble quantity has any validity or can be viewed as anything physical at all.

In order to meaningfully answer this question we should ask what in fact the functional in question represents on the level of a single trajectory. Writing the functional as the contribution of two terms

based on initial distributions and path probability functionals as follows

$$\mathcal{A}[\vec{\boldsymbol{x}}] = \ln\left[\frac{\mathcal{P}^{\mathrm{F}}[\vec{\boldsymbol{x}}]_{p_0^{\mathrm{F}}}}{\mathcal{P}^{*}[\vec{\boldsymbol{x}}^{*}]_{p_0^{*}}}\right] = \ln\left[\frac{p^{\mathrm{F}}(\boldsymbol{x}(0),0)}{p^{\mathrm{F}}(\boldsymbol{x}(\tau),\tau)}\right] + \ln\left[\frac{\mathcal{P}^{\mathrm{F}}[\vec{\boldsymbol{x}}|\boldsymbol{x}(0)]}{\mathcal{P}^{\mathrm{R}}[\vec{\boldsymbol{x}}^{\mathrm{R}}|\boldsymbol{x}(\tau)]}\right] \tag{3.20}$$

we may ask what such contributions each physically represent. The second, formed of path probability functionals providing the uncertainty in the path evolution is reasoned to be, quite generally, a path dependent entropy change of the environment based on an assumption of local detailed balance [12, 25, 49] such that we write

$$\Delta\mathcal{S}_{\mathrm{med}} = k_B \ln\left[\frac{\mathcal{P}^{\mathrm{F}}[\vec{\boldsymbol{x}}|\boldsymbol{x}(0)]}{\mathcal{P}^{\mathrm{R}}[\vec{\boldsymbol{x}}^{\mathrm{R}}|\boldsymbol{x}(\tau)]}\right]. \tag{3.21}$$

The defining feature of such a quantity is that given well defined transition rates or path probability functionals, which in turn might possibly be written in terms of defined observables, a microscopically consistent quantity can be associated unambiguously with that behaviour. That is to say (indeed for any choice of $\mathcal{P}^{*}[\vec{\boldsymbol{x}}^{*}|\boldsymbol{x}^{*}(0)]$) such a term is of a form that means it contributes based only on the system behaviour $\boldsymbol{x}(t)$ and not on some additionally specified function meaning it can in principle represent a physically relevant quantity that can be measured or counted. For example, for a large thermal environment that possesses a defined temperature, one expects its entropy change to be the heat flow scaled by its inverse temperature. This quantity, in principle is well defined even for single trajectories as the heat could be computed by energetic arguments. Because of this ability to identify such a contribution with real measurable quantities both it and its analogue in deterministic thermostatted dynamics, the phase space contraction, are also the subject of the so-called asymptotic fluctuation theorems [28, 31]. The nature of these relations are very similar to those we have outlined, but are considered in the long time limit where their validity becomes apparent if one observes that the contribution formed of initial distributions is unimportant. This however, is not a general feature of stochastic systems with unbounded state space and so is frequently observed to 'fail' [68–72]. However, our focus of interest lies in our original functional which contains an additional term consisting solely of the initial distributions used in the construction of $\mathcal{A}[\vec{\boldsymbol{x}}]$. This term, however, does not have the properties of the environmental entropy change; one cannot associate a unique contribution with some behaviour $\boldsymbol{x}(t)$, but instead one has to consider the form of the initial probability distributions which, depending on your point of view, might represent an arbitrary conditioning, a degree of confidence in a measurement or the behaviour of an ensemble. It should be noted that, subject to the ergodic constraints, any pair of initial distributions, not just those used above, ensure the fluctuation theorems previously derived pointing perhaps to the assertion that they are merely arbitrary boundaries without particular physical meaning. This point may seem especially relevant when one considers their dependence on something less concrete than just the path, unlike the environmental entropy change, providing a certain motivation in the pursuit of the asymptotic fluctuation theorems [49]. This, in general when the distributions are explicitly arbitrary, which unlike as outlined above, was the role they played in earlier work [49], is a fair assessment, however more recently, again perhaps following from increased real world experimental resolution, it has come to be accepted that there is a genuine and non-contrived physical interpretation of these terms when chosen appropriately. Such a choice is based on the notion that the coarse grained Gibbs or Shannon entropy of stochastic systems describing the intrinsic entropy of a system, which may be generalised to non-equilibrium ensembles by being made time dependent,

$$S(t) = -\int d\boldsymbol{x}(t)\; p^{\mathrm{F}}(\boldsymbol{x}(t),t)\ln p^{\mathrm{F}}(\boldsymbol{x}(t),t), \tag{3.22}$$

can be interpreted as the instantaneous average of a microscopic *path dependent* quantity $-\ln p^{\mathrm{F}}(\boldsymbol{x}(t),t)$. Consequently the change in such a fluctuating quantity over a time τ, $\ln(p^{\mathrm{F}}(\boldsymbol{x}(0),0)/p^{\mathrm{F}}(\boldsymbol{x}(\tau),\tau))$, represents a microscopic contribution to the change in the quantity of information, or entropy, required to describe the mean instantaneous probabilistic behaviour of the system. Such a concept was first hinted at in [8, 73], but introduced explicitly some time later in [12]. The boundary contributions p_0^{F} and p_0^* only take this form when we utilise the choice that corresponds to the measurement of irreversibility of the process, namely $p^*(\boldsymbol{x}^*(0)) = p^{\mathrm{F}}(\boldsymbol{x}(\tau),\tau)$. Since the change in this microscopic entropy production associated with the system is incorporated in the functional as described in equation (3.20), along with the environmental, or medium, entropy change, and that in such a model the universe consists only of the system and the medium, we can justify the earlier claim that the functional $\mathcal{A}[\vec{\boldsymbol{x}}]$ is a microscopic contribution of total entropy production of the universe with structure as below

$$\begin{aligned}\Delta\mathcal{S}_{\mathrm{tot}} &= k_B\mathcal{A}[\vec{\boldsymbol{x}}] \\ &= k_B \ln\left[\frac{\mathcal{P}^{\mathrm{F}}[\vec{\boldsymbol{x}}]_{p_0^{\mathrm{F}}}}{\mathcal{P}^{\mathrm{R}}[\vec{\boldsymbol{x}}^{\mathrm{R}}]_{p_\tau^{\mathrm{F}}}}\right] \\ &= k_B \ln\left[\frac{p^{\mathrm{F}}(\boldsymbol{x}(0),0)}{p^{\mathrm{F}}(\boldsymbol{x}(\tau),\tau)}\right] + k_B \ln\left[\frac{\mathcal{P}^{\mathrm{F}}[\vec{\boldsymbol{x}}|\boldsymbol{x}(0)]}{\mathcal{P}^{\mathrm{R}}[\vec{\boldsymbol{x}}^{\mathrm{R}}|\boldsymbol{x}(\tau)]}\right] \\ &= \Delta\mathcal{S}_{\mathrm{sys}} + \Delta\mathcal{S}_{\mathrm{med}}. \end{aligned} \tag{3.23}$$

3.2.1 The Integral Fluctuation Theorem for Total Entropy Production

As a preliminary consequence of being able to identify an entropy production in such a form we identify, perhaps, the most basic and general of the fluctuation relations since it relies only on the form of the functional $\mathcal{A}[\vec{\boldsymbol{x}}]$. By identifying the total entropy production in the integral relation of Eq. (3.10) such that $k_B\mathcal{A}[\vec{\boldsymbol{x}}] = \Delta\mathcal{S}_{\mathrm{tot}}$ we can write [12]

$$\langle \exp[-k_B^{-1}\Delta\mathcal{S}_{\mathrm{tot}}]\rangle_{p_0^{\mathrm{F}}}^{\mathrm{F}} = 1. \tag{3.24}$$

The most important consequence of such a relation is the bound it places on the mean quantity $\langle\Delta\mathcal{S}_{\mathrm{tot}}\rangle_{p_0^{\mathrm{F}}}^{\mathrm{F}}$. Because of the convexity of the exponential function it falls within the remit of Jensen's inequality, allowing us to write

$$1 \geq \exp[-k_B^{-1}\langle\Delta\mathcal{S}_{\mathrm{tot}}\rangle_{p_0^{\mathrm{F}}}^{\mathrm{F}}] \tag{3.25}$$

consequently leading us to conclude

$$\langle\Delta\mathcal{S}_{\mathrm{tot}}\rangle_{p_0^{\mathrm{F}}}^{\mathrm{F}} \geq 0. \tag{3.26}$$

This is, in essence, the statement of the second law when we explicitly consider fluctuations in the dynamics we consider. Instead of expecting the second law to be a rigorous inequality of the form $\Delta S \geq 0$, by defining a trajectory dependent entropy we cast the second law in the statistical terms under which it should be understood, allowing an entropy production to be either positive or negative, but recovering a positive expectation by means of a statistical bound. We note that this is true of any quantity obeying an integral fluctuation theorem; if we can express it as the logarithm of a ratio of path probabilities, then the physical quantity it corresponds to has a rigorous positive expectation.

3.3 Stochastic Energetics

In order to make physical sense of a stochastic entropy production we should be able to attribute thermodynamic quantities to the dynamics which describe the system we wish to model. It is most helpful if we consider a particular stochastic dynamics as a framework to illustrate their place within the fluctuation theorems. Such a paradigmatic approach to such a question was introduced by Sekimoto [34, 45, 74] entitled 'stochastic energetics', the fundamentals of which shall be paraphrased here, and the use of it within the context of entropy production and fluctuation theorems is generally called 'stochastic thermodynamics' [12, 33]. The paradigm utilised in stochastic energetics is that of a simple one dimensional over-damped Langevin equation of the form

$$\frac{dx}{dt} = \frac{1}{m\gamma}\mathcal{F}(x, \lambda^{\mathrm{F}}(t)) + \sqrt{\frac{2k_BT}{m\gamma}}\Gamma(t) \tag{3.27}$$

where we follow the literature so that the time dependence of the dynamics, often a dependence in a confining potential or applied force, is provided through the switching protocol, $\lambda^{\mathrm{F}}(t)$, that represents an external agent or protocol which can then drive the system. We note that we could write

$$\frac{dx}{dt} = -\frac{1}{m\gamma}\frac{\partial V(x, \lambda_0^{\mathrm{F}}(t))}{\partial x} + \frac{1}{m\gamma}\mathcal{F}_{nc}(x, \lambda_1^{\mathrm{F}}(t)) + \sqrt{\frac{2k_BT}{m\gamma}}\Gamma(t) \tag{3.28}$$

so that $\mathcal{F}(x, \lambda^{\mathrm{F}}(t))$ is the total force on the particle with a contribution from a potential and a separate non-conservative force each controlled by an individual protocol $\lambda_0^{\mathrm{F}}(t)$ and $\lambda_1^{\mathrm{F}}(t)$, which could, for example, be used to model a trapped particle in a steady flow. Following Sekimoto, we delay interpretation as an SDE so as to explicitly consider force terms, yet proceed with the understanding that any stochastic integrals are chosen to follow Stratonovich rules such that the normal rules of calculus apply (for details see Sect. 2.2.2). Rearranging such that terms have the unit of force we have

$$0 = -\left(-m\gamma\frac{dx}{dt} + \sqrt{2k_BTm\gamma}\Gamma(t)\right) + \frac{\partial V(x, \lambda_0^{\mathrm{F}}(t))}{\partial x} - \mathcal{F}_{nc}(x, \lambda_1^{\mathrm{F}}(t)). \tag{3.29}$$

Now by multiplying by an increment dx under Stratonovich such that

$$0 = -\left(-m\gamma\frac{dx}{dt} + \sqrt{2k_BTm\gamma}\Gamma(t)\right) \circ dx + \frac{\partial V(x, \lambda_0^{\mathrm{F}}(t))}{\partial x} \circ dx - \mathcal{F}_{nc}(x, \lambda_1^{\mathrm{F}}(t)) \circ dx \tag{3.30}$$

the individual increments of a force multiplied by a displacement can be thought of as terms in an energy balance equation. Sekimoto postulated that despite being probabilistic the force terms should always obey the law of action and reaction. Consequently the term in parentheses in Eq. (3.29), identified as the negative of the force due to the environment since it alone contains both the fluctuation and dissipative terms, must be the reaction force the particle exerts on the environment. Consequently the work done by this force (when multiplied by dx under Stratonovich) is the energy lost to the environment. It is this quantity we therefore associate with heat flow to the environment and thus wish to relate to the entropy production associated with it. To avoid confusion, we label the microscopic gain of heat in the system (or loss of heat for the bath) dQ_{sys} and the corresponding heat flow to the environment dQ such that $dQ = -dQ_{\mathrm{sys}}$, a notation that we will retain throughout. Following Stratonovich once more we have

$$dV = \frac{\partial V(x, \lambda_0^{\mathrm{F}}(t))}{\partial x} \circ dx + \frac{\partial V(x, \lambda_0^{\mathrm{F}}(t))}{\partial \lambda_0^{\mathrm{F}}(t)} \circ d\lambda_0^{\mathrm{F}}(t). \tag{3.31}$$

After substituting we then have

$$0 = -dQ_{\rm sys} + dV - \frac{\partial V(x,\lambda_0^{\rm F}(t))}{\partial \lambda_0^{\rm F}(t)} \circ d\lambda_0^{\rm F}(t) - \mathcal{F}_{nc}(x,\lambda_1^{\rm F}(t)) \circ dx. \tag{3.32}$$

This then allows us to identify the expected expression for the work done from basic mechanics

$$dW = \frac{\partial V(x,\lambda_0^{\rm F}(t))}{\partial \lambda_0^{\rm F}(t)} \circ d\lambda_0^{\rm F}(t) + \mathcal{F}_{nc}(x,\lambda_1^{\rm F}(t)) \circ dx \tag{3.33}$$

which means that following a substitution, the heat flow into the system is given by

$$dQ_{\rm sys} = \frac{\partial V(x,\lambda_0^{\rm F}(t))}{\partial x} \circ dx - \mathcal{F}_{nc}(x,\lambda_1^{\rm F}(t)) \circ dx = -\mathcal{F}(x,\lambda^{\rm F}(t)) \circ dx = \left(-m\gamma\frac{dx}{dt} + \sqrt{2k_BTm\gamma}\Gamma(t)\right) \circ dx. \tag{3.34}$$

Interpretation in this sense allows us to consider a first law equality in the form

$$dV = dQ_{\rm sys} + dW \tag{3.35}$$

or

$$dV + dQ = dW \tag{3.36}$$

where the internal energy is simply the potential such that $dU = dV$.

Note on the Stratonovich Interpretation of Work Like Terms

It is important to note here the implications of an interpretation according to Itō or Stratonovich for the work like quantities arising from spatially dependent forces multiplied by the increment dx. [2] As highlighted in previous sections, interpreting a stochastic differential equation in one sense over another will generally lead to *measurably different* behaviour in the resulting stochastic variable as can be shown by Eq. (2.88). Whilst this is not the case for simple Brownian motion the issue does arise when performing a multiplication of the form $\mathcal{F}(x) * dx$ since both integrand and integrator are now non-differentiable. Considering the simple case of a single conservative force acting upon the particle, following the interpretation due to Stratonovich produces the expression for the heat

$$dQ_{\rm sys} = \frac{\partial V(x,\lambda_0^{\rm F}(t))}{\partial x} \circ dx. \tag{3.37}$$

One might ask, however, what the form of the heat is under an Itō interpretation. We can certainly transform this quantity according to Eq. (2.93) and produce an equivalent quantity. Indeed if we can form an energy balance equation without Sekimoto's arguments that use a multiplied force, the result should be equivalent for simple Brownian motion. We can do this for the case of a single conservative force by considering the evolution of V according to Itō's lemma such that

$$dV = \frac{\partial V(x,t)}{\partial t}dt + \frac{\partial V(x,t)}{\partial x}dx + \frac{k_BT}{m\gamma}\frac{\partial^2 V(x,t)}{\partial x^2}dt. \tag{3.38}$$

[2] An example of improperly taking this into account is evidenced by the confusion and discussion in [75–79]. Aside from the improper use of mechanical work over thermodynamic work, as commented on, it is the naive interpretation of this work using Itō rules, which was not noticed, which results in an erroneous conclusion that work relations are only valid near equilibrium.

We can consistently identify the internal energy as the total derivative $dU = dV$ and the work as

$$dW = \frac{\partial V(x,\lambda_0^{\mathrm{F}}(t))}{\partial \lambda_0^{\mathrm{F}}(t)} d\lambda_0^{\mathrm{F}}(t) = \frac{\partial V(x,\lambda_0^{\mathrm{F}}(t))}{\partial \lambda_0^{\mathrm{F}}(t)} \frac{d\lambda_0^{\mathrm{F}}(t)}{dt} dt = \frac{\partial V(x,\lambda_0^{\mathrm{F}}(t))}{\partial t} dt \tag{3.39}$$

which is independent of the interpretation owing to its lack of dependence on the increment dx. As such we can construct an energy balance equation yielding

$$dQ_{\mathrm{sys}} = \frac{\partial V(x,\lambda_0^{\mathrm{F}}(t))}{\partial x} dx + \frac{k_B T}{m\gamma} \frac{\partial^2 V(x,\lambda_0^{\mathrm{F}}(t))}{\partial x^2} dt \tag{3.40}$$

in agreement with Eq. (2.93). However if we try to introduce non-conservative forces in such an approach the result loses its invariance since we must decide whether the previously unambiguous work is

$$dW = \frac{\partial V(x,\lambda_0^{\mathrm{F}}(t))}{\partial t} dt + \mathcal{F}_{nc}(x,\lambda_1^{\mathrm{F}}(t)) \circ dx \tag{3.41}$$

or

$$dW = \frac{\partial V(x,\lambda_0^{\mathrm{F}}(t))}{\partial t} dt + \mathcal{F}_{nc}(x,\lambda_1^{\mathrm{F}}(t)) dx, \tag{3.42}$$

or indeed any other interpretation of the stochastic integral. Consequently we must return to the force balance equation to assure generality. However, if we follow Sekimoto, but instead use an Itō interpretation we come across a problem even for the case of a single conservative force. Constructing the energy balance equation, but multiplying using an Itō convention we obtain

$$\begin{aligned} 0 &= -\left(-m\gamma \frac{dx}{dt} + \sqrt{2k_B T m\gamma}\Gamma(t)\right) dx + \frac{\partial V(x,\lambda_0^{\mathrm{F}}(t))}{\partial x} dx \\ &= -dQ_{\mathrm{sys}} + \frac{\partial V(x,\lambda_0^{\mathrm{F}}(t))}{\partial x} dx, \end{aligned} \tag{3.43}$$

however we also expect $V(x,\lambda_0^{\mathrm{F}}(t))$ to evolve according to Itō's Lemma in which we can identify the work done such that

$$0 = -dQ_{\mathrm{sys}} + dV - dW - \frac{k_B T}{m\gamma} \frac{\partial^2 V(x,t)}{\partial x^2} dt. \tag{3.44}$$

This is not compatible with the first law in as much as we can no longer associate the heat as being the energy dissipated by the reaction force. Consequently, for complete consistency and extension to non-conservative forces we must consider the thermodynamic quantities as the result of Stratonovich summations from which we can then convert in Itō form if desired from Eq. (2.91). We therefore have for the heat under both conservative and non-conservative forces

$$\begin{aligned} dQ_{\mathrm{sys}} &= \frac{\partial V(x,\lambda_0^{\mathrm{F}}(t))}{\partial x} \circ dx - \mathcal{F}_{nc}(x,\lambda_1^{\mathrm{F}}(t)) \circ dx \\ &= \frac{\partial V(x,\lambda_0^{\mathrm{F}}(t))}{\partial x} dx + \frac{k_B T}{m\gamma} \frac{\partial^2 V(x,\lambda_0^{\mathrm{F}}(t))}{\partial x^2} dt - \mathcal{F}_{nc}(x,\lambda_1^{\mathrm{F}}) dx - \frac{k_B T}{m\gamma} \frac{\partial \mathcal{F}_{nc}(x,\lambda_1^{\mathrm{F}})}{\partial x} dt. \end{aligned} \tag{3.45}$$

Since we have now identified the heat flow (to the Langevin particle), we may identify what we expect to be the entropy production in the medium, since we have idealised noise with a defined temperature. As such we expect an increment in the medium entropy change to be given by

$$d\Delta\mathcal{S}_{\mathrm{med}} = \frac{-dQ_{\mathrm{sys}}}{T} = -\frac{1}{T} \frac{\partial V(x,\lambda_0^{\mathrm{F}}(t))}{\partial x} \circ dx + \frac{\mathcal{F}_{nc}(x,\lambda_1^{\mathrm{F}}(t))}{T} \circ dx. \tag{3.46}$$

To verify that such a quantity is genuinely the subject of the fluctuation relations however, we must have

$$k_B \ln \frac{\mathcal{P}^{\mathrm{F}}[\vec{\boldsymbol{x}}|\boldsymbol{x}(0)]}{\mathcal{P}^{\mathrm{R}}[\vec{\boldsymbol{x}}^{\mathrm{R}}|\boldsymbol{x}(\tau)]} = \int_0^\tau d\Delta\mathcal{S}_{\mathrm{med}}. \tag{3.47}$$

Since the dynamics are Markovian, we need only ensure this for an infinitesimal increment, for which we can easily find the result by comparing two short time propagators of the form found in Eq. (2.142), indeed using any choice of evaluation point r, which we find to first order in dt to agree with the expected result. We illustrate deliberately without rigour in order to avoid repetition later with a more general result how this can be found. We point out the propagator for the forward path can be written in the form

$$p^{\mathrm{F}}(x', t+dt|x,t) = \sqrt{\frac{m\gamma}{4\pi k_B T dt}} \exp\left[-\frac{m\gamma\left((x'-x) - \frac{1}{m\gamma}\mathcal{F}(x,\lambda^{\mathrm{F}}(t))dt\right)^2}{4k_B T dt}\right] \tag{3.48}$$

and similarly for the reverse path required for the increment in medium entropy change

$$p^{\mathrm{R}}(x'^{\mathrm{R}}, (\tau-t)+dt|x^{\mathrm{R}}, \tau-t) = \sqrt{\frac{m\gamma}{4\pi k_B T dt}} \exp\left[-\frac{m\gamma\left((x'^{\mathrm{R}}-x^{\mathrm{R}}) - \frac{1}{m\gamma}\mathcal{F}(x^{\mathrm{R}},\lambda^{\mathrm{R}}(\tau-t))dt\right)^2}{4k_B T dt}\right] \tag{3.49}$$

which we may write

$$p^{\mathrm{R}}(x, t+dt|x', t) = \sqrt{\frac{m\gamma}{4\pi k_B T dt}} \exp\left[-\frac{m\gamma\left(-(x'-x) - \frac{1}{m\gamma}\mathcal{F}(x',\lambda^{\mathrm{F}}(t))dt\right)^2}{4k_B T dt}\right]. \tag{3.50}$$

Writing $dx = x' - x$, and constructing the medium entropy change contribution we find

$$\begin{aligned}
\frac{1}{k_B} d\Delta\mathcal{S}_{\mathrm{med}} &= \frac{m\gamma}{4k_B T}\left(\frac{(dx)^2}{dt} - \frac{(dx)^2}{dt}\right. \\
&\qquad \left. + \frac{2}{m\gamma}\left(2\mathcal{F}(x,\lambda^{\mathrm{F}}(t)) + d\left(\mathcal{F}(x,\lambda^{\mathrm{F}}(t))\right)\right)dx + d\left(\left(\frac{\mathcal{F}(x,\lambda^{\mathrm{F}}(t))}{m\gamma}\right)^2\right)dt\right) \\
&= \frac{\mathcal{F}(x,\lambda^{\mathrm{F}}(t))}{k_B T} dx + \frac{1}{m\gamma}\frac{\partial \mathcal{F}(x,\lambda^{\mathrm{F}}(t))}{\partial x} dt + O(dt^{3/2}) \\
&= \frac{\mathcal{F}(x,\lambda^{\mathrm{F}}(t))}{k_B T} \circ dx \\
&= -\frac{dQ_{\mathrm{sys}}}{k_B T}
\end{aligned} \tag{3.51}$$

as required.

Full Phase Space Stochastic Energetics

We note that the preceding arguments for the identify for heat and work etc. can be extended to the full phase space Langevin equation which is given by

$$\begin{aligned}
m\frac{dv}{dt} &= \mathcal{F}(x,\lambda^{\mathrm{F}}(t)) - m\gamma\frac{dx}{dt} + \sqrt{2k_B T(x,t)m\gamma}\Gamma(t) \\
&= -\frac{\partial V(x,\lambda_0^{\mathrm{F}}(t))}{\partial x} + \mathcal{F}_{nc}(x,\lambda_1^{\mathrm{F}}(t)) - m\gamma\frac{dx}{dt} + \sqrt{2k_B T(x,t)m\gamma}\Gamma(t).
\end{aligned} \tag{3.52}$$

Following later work by Sekimoto where a full phase space consideration was given [45] a similar energy balance argument can be given, again using Stratonovich calculus. By grouping in terms of force and reaction force we obtain

$$0 = -\left(-m\gamma\frac{dx}{dt} + \sqrt{2k_BT(x,t)m\gamma}\Gamma(t)\right) + \frac{\partial V(x,\lambda_0^{\rm F}(t))}{\partial x} - \mathcal{F}_{nc}(x,\lambda_1^{\rm F}(t)) + m\frac{dv}{dt}. \tag{3.53}$$

Again we multiply by the increment dx to obtain the energy balance

$$\begin{aligned} 0 &= -\left(-m\gamma\frac{dx}{dt} + \sqrt{2k_BT(x,t)m\gamma}\Gamma(t)\right)\circ dx + \frac{\partial V(x,\lambda_0^{\rm F}(t))}{\partial x}\circ dx - \mathcal{F}_{nc}(x,\lambda_1^{\rm F}(t))\circ dx + m\frac{dv}{dt}\circ dx \\ &= -dQ_{\rm sys} + \frac{\partial V(x,\lambda_0^{\rm F}(t))}{\partial x}\circ dx - \mathcal{F}_{nc}(x,\lambda_1^{\rm F}(t))\circ dx + m\frac{dv}{dt}\circ dx \end{aligned} \tag{3.54}$$

where again the first term is the heat flow into the bath ($dQ = -dQ_{\rm sys}$) since it can be interpreted as the reaction force to the environment multiplied by increment dx. Accordingly we find that

$$\begin{aligned} dQ_{\rm sys} &= \left(-m\gamma\frac{dx}{dt} + \sqrt{2k_BT(x,t)m\gamma}\Gamma(t)\right)\circ dx \\ &= \frac{\partial V(x,\lambda_0^{\rm F}(t))}{\partial x}\circ dx - \mathcal{F}_{nc}(x,\lambda_1^{\rm F}(t))\circ dx + m\frac{dv}{dt}\circ dx. \end{aligned} \tag{3.55}$$

We mention that, strictly, the third expression in this final line is the only one that requires the distinction between integration schemes since x is now an integrated stochastic variable removing the ambiguity. This third term, however, using the Stratonovich integration scheme can be written

$$m\frac{dv}{dt}\circ dx = m\frac{dv}{dt}\circ vdt = mv\circ dv = d\left(\frac{mv^2}{2}\right) \tag{3.56}$$

allowing us to write, removing the Stratonovich notation where it is not needed,

$$\begin{aligned} dQ_{\rm sys} &= \frac{\partial V(x,\lambda_0^{\rm F}(t))}{\partial x}dx - \mathcal{F}_{nc}(x,\lambda_1^{\rm F}(t))dx + d\left(\frac{mv^2}{2}\right) \\ &= -\mathcal{F}(x,\lambda^{\rm F}(t))dx + d\left(\frac{mv^2}{2}\right). \end{aligned} \tag{3.57}$$

Finally we mention that this result too can be captured by comparing logarithmic path probabilities [80].

3.4 Specific Fluctuation Theorems

Such an identification of the heat transfer as the medium entropy change along with a consistent notion of the first law on the level of the fluctuating trajectory allows us now to explore the implications of the fluctuation theorem in different circumstances. Starting from, and examining the form of, Eq. (3.9) we see that, in general, it relates the distributions of two distinct physical observables, $\mathcal{A}[\vec{\boldsymbol{x}}]$ and $\hat{\mathcal{A}}[\vec{\boldsymbol{x}}^*]$, of which the first we have argued, for certain choices of protocol, initial distributions etc., is the microscopic total entropy production of system and environment along a stochastic trajectory. The second, however, being a functional defined using the reverse path and produced under the reverse dynamics would seem to encourage a consideration of the same thermodynamic quantity, the microscopic entropy production, of the reverse path under the reverse process. As such we examine the form of $k_B\hat{\mathcal{A}}[\vec{\boldsymbol{x}}^*] = \Delta\hat{\mathcal{S}}_{\rm tot}$ which

is given as

$$\frac{\Delta\hat{\mathcal{S}}_{\rm tot}}{k_B} = \hat{\mathcal{A}}[\vec{\boldsymbol{x}}^{\rm R}] = \ln\left[\frac{\mathcal{P}^{\rm R}[\vec{\boldsymbol{x}}^{\rm R}]_{p^{\rm F}_\tau}}{\mathcal{P}^{\rm F}[\vec{\boldsymbol{x}}]_{p^{\rm F}_0}}\right] = \ln\left[\frac{p^{\rm F}(\boldsymbol{x}(\tau),\tau)}{p^{\rm F}(\boldsymbol{x}(0),0)}\right] + \ln\left[\frac{\mathcal{P}^{\rm R}[\vec{\boldsymbol{x}}^{\rm R}|\boldsymbol{x}^{\rm R}(0)]}{\mathcal{P}^{\rm F}[\vec{\boldsymbol{x}}|\boldsymbol{x}(0)]}\right]. \tag{3.58}$$

We now contrast this with the microscopic entropy production of the reverse path in the reverse process by considering the functional $\mathcal{A}[\vec{\boldsymbol{x}}]$ constructed over the reverse process instead of the forward one where we denote this microscopic entropy production $\Delta\mathcal{S}^{\rm R}_{\rm tot} = k_B\mathcal{A}^{\rm R}[\vec{\boldsymbol{x}}^{\rm R}]$. This has the form

$$\frac{\Delta\mathcal{S}^{\rm R}_{\rm tot}}{k_B} = \mathcal{A}^{\rm R}[\vec{\boldsymbol{x}}^{\rm R}] = \ln\left[\frac{\mathcal{P}^{\rm R}[\vec{\boldsymbol{x}}^{\rm R}]_{p^{\rm R}_0}}{\mathcal{P}^{\rm F}[\vec{\boldsymbol{x}}]_{p^{\rm R}_\tau}}\right] = \ln\left[\frac{p^{\rm R}(\boldsymbol{x}^{\rm R}(0),0)}{p^{\rm R}(\boldsymbol{x}^{\rm R}(\tau),\tau)}\right] + \ln\left[\frac{\mathcal{P}^{\rm R}[\vec{\boldsymbol{x}}^{\rm R}|\boldsymbol{x}^{\rm R}(0)]}{\mathcal{P}^{\rm F}[\vec{\boldsymbol{x}}|\boldsymbol{x}(0)]}\right] \tag{3.59}$$

where analogously to the role of $p^{\rm F}(\boldsymbol{x}(t),t)$ in the boundary terms of the total entropy production of the forward process, $p^{\rm R}(\boldsymbol{x}^{\rm R}(t),t)$ is the solution of the Kolmogorov forward equation for the reverse process. By our specification of the reverse process we have $p^{\rm R}(\boldsymbol{x}^{\rm R}(0),0) = p^{\rm F}(\boldsymbol{x}^{\rm F}(\tau),\tau)$. Consequently we identify the equivalence of $\hat{\mathcal{A}}[\vec{\boldsymbol{x}}^{\rm R}]$ and $\mathcal{A}^{\rm R}[\vec{\boldsymbol{x}}^{\rm R}]$, or rather $\Delta\hat{\mathcal{S}}_{\rm tot}$ and $\Delta\mathcal{S}^{\rm R}_{\rm tot}$, *if and only if* $p^{\rm F}(\boldsymbol{x}(0),0) = p^{\rm R}(\boldsymbol{x}^{\rm R}(\tau),\tau)$. That is to say the transient fluctuation theorem explicitly considers distributions of entropy production in the forward and reverse processes *if* after the forward process, the reverse process returns the probability distribution to the distribution that acted as the initial distribution of the forward process. This does not, however, occur in any particular generality. If this did occur, in general, (as for example in Liouville's equation) it would imply complete reversibility in the dynamics and thus render such a definition of entropy as defined here, inappropriate. As such, many of the distinct fluctuation relations in the literature arise from specific situations where such a relation does hold.

3.4.1 Work Relations

In thermodynamics the work performed on a system is a quantity of great importance and it is most revealing that we may derive fluctuation theorems in such a way that they concern it. The most straightforward, or general, way to do so is to consider the relation

$$\langle g(\mathcal{C}[\vec{\boldsymbol{x}}^{\rm R}])\rangle^{\rm R}_{p^{\rm R}_0} = \langle g(\mathcal{B}[\vec{\boldsymbol{x}}])e^{-\mathcal{A}[\vec{\boldsymbol{x}}]}\rangle^{\rm F}_{p^{\rm F}_0} \tag{3.60}$$

and notice that the work done on the system obeys the symmetry requirements $\mathcal{C}[\vec{\boldsymbol{x}}^{\rm R}] = \mathcal{B}[\vec{\boldsymbol{x}}]$ when we choose $\mathcal{B}[\vec{\boldsymbol{x}}] = \Delta W$ and $\mathcal{C}[\vec{\boldsymbol{x}}^{\rm R}] = -\Delta W^{\rm R}$ and then choosing $p^{\rm R}_0(\boldsymbol{x}(0))$, $p^{\rm F}_0(\boldsymbol{x}(0))$ and $\mathcal{A}[\vec{\boldsymbol{x}}]$ to relate the nature of the averages with the appropriate functional. In general then we may write

$$\langle g(-\Delta W^{\rm R})\rangle^{\rm R}_{p^{\rm F}_\tau} = \langle g(\Delta W)e^{-k_B^{-1}\Delta\mathcal{S}_{\rm tot}}\rangle^{\rm F}_{p^{\rm F}_0}, \tag{3.61}$$

for which an important choice of g is the delta function allowing us to relate the distributions in each process through

$$p^{\rm R}(\Delta W^{\rm R} = -A)_{p^{\rm F}_\tau} = p^{\rm F}(\Delta W = A)_{p^{\rm F}_0}e^{-k_B^{-1}\Delta\mathcal{S}_{\rm tot}}, \tag{3.62}$$

where, by definition, $\Delta\mathcal{S}_{\rm tot}$ is the total entropy production produced by the same path that led to a work of $\Delta W = A$, with reverse process again starting from the distribution at the end of the forward process. If we now insist that, for example, the initial and final distributions both be equilibrium we must in turn adjust the boundary terms in the functional $\mathcal{A}[\vec{\boldsymbol{x}}]$ and the initial distributions in the averaging procedure

such that we now have

$$p^{\mathrm{R}}(\Delta W^{\mathrm{R}} = -A)_{p_\tau^{\mathrm{F,eq}}} = p^{\mathrm{F}}(\Delta W = A)_{p_0^{\mathrm{F,eq}}} e^{-k_B^{-1}\Delta \mathcal{S}_{\mathrm{tot}}^{\mathrm{eq}\to\mathrm{eq}}}, \tag{3.63}$$

where again $\Delta \mathcal{S}_{\mathrm{tot}}^{\mathrm{eq}\to\mathrm{eq}}$ takes the value arising from the path that produced $\Delta W = A$ in the forward dynamics. However, we also know that the transient fluctuation theorem for total entropy production holds when the reverse process returns the final distribution to the initial distribution; something that we can do trivially for equilibrium to equilibrium processes since we can simply allow the system to relax after any driving. As such we also expect the following relation to hold from Eq. (3.9)

$$p^{\mathrm{R}}(\Delta \mathcal{S}_{\mathrm{tot}}^{\mathrm{R,eq}\to\mathrm{eq}} = -A)_{p_0^{\mathrm{R,eq}}} = e^{-k_B^{-1}A}\, p^{\mathrm{F}}(\Delta \mathcal{S}_{\mathrm{tot}}^{\mathrm{eq}\to\mathrm{eq}} = A)_{p_0^{\mathrm{F,eq}}}. \tag{3.64}$$

The two are clearly related motivating a consideration of the equilibrium to equilibrium entropy production. To do so we simply need consider the initial and final distributions which represent the canonical ensemble. As before we consider the system energy for an over-damped system to be entirely described by the potential $V(x, \lambda_0^{\mathrm{F}}(t))$ such that we write[3]

$$p^{\mathrm{F,eq}}(x(t), \lambda_0^{\mathrm{F}}(t)) = \frac{1}{Z(\lambda_0^{\mathrm{F}}(t))} \exp\left[-\frac{V(x(t), \lambda_0^{\mathrm{F}}(t))}{k_B T}\right], \tag{3.65}$$

where Z is the partition function, uniquely defined by $\lambda_0^{\mathrm{F}}(t)$, which can in general be related to the Helmholtz free energy through the relation

$$F(\lambda_0^{\mathrm{F}}(t)) = -k_B T \ln Z(\lambda_0^{\mathrm{F}}(t)). \tag{3.66}$$

Let us now define the initial and final distributions to be given by the respective equilibria defined by the protocol at the start and finish of the forward process and the same temperature

$$\begin{aligned} p_0^{\mathrm{F,eq}}(\boldsymbol{x}(0)) &= p^{\mathrm{F,eq}}(\boldsymbol{x}(0), \lambda_0^{\mathrm{F}}(0)) = \exp\left[\frac{F(\lambda_0^{\mathrm{F}}(0)) - V(x(0), \lambda_0^{\mathrm{F}}(0))}{k_B T}\right] \\ p_0^{\mathrm{R,eq}}(\boldsymbol{x}^{\mathrm{R}}(0)) &= p_\tau^{\mathrm{F,eq}}(\boldsymbol{x}(\tau)) = p^{\mathrm{F,eq}}(\boldsymbol{x}(\tau), \lambda_0^{\mathrm{F}}(\tau)) = \exp\left[\frac{F(\lambda_0^{\mathrm{F}}(\tau)) - V(x(\tau), \lambda_0^{\mathrm{F}}(\tau))}{k_B T}\right]. \end{aligned} \tag{3.67}$$

Using these boundary terms we may construct the total entropy production by first considering the system entropy change $\Delta \mathcal{S}_{\mathrm{sys}}$ as

$$\begin{aligned} \Delta \mathcal{S}_{\mathrm{sys}} &= k_B \ln\left(\frac{p_0^{\mathrm{F,eq}}(\boldsymbol{x}(0))}{p_0^{\mathrm{R,eq}}(\boldsymbol{x}^{\mathrm{R}}(0))}\right) = k_B \ln\left(\frac{\exp\left[\frac{F(\lambda_0^{\mathrm{F}}(0)) - V(x(0), \lambda_0^{\mathrm{F}}(0))}{k_B T}\right]}{\exp\left[\frac{F(\lambda_0^{\mathrm{F}}(\tau)) - V(x(\tau), \lambda_0^{\mathrm{F}}(\tau))}{k_B T}\right]}\right) \\ &= \frac{1}{T}\left(-F(\lambda_0^{\mathrm{F}}(\tau)) + F(\lambda_0^{\mathrm{F}}(0)) + V(x(\tau), \lambda_0^{\mathrm{F}}(\tau)) - V(x(0), \lambda_0^{\mathrm{F}}(0))\right) \\ &= \frac{\Delta V - \Delta F}{T}. \end{aligned} \tag{3.68}$$

The medium entropy change is as we defined previously and can be written

$$\Delta \mathcal{S}_{\mathrm{med}} = \frac{\Delta Q}{T} = \frac{\Delta W - \Delta V}{T} \tag{3.69}$$

[3] We note a difference in notation, specifically in the second argument used here and throughout, between stationary (including equilibrium) states, in principle uniquely described by the protocol, for which we write $p^{\mathrm{F,eq}}(x(t), \lambda^{\mathrm{F}}(t))$ or $p^{\mathrm{F,st}}(x(t), \lambda^{\mathrm{F}}(t))$ and the general solution to the relevant forward equation, $p^{\mathrm{F}}(x(t), t)$, which cannot and as such utilises time as the second argument.

where ΔW, for an over-damped particle for example, is the work given earlier in Eq. (3.33), but we now emphasise that this term contains contributions due to changes in the potential and due to the external force $\mathcal{F}_{nc}$. We thus further define two new quantities ΔW_0 and ΔW_1 such that $\Delta W = \Delta W_0 + \Delta W_1$ with

$$\Delta W_0 = \int_0^\tau \frac{\partial V(x(t), \lambda_0^{\mathrm{F}}(t))}{\partial \lambda_0^{\mathrm{F}}(t)} \frac{d\lambda_0^{\mathrm{F}}(t)}{dt} dt \tag{3.70}$$

and

$$\Delta W_1 = \int_0^\tau \mathcal{F}_{nc}(x(t), \lambda_1^{\mathrm{F}}(t)) \circ dx. \tag{3.71}$$

ΔW_0 and ΔW_1 are not defined in the same way with ΔW_0 being found more often in thermodynamics and ΔW_1 being a familiar definition from mechanics; one may therefore refer to these definitions as thermodynamic and mechanical work respectively. The total entropy production in this case is simply given by

$$\Delta \mathcal{S}_{\mathrm{tot}}^{\mathrm{eq}\to\mathrm{eq}} = \frac{\Delta W - \Delta F}{T}. \tag{3.72}$$

Additionally, since we have met the conditions required to relate $\Delta\hat{\mathcal{S}}$ and $\Delta\mathcal{S}^{\mathrm{R}}$ we can write

$$\Delta \mathcal{S}_{\mathrm{tot}}^{\mathrm{R,eq}\to\mathrm{eq}} = \frac{\Delta W^{\mathrm{R}} - \Delta F^{\mathrm{R}}}{T} = -\frac{\Delta W - \Delta F}{T}. \tag{3.73}$$

As such we consider Eqs. (3.63) and (3.64) to describe precisely the same thing characterised by

$$p^{\mathrm{R}}((\Delta W^{\mathrm{R}} - \Delta F^{\mathrm{R}})/T = -A)_{p_\tau^{\mathrm{F,eq}}} = p^{\mathrm{F}}((\Delta W - \Delta F)/T = A)_{p_0^{\mathrm{F,eq}}} e^{-k_B^{-1}A} \tag{3.74}$$

noting that the free energy change ΔF and temperature T are just numbers and so can be excluded from the distributions and that such an expression is entirely equivalent to the transient fluctuation theorem for entropy production for a protocol that enables relaxation to equilibrium at the end of both forward and reverse processes. In practice such a protocol would initially be held at some value to allow relaxation in the reverse process then made time dependent to perform work and driving and then held at the final value to allow relaxation in the forward process.

The Crooks Work Relation and Jarzynski Equality

The specific relations that now follow arise by imposing certain constraints on the process we consider. First we consider the situation where the external force $\mathcal{F}_{nc}(x, \lambda_1^{\mathrm{F}}(t)) = 0$ and so all work is performed conservatively through the potential such that $\Delta W = \Delta W_0$. For such a process we write the total entropy production of the corresponding equilibrium to equilibrium process as

$$\Delta \mathcal{S}_{\mathrm{tot}}^{\mathrm{eq}\to\mathrm{eq}} = \frac{\Delta W_0 - \Delta F}{T} \tag{3.75}$$

allowing us to find [8]

The Crooks Work Relation:

$$\frac{p^{\mathrm{F}}\left(\Delta W_0 = A\right)_{p_0^{\mathrm{F,eq}}}}{p^{\mathrm{R}}\left(\Delta W_0^{\mathrm{R}} = -A\right)_{p_\tau^{\mathrm{F,eq}}}} = \exp\left[\frac{A - \Delta F}{k_B T}\right]. \tag{3.76}$$

Rearranging and integrating over all ΔW on both sides, and taking the deterministic ΔF out of the path integral then yields an expression for the average over the forward process called [7, 10, 81]

The Jarzynski Equality:

$$\langle \exp\left(-\Delta W_0/k_BT\right)\rangle^{\mathrm{F}}_{p_0^{\mathrm{F,eq}}} = \exp\left(-\Delta F/k_BT\right). \tag{3.77}$$

We point out that we can formulate these two relations by considering the entropy production of an equilibrium to equilibrium process, but the symmetries they describe hold for any process as long as the relevant initial conditions are prescribed as being equilibrium (as indicated by the relevant subscripts). Consequently, the system does not need to be in equilibrium at the end of both the forward and reverse process. Historically this has had one particularly important consequence: the results hold for driving, in principle, arbitrarily far from equilibrium. This is widely summed up as the ability to obtain equilibrium information from non-equilibrium averaging since, upon examining the form of the Jarzynski equality, we can compute the free energy difference by taking an average of the exponentiated work done in the course of some non-equilibrium process. It should be noted however, that one must exercise caution if performing such an average since it is patently dominated by very rare, negative entropy fluctuations meaning in practice one may need to perform a very large number of realisations in order to yield a reliable average from such a procedure. Alternatively, if one can perform the reverse process, one may turn to the Crooks relation and identify the point where both distributions cross in order to find the free energy change. Finally we note both the Crooks relation and Jarzynski equality can be seen as refinements to the usual second law. The Crooks relation quantitatively relates the probability we would extract more work than we put in, in for example, the compression then expansion of a gas and applying Jensen's inequality to the Jarzynski equality yields

$$\langle \Delta W_0\rangle^{\mathrm{F}}_{p_0^{\mathrm{F,eq}}} \geq \Delta F \tag{3.78}$$

reducing to the usual thermodynamic measure of reversibility in the thermodynamic limit with the equality holding for quasi-static reversible processes.

3.4.2 Fluctuation Relations for Mechanical Work

A similar, but subtly different circumstance to that of the Jarzynski and Crooks relations is that where we consider a driving process that again starts in equilibrium, but this time keeps the protocol $\lambda_0^{\mathrm{F}}(t)$ held fixed such that all work is performed by the externally applied force $\mathcal{F}_{nc}(x, \lambda_1^{\mathrm{F}}(t))$ meaning that $\Delta W = \Delta W_1$. Once again we may find a fluctuation relation concerning a work equivalent to the entropy production of some equilibrium to equilibrium process. This time however, the corresponding equilibrium at the start of both the forward and reverse processes are identical and correspond, assuming an equilibrium exists, to the same value of both protocols $\lambda_0^{\mathrm{F}}(t)$ and $\lambda_1^{\mathrm{F}}(t)$. For such a process we find by similar means

$$\Delta \mathcal{S}_{\mathrm{tot}} = \frac{\Delta W_1}{T} \tag{3.79}$$

since the free energy difference between the same equilibrium states vanishes. This in turn leads to a set of fluctuation relations [82–85] which may collectively be referred to as

Bochkov-Kuzovlev relations for mechanical work:

$$\frac{p^{\mathrm{F}}(\Delta W_1 = A)_{p_0^{\mathrm{F,eq}}}}{p^{\mathrm{R}}(\Delta W_1^{\mathrm{R}} = -A)_{p_\tau^{\mathrm{F,eq}}}} = \exp\left[\frac{A}{k_BT}\right] \tag{3.80}$$

and

$$\langle \exp\left(-\Delta W_1/k_B T\right)\rangle^{\mathrm{F}}_{p_0^{\mathrm{F,eq}}} = 1. \tag{3.81}$$

For the same reasons as in the Jarzynski and Crooks relations they are valid for all times and thus hold as a non-equilibrium result. Taking in particular the integrated relation and comparing with the Jarzynski equality in Eq. (3.77) one may think there is an inconsistency. Both are valid for all times and arbitrary driving and concern the work done under the constraint that both start in equilibrium, yet on first inspection they seem to be saying different things. But recall our distinction between the work ΔW_0 and ΔW_1 from Eqs. (3.70) and (3.71); there are two distinct ways to describe work on such a particle. If one performs work ΔW_0 one necessarily changes the form of the system energy whereas the application of work ΔW_1 leaves the form of the system energy unchanged. The difference is manifest in the two different integrated relations because their derivations exploit the fact that the Hamiltonian, which represents the system energy, appears in initial and final distributions. To clarify, as written the Jarzynski equality explicitly concerns driving where the application of any work also changes the Hamiltonian and thus the equilibrium state. On the other hand the relations for W_1 concern work as the path integral of an external force such that the Hamiltonian remains unchanged for the entire process. We briefly note that the use of this difference in definition of work led to some discussion [86–89], but as illustrated in [82, 83] it is something of an aesthetic decision (with resulting changes of form in the work relation) concerning how heat exchanges are divided into a work and internal energy contribution, but the result which leads to a work relation at all, the form of the heat and its relation to ratios of path probabilities, is the same regardless of interpretation.

Of course, there is nothing in the derivation of either of these relations that precludes the possibility of both types of work to be performed at the same time and so using the same arguments we arrive at

$$\frac{p^{\mathrm{F}}(\Delta W = A)_{p_0^{\mathrm{F,eq}}}}{p^{\mathrm{R}}(\Delta W^{\mathrm{R}} = -A)_{p_\tau^{\mathrm{F,eq}}}} = \exp\left[(A - \Delta F)/k_B T\right] \tag{3.82}$$

and

$$\left\langle \exp\left(-\frac{\Delta W - \Delta F}{k_B T}\right)\right\rangle^{\mathrm{F}}_{p_0^{\mathrm{F,eq}}} = 1 \tag{3.83}$$

again under the constraint that the system be initially prepared in equilibrium.

3.4.3 Kawasaki Relation

A result in the same spirit of these work relations is the non-equilibrium response relation which is achieved by using the relation in Eq. (3.14) for a given function. Since we have $\boldsymbol{x}^*(t) = \boldsymbol{x}^{\mathrm{R}}(t)$ the appropriate transform of the function $g(x(t))$ in the forward process is $g(x^{\mathrm{R}}(\tau - t))$ in the reverse process. As such we have

$$\langle g(\boldsymbol{x}^{\mathrm{R}}(\tau - t))\rangle^{\mathrm{R}}_{p_\tau^{\mathrm{F}}} = \langle g(\boldsymbol{x}(t)) e^{-k_B^{-1}\Delta \mathcal{S}_{\mathrm{tot}}}\rangle^{\mathrm{F}}_{p_0^{\mathrm{F}}}. \tag{3.84}$$

If one then specifies that the function is evaluated at the end of the process, that is $t = \tau$, and that we construct the functional $\mathcal{A}[\vec{\boldsymbol{x}}]$ so that it corresponds to equilibrium initial distributions and that the process is isothermal we may simplify to find

$$\langle g(\boldsymbol{x}^{\mathrm{R}}(0))\rangle^{\mathrm{R}}_{p_\tau^{\mathrm{F,eq}}} = \langle g(\boldsymbol{x}(\tau)) e^{-(\Delta W - \Delta F)/k_B T}\rangle^{\mathrm{F}}_{p_0^{\mathrm{F,eq}}}. \tag{3.85}$$

We then note that the left hand side is simply an instantaneous equilibrium average of the function $g(\boldsymbol{x})$ appropriate to the value of the protocol corresponding to the end of the forward process allowing us to relate it to the non-equilibrium average on the right hand side.

3.4.4 Generalised Crooks Relation

Exploiting the fact we have just seen that the equilibrium to equilibrium entropy production satisfies the symmetry requirements for $\mathcal{B}[\vec{\boldsymbol{x}}]$ and $\mathcal{C}[\vec{\boldsymbol{x}}^*]$ we can apply this more generally without the isothermal condition. As such we can write, as a slight generalisation of the Crooks relation

$$p^{\mathrm{R}}(\Delta\mathcal{S}_{\mathrm{tot}}^{\mathrm{R,eq\to eq}} = -A)_{p_\tau^{\mathrm{F}}} = p^{\mathrm{F}}(\Delta\mathcal{S}_{\mathrm{tot}}^{\mathrm{eq\to eq}} = A)_{p_0^{\mathrm{F}}} e^{-k_B^{-1}\Delta\mathcal{S}_{\mathrm{tot}}}, \tag{3.86}$$

again where $\Delta\mathcal{S}_{\mathrm{tot}}$ takes the value that would be realised from the path that led to $\Delta\mathcal{S}_{\mathrm{tot}}^{\mathrm{eq\to eq}} = A$ in the forward dynamics, but this time allow a protocol dependent temperature such that the total entropy production generalises to

$$\begin{aligned}\Delta\mathcal{S}_{\mathrm{tot}} &= \Delta\mathcal{S}_{\mathrm{sys}} + \int_0^\tau \frac{dQ}{T} \\ &= \Delta\mathcal{S}_{\mathrm{sys}} - \int_0^\tau \frac{dQ_{\mathrm{sys}}}{T}.\end{aligned} \tag{3.87}$$

In the same way that the Jarzynski equality generalised the irreversibility statement found in classical thermodynamics $\Delta W \geq \Delta F$ to $\langle \Delta W \rangle \geq \Delta F$ by the convexity of the exponential function, because the mean of the equilibrium system entropy is the well defined equilibrium Gibbs entropy one has a Clausius like statement

$$\Delta S = \langle \Delta\mathcal{S}_{\mathrm{sys}}^{\mathrm{eq\to eq}} \rangle \geq \left\langle \int_0^\tau \frac{dQ_{\mathrm{sys}}}{T} \right\rangle_{p_0^{\mathrm{F,eq}}}^{\mathrm{F}} \tag{3.88}$$

and for a cyclic process

$$0 \geq \left\langle \int_0^\tau \frac{dQ_{\mathrm{sys}}}{T} \right\rangle_{p_0^{\mathrm{F,eq}}}^{\mathrm{F}} \tag{3.89}$$

which reduces to the Clausius inequality in the thermodynamic limit.

3.4.5 Fluctuation Theorems for Entropy Production

Understanding that the transient fluctuation theorem concerns total entropy production when the reverse process leaves the resultant distribution in that found at the beginning of the forward process, we have seen that work relations can be constructed by meeting this criterion using equilibrium initial distributions. We now consider more generally scenarios where this result holds, but we do not require equilibrium distributions. As perhaps the most general statement of the above reasoning we can state that if we have the condition

$$\begin{aligned}p^{\mathrm{F}}(\boldsymbol{x}^{\mathrm{R}}(\tau), 0) &= \int d\boldsymbol{x}(0) \int [\mathcal{D}\boldsymbol{x}^{\mathrm{R}}]\, p^{\mathrm{R}}(\boldsymbol{x}^{\mathrm{R}}(0), 0) p^{\mathrm{R}}[\vec{\boldsymbol{x}}^{\mathrm{R}} | \boldsymbol{x}^{\mathrm{R}}(0)] \\ &= \int d\vec{\boldsymbol{x}}^{\mathrm{R}}_{\neq \boldsymbol{x}^{\mathrm{R}}(\tau)}\, p^{\mathrm{R}}(\boldsymbol{x}^{\mathrm{R}}(0), 0) p^{\mathrm{R}}[\vec{\boldsymbol{x}}^{\mathrm{R}} | \boldsymbol{x}^{\mathrm{R}}(0)]\end{aligned} \tag{3.90}$$

then we have the following symmetry relation between the entropy produced in the forward and reverse processes

$$p^{\mathrm{R}}(\Delta\mathcal{S}_{\mathrm{tot}}^{\mathrm{R}}[\vec{\boldsymbol{x}}^{\mathrm{R}}] = -A)_{p_\tau^{\mathrm{F}}} = e^{-k_B^{-1}A}\, p^{\mathrm{F}}(\Delta\mathcal{S}_{\mathrm{tot}}[\vec{\boldsymbol{x}}] = A)_{p_0^{\mathrm{F}}}. \tag{3.91}$$

The challenge of utilising such a symmetry then becomes one of finding situations that meet its requirements. One of the most straightforward ways to achieve this however, reveals an even more direct symmetry by insisting that the evolution under the forward process is indistinguishable from that under the reverse process and that $p_\tau^{\mathrm{F}} = p_0^{\mathrm{F}}$. Mathematically the indistinguishability of the dynamics is the requirement $\mathcal{P}^{\mathrm{R}}[\vec{\boldsymbol{x}}^*|\boldsymbol{x}^*(0)] = \mathcal{P}^{\mathrm{F}}[\vec{\boldsymbol{x}}^*|\boldsymbol{x}^*(0)]$. Given these conditions, evolution from the initial distribution will trivially result in the final distribution and evolution under the reverse process from the final distribution will result in the initial distribution. If we consider in more detail the requirements for such behaviour we understand there are two main ways in which this can be achieved. Given that the initial and final distributions are the same the first way is to require a constant protocol $\lambda^{\mathrm{F}}(t)$. In this way the forward process is trivially the same as the reverse process. Alternatively we could require the protocol to be time symmetric such that $\lambda^{\mathrm{F}}(t) = \lambda^{\mathrm{F}}(\tau - t) = \lambda^{\mathrm{R}}(t)$. In both situations the forward and reverse processes are entirely indistinguishable. As such, by careful construction we can, in these specific circumstances, relate the probability of seeing a positive entropy production to that of a negative entropy production over the *same forward process* allowing us from Eq. (3.9) to write a [12]

Detailed fluctuation theorem (DFT):

$$p^{\mathrm{F}}(\Delta\mathcal{S}_{\mathrm{tot}} = -A)_{p_0^{\mathrm{F}}} = \mathrm{e}^{-k_B^{-1}A} p^{\mathrm{F}}(\Delta\mathcal{S}_{\mathrm{tot}} = A)_{p_0^{\mathrm{F}}}. \tag{3.92}$$

Physically the two situations we have considered correspond to

- $p_0^{\mathrm{F}} = p_\tau^{\mathrm{F}} = p_0^{\mathrm{F,st}} = p^{\mathrm{F,st}}(\boldsymbol{x}, \lambda^{\mathrm{F}}(0))$, $\lambda^{\mathrm{F}}(t) = \text{const}$:
  To satisfy such criteria the system must be in a steady state, that is all intrinsic system properties (probability distribution, mean system entropy, mean system energy etc.) must remain constant over the process. The simplest steady state is equilibrium which trivially has zero entropy production in detail for all trajectories. However, a *non-equilibrium* steady state can be achieved by breaking detailed balance through some constraint which prevents the equilibration. The mean entropy production rate of these states is constant, non-zero and, as we have now shown, there is an explicit exponential symmetry in the probability of positive and negative fluctuations. For clarity, path integrals that utilise a constant protocol in this way, such that they approach stationary states, are indicated by the notation $\langle\rangle^{\mathrm{F,st}}$.

- $p_0^{\mathrm{F}} = p_\tau^{\mathrm{F}}$, $\lambda^{\mathrm{F}}(t) = \lambda^{\mathrm{R}}(t)$:
  This condition can be achieved in a system that is being periodically driven characterised by a time symmetric $\lambda^{\mathrm{F}}(t)$. If from some starting point we allow the system to undergo an arbitrarily large number of periods of driving it will arrive at a so-called non-equilibrium oscillatory state such that $p^{\mathrm{F}}(x,t) = p^{\mathrm{F}}(x, t + t_p)$ where t_p is the period of oscillation. In this state we can expect the above relation to hold for integer multiples of period t_p starting from a time such that $\lambda^{\mathrm{F}}(t) = \lambda^{\mathrm{R}}(t)$.

Such a result, because it directly concerns entropy production in the steady state, is frequently considered to be a finite time analogue of the asymptotic fluctuation theorems that hold in steady states (and the approach to them) in the limit $t \to \infty$. Considering, in particular, the analogous relation of Eq. (3.16) that corresponds to a non-equilibrium stationary state we have

$$\langle e^{-(1-\lambda)k_B^{-1}\Delta\mathcal{S}_{\mathrm{tot}}}\rangle^{\mathrm{F,st}}_{p_0^{\mathrm{F,st}}} = \langle e^{-\lambda k_B^{-1}\Delta\mathcal{S}_{\mathrm{tot}}}\rangle^{\mathrm{F,st}}_{p_0^{\mathrm{F,st}}}. \tag{3.93}$$

By some manipulation one can show that this fluctuation theorem reproduces the Kubo relations for transport coefficients when taken close to equilibrium giving an insight into its place as a genuine non-equilibrium result and to do so we loosely follow the procedures documented in, for example, [90, 91].

Starting with the symmetry in Eq. (3.93) we may express the same symmetry in the scaled cumulant generating function

$$g(\lambda) = g(1-\lambda) = -\tau^{-1} \ln\langle \exp\left[-\lambda k_B^{-1} \Delta \mathcal{S}_{\rm tot}\right]\rangle_{p_0^{\rm F,st}}^{\rm F,st} \tag{3.94}$$

with

$$g^k = \frac{\partial^k g}{\partial \lambda^k}\Big|_{\lambda=0} = (-1)^{k-1}\tau^{-1}\langle (k_B^{-1}\Delta \mathcal{S}_{\rm tot})^k\rangle_c \tag{3.95}$$

where $\langle (k_B^{-1}\Delta \mathcal{S}_{\rm tot})^k\rangle_c$ denotes the $k^{\rm th}$ cumulant of $\langle k_B^{-1}\Delta \mathcal{S}_{\rm tot}\rangle_{p_0^{\rm F}}^{\rm F}$. We then assume that close to equilibrium, in the regime of linear irreversible thermodynamics, the mean path dependent total entropy production is equivalent to the positive internal entropy production of near equilibrium states and is thus given by

$$\langle \Delta \mathcal{S}_{\rm tot}\rangle_{p_0^{\rm F}}^{\rm F,st} \simeq \int_0^\tau \frac{\langle J_{\rm th}(t)\rangle \mathcal{F}_{\rm th}}{T} dt = \int_0^\tau \langle \dot{\mathcal{S}}_{\rm tot}(t)\rangle dt \tag{3.96}$$

for a small thermodynamic force $\mathcal{F}_{\rm th}$ and conjugate flux $J_{\rm th}$ such that $\langle \dot{\mathcal{S}}_{\rm tot}(t)\rangle$ is small. Explicitly, this is the assumed ability to write the medium entropy change as

$$\Delta \mathcal{S}_{\rm med} = -\Delta \mathcal{S}_{\rm sys}^{\rm eq} + \frac{\mathcal{F}_{\rm th}}{T}\int_0^\tau J_{\rm th}(t) dt, \tag{3.97}$$

where $\Delta \mathcal{S}_{\rm sys}^{\rm eq}$ is the system entropy change one expects in equilibrium from the path that generates $\Delta \mathcal{S}_{\rm med}$, so that as we approach equilibrium the above reasoning holds. For example, for the underdamped Langevin equation subject to both conservative forces arising as a potential in a Hamiltonian and a non-conservative force which does not and gives rise to a physical particle current we have

$$\Delta \mathcal{S}_{\rm med} = \underbrace{-\frac{1}{T}\int_0^\tau d\left(\frac{mv^2}{2}\right) - \frac{1}{T}\int_0^\tau \frac{\partial V(x)}{\partial x} dx}_{-\Delta \mathcal{S}_{\rm sys}^{\rm eq}} + \frac{\mathcal{F}_{\rm nc}}{T}\int_0^\tau v dt \tag{3.98}$$

where the flux is simply the particle current v and the system entropy change, in equilibrium, is the change in the Hamiltonian scaled by the environmental temperature. If we then recognise that in the stationary state we may write

$$(-1)^{k-1} g^k = \tau^{-1}\langle (k_B^{-1}\Delta \mathcal{S}_{\rm tot})^k\rangle_c = \langle (k_B^{-1}\dot{\mathcal{S}}_{\rm tot}(t))^k\rangle_c \tag{3.99}$$

we may, after identifying the behaviour $g^k \sim \langle k_B^{-k}\dot{\mathcal{S}}_{\rm tot}^k\rangle$ valid in the small $\mathcal{F}_{\rm th}$, small $\langle \dot{\mathcal{S}}_{\rm tot}(t)\rangle$ limit, represent $g(\lambda)$ well by the expansion

$$g(\lambda) \simeq g^0 + g^1\lambda + g^2\frac{\lambda^2}{2} + \dots . \tag{3.100}$$

However, we have $g(\lambda) = g(1-\lambda)$ and so have the relation $g^2 = -2g^1$. In turn we note this implies Gaussianity in the linear regime since $2k_B^{-1}\langle \Delta \mathcal{S}_{\rm tot}\rangle_c = k_B^{-2}\langle \Delta \mathcal{S}_{\rm tot}^2\rangle_c$ and allows us to write

$$2k_B^{-1}\tau^{-1}\langle \Delta \mathcal{S}_{\rm tot}\rangle_{p_0^{\rm F,st}}^{\rm F,st} = k_B^{-2}\tau^{-1}\left\langle (\Delta \mathcal{S}_{\rm tot} - \langle \Delta \mathcal{S}_{\rm tot}\rangle)^2\right\rangle_{p_0^{\rm F,st}}^{\rm F,st}. \tag{3.101}$$

We can write this however, after eliminating τ and k_B and using $\Delta\dot{\mathcal{S}}_{\rm tot}(t) = \dot{\mathcal{S}}_{\rm tot}(t) - \langle\dot{\mathcal{S}}_{\rm tot}(t)\rangle$, as

$$2\left\langle\Delta\mathcal{S}_{\rm tot}\right\rangle_{p_0^{\rm F,st}}^{\rm F,st} = k_B^{-1}\left\langle\left(\int_0^\tau dt_1\ \left(\dot{\mathcal{S}}_{\rm tot}(t_1) - \langle\dot{\mathcal{S}}_{\rm tot}(t_1)\rangle\right)\right)\left(\int_0^\tau dt_2\ \left(\dot{\mathcal{S}}_{\rm tot}(t_2) - \langle\dot{\mathcal{S}}_{\rm tot}(t_2)\rangle\right)\right)\right\rangle_{p_0^{\rm F,st}}^{\rm F,st}$$
$$= k_B^{-1}\int_0^\tau dt_1\int_0^\tau dt_2\langle\Delta\dot{\mathcal{S}}_{\rm tot}(t_1)\Delta\dot{\mathcal{S}}_{\rm tot}(t_2)\rangle_{p_0^{\rm F,st}}^{\rm F,st}. \tag{3.102}$$

Now, since we are in the steady state we have the following properties of a correlation function

$$\langle A(t_1)A(t_2)\rangle = \langle A(t_1-t_2)A(0)\rangle = \langle A(t_2-t_1)A(0)\rangle \tag{3.103}$$

such that

$$\int_0^\tau dt_1\int_0^\tau dt_2\langle A(t_1)A(t_2)\rangle = \int_0^\tau dt_1\int_0^\tau dt_2\langle A(t_1-t_2)A(0)\rangle$$
$$= \int_0^\tau dt_1\int_{t_1-\tau}^{t_1} dt'\langle A(t')A(0)\rangle \tag{3.104}$$

achieved by the change of variable $t_2 \to t' = t_1 - t_2$. Changing the order of integration and separating into positive and negative t' domains we then have

$$\int_0^\tau dt_1\int_{t_1-\tau}^{t_1} dt'\langle A(t')A(0)\rangle = \int_{-\tau}^0 dt'\int_0^{t'+\tau} dt_1\langle A(t')A(0)\rangle + \int_0^\tau dt'\int_{t'}^\tau dt_1\langle A(t')A(0)\rangle. \tag{3.105}$$

Changing variable $t' \to -t'$ in the first of these integrals then yields

$$\int_0^\tau dt_1\int_{t_1-\tau}^{\tau} dt'\langle A(t')A(0)\rangle = \int_0^\tau dt'\int_0^{\tau-t'} dt_1\langle A(-t')A(0)\rangle + \int_0^\tau dt'\int_{t'}^\tau dt_1\langle A(t')A(0)\rangle$$
$$= \int_0^\tau dt'\int_0^{\tau-t'} dt_1\langle A(t')A(0)\rangle + \int_0^\tau dt'\int_{t'}^\tau dt_1\langle A(t')A(0)\rangle \tag{3.106}$$

by virtue of Eq. (3.103). Changing variable once more in the first integral $t_1 \to x = t_1 + t'$ we have

$$\int_0^\tau dt_1\int_{t_1-\tau}^{\tau} dt'\langle A(t')A(0)\rangle = \int_0^\tau dt'\int_{t'}^\tau dx\langle A(t')A(0)\rangle + \int_0^\tau dt'\int_{t'}^\tau dt_1\langle A(t')A(0)\rangle$$
$$= 2\int_0^\tau dt'\int_{t'}^\tau dt_1\langle A(t')A(0)\rangle. \tag{3.107}$$

This is then simply

$$2\int_0^\tau dt'\int_{t'}^\tau dt_1\langle A(t')A(0)\rangle = 2\tau\int_0^\tau dt'\langle A(t')A(0)\rangle - 2\int_0^\tau dt't'\langle A(t')A(0)\rangle. \tag{3.108}$$

Inserting this into the relation between mean and variance of the entropy production we then have

$$\langle\Delta\mathcal{S}_{\rm tot}\rangle_{p_0^{\rm F,st}}^{\rm F,st} = k_B^{-1}\tau\int_0^\tau dt'\langle\Delta\dot{\mathcal{S}}_{\rm tot}(t')\Delta\dot{\mathcal{S}}_{\rm tot}(0)\rangle_{p_0^{\rm F,st}}^{\rm F,st} - k_B^{-1}\int_0^\tau dt't'\langle\Delta\dot{\mathcal{S}}_{\rm tot}(t')\Delta\dot{\mathcal{S}}_{\rm tot}(0)\rangle_{p_0^{\rm F,st}}^{\rm F,st} \tag{3.109}$$

or

$$\langle\dot{\mathcal{S}}_{\rm tot}\rangle_{p_0^{\rm F,st}}^{\rm F,st} = k_B^{-1}\int_0^\tau dt'\langle\Delta\dot{\mathcal{S}}_{\rm tot}(t')\Delta\dot{\mathcal{S}}_{\rm tot}(0)\rangle_{p_0^{\rm F,st}}^{\rm F,st} - \frac{1}{\tau k_B}\int_0^\tau dt't'\langle\Delta\dot{\mathcal{S}}_{\rm tot}(t')\Delta\dot{\mathcal{S}}_{\rm tot}(0)\rangle_{p_0^{\rm F,st}}^{\rm F,st}. \tag{3.110}$$

In the limit $\tau \to \infty$ the second integral vanishes allowing us to write

$$\langle \dot{\mathcal{S}}_{\rm tot}(\tau)\rangle^{\rm F,st}_{p_0^{\rm F,st}} = k_B^{-1}\int_0^\infty dt \langle \Delta\dot{\mathcal{S}}_{\rm tot}(t)\Delta\dot{\mathcal{S}}_{\rm tot}(0)\rangle^{\rm F,st}_{p_0^{\rm F,st}} \tag{3.111}$$

or arranging in terms of fluxes and forces

$$\langle J_{\rm th}\rangle^{\rm F,st}_{p_0^{\rm F,st}} = L(\mathcal{F}_{\rm th}=0)\mathcal{F}_{\rm th} = \frac{\mathcal{F}_{\rm th}}{k_BT}\int_0^\infty dt\, \langle \Delta J_{\rm th}(t)\Delta J_{\rm th}(0)\rangle^{\rm F,st}_{p_0^{\rm F,st}} \tag{3.112}$$

which is the generic Green-Kubo relation for linear transport coefficients [92] in the limit $\mathcal{F}_{\rm th} \to 0$ such that $p_0^{\rm F,st} \to p_0^{\rm F,eq}$ when ignoring $O(\mathcal{F}_{\rm th}^2)$ terms.

3.5 Entropy Production Theorems and an Alternative Division of the Total Entropy Production

So far we have considered a functional that represents the total entropy production of a stochastic system and seen justification based on the notion of a division of entropy production into a heat flow associated with a medium and a system entropy term providing a physical rationale for boundary terms. However, once such reasoning is accepted one may go further and define further quantities that constitute the total entropy production which may provide some physical insight into the origins of that entropy production.

3.5.1 Division Based on the Existence of Non-equilibrium Stationary States

Division of Heat Flow

An operational formalism of non-equilibrium, or rather steady state, thermodynamics was presented by Oono and Paniconi [35] and consisted of postulating that the heat flow out of the system to the environment could be divided into two distinct contributions called the 'excess' and 'house-keeping' heat transfers (to the environment), which we write shorthand as simply excess and house-keeping heats, such that

$$\Delta Q = \Delta Q_{\rm ex} + \Delta Q_{\rm hk}. \tag{3.113}$$

It was argued that in order to maintain any non-equilibrium stationary state there would be need to be a constant dissipation of heat to break detailed balance and that this contribution was the house-keeping heat. Consequently the remainder, the excess heat, would be a bounded contribution which would act much more like the total heat flow between equilibrium stationary states.

Hatano-Sasa Equality and Speck-Seifert Relations

Introducing the notation of a non-equilibrium potential related to the stationary state through

$$p^{\rm F,st}(\boldsymbol{x},\lambda^{\rm F}(t)) = \exp\left[-\phi(\boldsymbol{x},\lambda^{\rm F}(t))\right], \tag{3.114}$$

Hatano and Sasa derived a relationship conspicuously similar in form to fluctuation theorems, under the paradigm of dynamics given by

$$dx = -\frac{1}{m\gamma}\frac{\partial V(x,\lambda^{\rm F}(t))}{\partial x}dt + \frac{\mathcal{F}_{nc}}{m\gamma}dt + \frac{k_BT}{m\gamma}dW, \tag{3.115}$$

given by

$$\left\langle \exp\left[-\int_0^\tau dt \frac{d\lambda^{\mathrm{F}}(t)}{dt}\frac{\partial\phi(x,\lambda^{\mathrm{F}}(t))}{\partial\lambda^{\mathrm{F}}(t)}\right]\right\rangle^{\mathrm{F}}_{p_0^{\mathrm{F,st}}} = 1 \tag{3.116}$$

originally considered as an extension of the Jarzynski equality, valid between non-equilibrium stationary states, as evidenced by the substitution $\phi(x,\lambda^{\mathrm{F}}(t)) = -\beta(F - V(x,\lambda^{\mathrm{F}}(t)))$ and $p_0^{\mathrm{F,st}} = p_0^{\mathrm{F,eq}}$. Further, by utilising the definition of the heat from stochastic energetics in Eq. (3.34), we may write

$$\begin{aligned} dQ &= \mathcal{F}(x)\circ dx \\ &= \mathcal{F}(x)\circ dx - k_BT\left(d\phi - \frac{\partial\phi(x,\lambda^{\mathrm{F}}(t))}{\partial x}\circ dx - \frac{\partial\phi(x,\lambda^{\mathrm{F}}(t))}{\partial\lambda^{\mathrm{F}}}d\lambda^{\mathrm{F}}\right) \\ &= \underbrace{\left(\mathcal{F}(x) + k_BT\frac{\partial\phi(x,\lambda^{\mathrm{F}}(t))}{\partial x}\right)\circ dx}_{dQ_{\mathrm{hk}}} \underbrace{-k_BTd\phi(x,\lambda^{\mathrm{F}}(t)) + k_BT\frac{\partial\phi(x,\lambda^{\mathrm{F}}(t))}{\partial\lambda^{\mathrm{F}}(t)}d\lambda^{\mathrm{F}}(t)}_{dQ_{\mathrm{ex}}} \end{aligned} \tag{3.117}$$

operationally defining, for the considered dynamics, the house-keeping and excess heats. It in turn allows one to write the Hatano-Sasa equality in terms of the non-equilibrium potential and excess heat [93]

$$\left\langle \exp\left[-\Delta\phi - \frac{\Delta Q_{\mathrm{ex}}}{k_BT}\right]\right\rangle^{\mathrm{F}}_{p_0^{\mathrm{F,st}}} = 1. \tag{3.118}$$

Subsequently Speck and Seifert showed, by demonstrating an explicit invariance in time of the exponential average, that, for all time, the corresponding house-keeping heat obeys [94]

$$\left\langle \exp\left[-\frac{\Delta Q_{\mathrm{hk}}}{k_BT}\right]\right\rangle^{\mathrm{F}}_{p_0^{\mathrm{F}}} = 1 \tag{3.119}$$

with the two relations together providing a framework and set of statistical restrictions on the quantities that play a role in the thermodynamics of transitions between steady states.

Adiabatic and Non-adiabatic Entropy Production

More recently Van den Broeck et al. [36], building upon the contribution of many others [95–97], have generalised the concept of splitting the heat flow, by applying such ideas to the total entropy production for arbitrary Markovian systems, without requiring the process in question to concern a transition between stationary states and going some way further in describing the distinct physical mechanisms by which entropy is generated. The fundamental division is as below

$$\begin{aligned} \frac{\Delta\mathcal{S}_{\mathrm{tot}}}{k_B} &= \ln\frac{\mathcal{P}^{\mathrm{F}}[\vec{\boldsymbol{x}}]_{p_0^{\mathrm{F}}}}{\mathcal{P}^{\mathrm{R}}[\vec{\boldsymbol{x}}^{\mathrm{R}}]_{p_\tau^{\mathrm{F}}}} \\ &= \ln\frac{\mathcal{P}^{\mathrm{F}}[\vec{\boldsymbol{x}}]_{p_0^{\mathrm{F}}}}{\mathcal{P}^{\mathrm{R,ad}}[\vec{\boldsymbol{x}}^{\mathrm{R}}]_{p_\tau^{\mathrm{F}}}} + \ln\frac{\mathcal{P}^{\mathrm{F}}[\vec{\boldsymbol{x}}]_{p_0^{\mathrm{F}}}}{\mathcal{P}^{\mathrm{F,ad}}[\vec{\boldsymbol{x}}]_{p_0^{\mathrm{F}}}} \\ &= \ln\frac{p^{\mathrm{F}}(\boldsymbol{x}(0),0)}{p^{\mathrm{F}}(\boldsymbol{x}(\tau),\tau)} + \ln\frac{\mathcal{P}^{\mathrm{F}}[\vec{\boldsymbol{x}}|\boldsymbol{x}(0)]}{\mathcal{P}^{\mathrm{R,ad}}[\vec{\boldsymbol{x}}^{\mathrm{R}}|\boldsymbol{x}^{\mathrm{R}}(0)]} + \ln\frac{\mathcal{P}^{\mathrm{F}}[\vec{\boldsymbol{x}}|\boldsymbol{x}(0)]}{\mathcal{P}^{\mathrm{F,ad}}[\vec{\boldsymbol{x}}|\boldsymbol{x}(0)]} \end{aligned} \tag{3.120}$$

where path probability functionals designated 'ad' are generated by the so-called adjoint or dual dynamics which produce the same stationary state as the normal dynamics, but with a reversed stationary flux. The justification of such a division is best illustrated in a discrete probability space where the transition

rates of such adjoint dynamics are given as

$$T^{\mathrm{ad}}(\boldsymbol{x}|\boldsymbol{x}',\lambda^{\mathrm{F}}(t)) = T(\boldsymbol{x}'|\boldsymbol{x},\lambda^{\mathrm{F}}(t))\frac{P^{\mathrm{F,st}}(\boldsymbol{x},\lambda^{\mathrm{F}}(t))}{P^{\mathrm{F,st}}(\boldsymbol{x}',\lambda^{\mathrm{F}}(t))} \tag{3.121}$$

and lead naturally to the above division when used in the construction of the total entropy production for a jump process consisting of N transitions, noting that the exponential integrated mean escape rates seen in Sect. 2.3.1 cancel [25, 36], to yield

$$\frac{\Delta\mathcal{S}_{\mathrm{tot}}}{k_B} = \ln\frac{P^{\mathrm{F}}(\boldsymbol{x}_0,0)}{P^{\mathrm{F}}(\boldsymbol{x}_N,\tau)} + \sum_{i=1}^{N}\ln\frac{P^{\mathrm{F,st}}(\boldsymbol{x}_i,\lambda^{\mathrm{F}}(t))}{P^{\mathrm{F,st}}(\boldsymbol{x}_{i-1},\lambda^{\mathrm{F}}(t))} + \sum_{i=1}^{N}\ln\frac{P^{\mathrm{F,st}}(\boldsymbol{x}_{i-1},\lambda^{\mathrm{F}}(t))T(\boldsymbol{x}_i|\boldsymbol{x}_{i-1},\lambda^{\mathrm{F}}(t))}{P^{\mathrm{F,st}}(\boldsymbol{x}_i,\lambda^{\mathrm{F}}(t))T(\boldsymbol{x}_{i-1}|\boldsymbol{x}_i,\lambda^{\mathrm{F}}(t))}, \tag{3.122}$$

with the first two terms comprising the so-called 'non-adiabatic' entropy production

$$\frac{\Delta\mathcal{S}_{\mathrm{na}}}{k_B} = \ln\frac{P^{\mathrm{F}}[\vec{\boldsymbol{x}}]_{P_0^{\mathrm{F}}}}{P^{\mathrm{R,ad}}[\vec{\boldsymbol{x}}^{\mathrm{R}}]_{P_\tau^{\mathrm{F}}}} = \ln\frac{P^{\mathrm{F}}(\boldsymbol{x}_0,0)}{P^{\mathrm{F}}(\boldsymbol{x}_N,\tau)} + \sum_{i=1}^{N}\ln\frac{P^{\mathrm{F,st}}(\boldsymbol{x}_i,\lambda^{\mathrm{F}}(t))}{P^{\mathrm{F,st}}(\boldsymbol{x}_{i-1},\lambda^{\mathrm{F}}(t))} \tag{3.123}$$

and the final term comprising the 'adiabatic' entropy production

$$\begin{aligned}\frac{\Delta\mathcal{S}_{\mathrm{a}}}{k_B} = \ln\frac{P^{\mathrm{F}}[\vec{\boldsymbol{x}}]_{P_0^{\mathrm{F}}}}{P^{\mathrm{F,ad}}[\vec{\boldsymbol{x}}]_{P_0^{\mathrm{F}}}} &= \ln\frac{P^{\mathrm{F}}(\boldsymbol{x}_0,0)}{P^{\mathrm{F}}(\boldsymbol{x}_0,0)} + \sum_{i=1}^{N}\ln\frac{P^{\mathrm{F,st}}(\boldsymbol{x}_{i-1},\lambda^{\mathrm{F}}(t))T(\boldsymbol{x}_i|\boldsymbol{x}_{i-1},\lambda^{\mathrm{F}}(t))}{P^{\mathrm{F,st}}(\boldsymbol{x}_i,\lambda^{\mathrm{F}}(t))T(\boldsymbol{x}_{i-1}|\boldsymbol{x}_i,\lambda^{\mathrm{F}}(t))} \\ &= \sum_{i=1}^{N}\ln\frac{P^{\mathrm{F,st}}(\boldsymbol{x}_{i-1},\lambda^{\mathrm{F}}(t))T(\boldsymbol{x}_i|\boldsymbol{x}_{i-1},\lambda^{\mathrm{F}}(t))}{P^{\mathrm{F,st}}(\boldsymbol{x}_i,\lambda^{\mathrm{F}}(t))T(\boldsymbol{x}_{i-1}|\boldsymbol{x}_i,\lambda^{\mathrm{F}}(t))}.\end{aligned} \tag{3.124}$$

Of note regarding their form is that the careful choice of boundary conditions enables both to comprise the total entropy production and thus considers such a division of entropy production (and heat flow) beyond that of transitions between steady states. This is evidenced by noting that non-adiabatic entropy production, like the argument of the Hatano-Sasa equality, comprises the excess heat and boundary terms, but the boundary terms represent the change in system entropy, defined generally, rather than the non-equilibrium potential, such that, for a thermal system, we have

$$\Delta\mathcal{S}_{\mathrm{na}} = \Delta\mathcal{S}_{\mathrm{sys}} + \frac{\Delta Q_{\mathrm{ex}}}{T}. \tag{3.125}$$

Further, since they have been constructed as functionals in the form found in Eq. (3.1) they naturally obey the appropriate fluctuation relations detailed earlier, with particular emphasis placed on the fact that they obey integral fluctuation theorems

$$\left\langle \exp\left[-k_B^{-1}\Delta\mathcal{S}_{\mathrm{na}}\right]\right\rangle_{P_0^{\mathrm{F}}}^{\mathrm{F}} = 1 \qquad\qquad \left\langle \exp\left[-k_B^{-1}\Delta\mathcal{S}_{\mathrm{a}}\right]\right\rangle_{P_0^{\mathrm{F}}}^{\mathrm{F}} = 1. \tag{3.126}$$

Consequently we have $\langle\Delta\mathcal{S}_{\mathrm{na}}\rangle_{P_0^{\mathrm{F}}}^{\mathrm{F}} \geq 0$, $\langle\Delta\mathcal{S}_{\mathrm{a}}\rangle_{P_0^{\mathrm{F}}}^{\mathrm{F}} \geq 0$ and $\langle\Delta\mathcal{S}_{\mathrm{tot}}\rangle_{P_0^{\mathrm{F}}}^{\mathrm{F}} = \langle\Delta\mathcal{S}_{\mathrm{na}}\rangle_{P_0^{\mathrm{F}}}^{\mathrm{F}} + \langle\Delta\mathcal{S}_{\mathrm{a}}\rangle_{P_0^{\mathrm{F}}}^{\mathrm{F}} \geq 0$ suggesting each portrays an aspect of irreversibility in a system which leads to a positive total entropy production.

The nature of the two contributions is easily understood by the consideration of some simple scenarios. Considering first the non-adiabatic contribution, one observes that it is in general non-zero except when $P^{\mathrm{F}}(\boldsymbol{x},0) = P^{\mathrm{F}}(\boldsymbol{x},\tau) = P^{\mathrm{F,st}}(\boldsymbol{x},\lambda(t))$ indicating that it is a contribution that arises when the system is out of stationarity. The adiabatic contribution, however, does not behave in this way and only vanishes

when the dynamics obey detailed balance

$$T(\boldsymbol{x}'|\boldsymbol{x},\lambda^{\mathrm{F}}(t))P^{\mathrm{F,st}}(\boldsymbol{x},\lambda^{\mathrm{F}}(t)) = T(\boldsymbol{x}|\boldsymbol{x}',\lambda^{\mathrm{F}}(t))P^{\mathrm{F,st}}(\boldsymbol{x}',\lambda^{\mathrm{F}}(t)), \tag{3.127}$$

the condition, in general, required for an equilibrium stationary state. The two mean positive contributions to total entropy production can then be summarised as a term relating to transient evolution of the probability distribution, vanishing in the stationary state and caused, physically, by a driving or a manipulation of the protocol and/or any instantaneous relaxation of the probability distribution and a term relating to the constant production of entropy in the stationary state caused by a breakage of detailed balance in the underlying dynamics.

Mean Contributions to Entropy Production in Continuous Systems

A result which helps to illustrate the nature of these contributions, beyond their microscopic forms, is that of their mean, or expected, contribution for which the form taken in continuous one dimensional systems as seen in [98] is most clear. First, we define the probability current or flux, $J(x,t)$, by writing the Fokker-Planck equation as a continuity equation

$$\frac{\partial p(x,t)}{\partial t} = -\frac{\partial J(x,t)}{\partial x}, \tag{3.128}$$

where we note the abbreviations $p(x,t) \equiv p^{\mathrm{F}}(x,t)$ and $p^{\mathrm{st}}(x,\lambda^{\mathrm{F}}(t)) \equiv p^{\mathrm{F,st}}(x,\lambda^{\mathrm{F}}(t))$ for brevity. We then proceed by starting with the mean entropy production rate for such systems, for which the form can be obtained by its identification with the time derivative of the generalised non-equilibrium Gibbs entropy [12],

$$\frac{1}{k_B}\frac{d\langle \Delta \mathcal{S}_{\mathrm{tot}}\rangle^{\mathrm{F}}_{p_0^{\mathrm{F}}}}{dt} = \int dx\ \frac{[J(x,t)]^2}{p(x,t)D}, \tag{3.129}$$

where D is the diffusion coefficient equal to half the second Kramers-Moyal coefficient. The division of the entropy production into an adiabatic and non-adiabatic contribution from such an expression was then provided in [98] by first writing

$$\begin{aligned}
\frac{1}{k_B}\frac{d\langle \Delta \mathcal{S}_{\mathrm{tot}}\rangle^{\mathrm{F}}_{p_0^{\mathrm{F}}}}{dt} &= \int dx\ \frac{p(x,t)}{D}\left(\frac{J(x,t)}{p(x,t)} - \frac{J^{\mathrm{st}}(x,\lambda^{\mathrm{F}}(t))}{p^{\mathrm{st}}(x,\lambda^{\mathrm{F}}(t))} + \frac{J^{\mathrm{st}}(x,\lambda^{\mathrm{F}}(t))}{p^{\mathrm{st}}(x,\lambda^{\mathrm{F}}(t))}\right)^2 \\
&= \int dx\ \frac{p(x,t)}{D}\left(\frac{J(x,t)}{p(x,t)} - \frac{J^{\mathrm{st}}(x,\lambda^{\mathrm{F}}(t))}{p^{\mathrm{st}}(x,\lambda^{\mathrm{F}}(t))}\right)^2 + \int dx\ \frac{p(x,t)}{D}\left(\frac{J^{\mathrm{st}}(x,\lambda^{\mathrm{F}}(t))}{p^{\mathrm{st}}(x,\lambda^{\mathrm{F}}(t))}\right)^2 \\
&\quad + \int dx\ \frac{2p(x,t)}{D}\frac{J^{\mathrm{st}}(x,\lambda^{\mathrm{F}}(t))}{p^{\mathrm{st}}(x,\lambda^{\mathrm{F}}(t))}\left(\frac{J(x,t)}{p(x,t)} - \frac{J^{\mathrm{st}}(x,\lambda^{\mathrm{F}}(t))}{p^{\mathrm{st}}(x,\lambda^{\mathrm{F}}(t))}\right).
\end{aligned} \tag{3.130}$$

Examining the final term we realise we may write it as

$$\begin{aligned}
\left(\frac{J(x,t)}{p(x,t)} - \frac{J^{\mathrm{st}}(x,\lambda^{\mathrm{F}}(t))}{p^{\mathrm{st}}(x,\lambda^{\mathrm{F}}(t))}\right) &= -D\left(\frac{1}{p(x,t)}\frac{\partial p(x,t)}{\partial x} - \frac{1}{p^{\mathrm{st}}(x,\lambda^{\mathrm{F}}(t))}\frac{\partial p^{\mathrm{st}}(x,\lambda^{\mathrm{F}}(t))}{\partial x}\right) \\
&= -D\frac{\partial}{\partial x}\left(\ln\frac{p(x,t)}{p^{\mathrm{st}}(x,\lambda^{\mathrm{F}}(t))}\right) \\
&= -D\frac{p^{\mathrm{st}}(x,\lambda^{\mathrm{F}}(t))}{p(x,t)}\frac{\partial}{\partial x}\left(\frac{p(x,t)}{p^{\mathrm{st}}(x,\lambda^{\mathrm{F}}(t))}\right).
\end{aligned} \tag{3.131}$$

Consequently the final term in Eq. (3.130) is given by

$$-\int dx\; 2J^{\rm st}(x,\lambda^{\rm F}(t))\frac{\partial}{\partial x}\left(\frac{p(x,t)}{p^{\rm st}(x,\lambda^{\rm F}(t))}\right)=0 \tag{3.132}$$

where the zero contribution arises due to vanishing of surface terms and the explicit condition

$$\frac{\partial J^{\rm st}(x,\lambda^{\rm F}(t))}{\partial x}=0. \tag{3.133}$$

Consequently, by the above construction, the first term in Eq. (3.130) must be the non-adiabatic contribution, vanishing uniquely in the stationary state rendering the second term the adiabatic contribution

$$\frac{1}{k_B}\frac{d\langle\Delta\mathcal{S}_{\rm a}\rangle^{\rm F}_{p_0^{\rm F}}}{dt}=\int dx\;\frac{p(x,t)}{D}\left(\frac{J^{\rm st}(x,t)}{p^{\rm st}(x,\lambda^{\rm F}(t))}\right)^2 \tag{3.134}$$

and

$$\frac{1}{k_B}\frac{d\langle\Delta\mathcal{S}_{\rm na}\rangle^{\rm F}_{p_0^{\rm F}}}{dt}=\int dx\;\frac{p(x,t)}{D}\left(\frac{J(x,t)}{p(x,t)}-\frac{J^{\rm st}(x,t)}{p^{\rm st}(x,\lambda^{\rm F}(t))}\right)^2 \tag{3.135}$$

each of which is rigorously positive and demonstrates the properties of each contribution. Examining further the non-adiabatic contribution we can write using Eq. (3.131)

$$\begin{aligned}\frac{1}{k_B}\frac{d\langle\Delta\mathcal{S}_{\rm na}\rangle^{\rm F}_{p_0^{\rm F}}}{dt}&=-\int dx\;p(x,t)\left(\frac{J(x,t)}{p(x,t)}-\frac{J^{\rm st}(x,t)}{p^{\rm st}(x,\lambda^{\rm F}(t))}\right)\frac{\partial}{\partial x}\left(\ln\frac{p(x,t)}{p^{\rm st}(x,\lambda^{\rm F}(t))}\right)\\&=-\left[p(x,t)\left(\frac{J(x,t)}{p(x,t)}-\frac{J^{\rm st}(x,t)}{p^{\rm st}(x,\lambda^{\rm F}(t))}\right)\left(\ln\frac{p(x,t)}{p^{\rm st}(x,\lambda^{\rm F}(t))}\right)\right]_{-\infty}^{\infty}\\&\quad+\int dx\;\left(\ln\frac{p(x,t)}{p^{\rm st}(x,\lambda^{\rm F}(t))}\right)\frac{\partial}{\partial x}\left(J(x,t)-\frac{J^{\rm st}(x,\lambda^{\rm F}(t))p(x,t)}{p^{\rm st}(x,\lambda^{\rm F}(t))}\right).\end{aligned} \tag{3.136}$$

Ignoring the surface terms we substitute using the continuity form of the Fokker-Planck equation and utilise Eq. (3.133) again to reach

$$\frac{1}{k_B}\frac{d\langle\Delta\mathcal{S}_{\rm na}\rangle^{\rm F}_{p_0^{\rm F}}}{dt}=\int dx\;\left(\ln\frac{p(x,t)}{p^{\rm st}(x,\lambda^{\rm F}(t))}\right)\left(-\frac{\partial p(x,t)}{\partial t}-J^{\rm st}(x,t)\frac{\partial}{\partial x}\frac{p(x,t)}{p^{\rm st}(x,\lambda^{\rm F}(t))}\right). \tag{3.137}$$

Taking the second term in this expression and integrating once more we find

$$\begin{aligned}&-\int dx\;\left(\ln\frac{p(x,t)}{p^{\rm st}(x,\lambda^{\rm F}(t))}\right)\left(J^{\rm st}(x,\lambda^{\rm F}(t))\frac{\partial}{\partial x}\frac{p(x,t)}{p^{\rm st}(x,\lambda^{\rm F}(t))}\right)\\&\qquad=\int dx\;\frac{p(x,t)}{p^{\rm st}(x,\lambda^{\rm F}(t))}\frac{\partial}{\partial x}\left(J^{\rm st}(x,t)\ln\frac{p(x,t)}{p^{\rm st}(x,\lambda^{\rm F}(t))}\right)\\&\qquad=\int dx\;J^{\rm st}(x,\lambda^{\rm F}(t))\frac{\partial}{\partial x}\left(\frac{p(x,t)}{p^{\rm st}(x,\lambda^{\rm F}(t))}\right)\end{aligned} \tag{3.138}$$

by the discarding of surface terms and application of Eq. (3.133). This contribution vanishes since it is of the same form as Eq. (3.132) and so we are finally left with

$$\frac{1}{k_B}\frac{d\langle\Delta\mathcal{S}_{\rm na}\rangle^{\rm F}_{p_0^{\rm F}}}{dt}=-\int dx\;\frac{\partial p(x,t)}{\partial t}\ln\frac{p(x,t)}{p^{\rm st}(x,\lambda^{\rm F}(t))} \tag{3.139}$$

which gives a concise description of the non-adiabatic contribution as the movement of the probability distribution towards the stationary solution. This will be the equilibrium distribution in the absence of

non-equilibrium conditions being applied such as non-conservative forces or temperature gradients. The above has been shown to apply quite generally with an equivalent quantity also being derived in the master equation approach [99].

3.6 Other Notable Theories of Entropy Production

3.6.1 Non-Thermal Divisions of Medium Entropy

The original division of entropy production into a system and a medium entropy change contribution is, in principle, defined quite generally, but some effort has been put into considering the physical origins of this contribution and the implication on the observed, or inferred, second law inequality. Such a situation arises when the medium entropy change contribution consists of a term which recognisably comprises the heat transfer to the environment as originally defined in stochastic energetics, but also a further contribution which cannot. Since it cannot be associated with a heat (or a defined temperature) such a contribution is a potentially athermal distinct mechanism for entropy production. A system investigated by Kim et al. [100, 101] was that of macromolecules described by a full phase space Langevin equation of the form

$$\begin{aligned}\frac{dx}{dt} &= v \\ \frac{dv}{dt} &= -\gamma v + \mathcal{F}_v(v) + \sqrt{\frac{2k_BT\gamma}{m}}\Gamma(t).\end{aligned} \tag{3.140}$$

With the inclusion of an extra velocity dependent force $\mathcal{F}_v(v)$ (which they mention in an example could be of the form $\gamma' v$ where γ' depends on a magnetic field) they found that what we have called the medium entropy change contribution, in the mean, contained the usual heat transfer term, which they went on to associate with the medium entropy change of others [12], but also a term based on any non-linearity of $\mathcal{F}_v(v)$ which they called the 'pumped' entropy contribution as it is a term that would, in addition to the stochastic thermostat, adjust the effective temperature of the particle. The term they observed had the form

$$\Delta\mathcal{S}_{\mathrm{pu}} = -\int_0^\tau \frac{\partial\mathcal{F}_v(v)}{\partial v}dt \tag{3.141}$$

and its inclusion led them to write the division of entropy as

$$\begin{aligned}\Delta\mathcal{S}_{\mathrm{pos}} &= \Delta\mathcal{S}_{\mathrm{sys}} + \Delta\mathcal{S}_{\mathrm{med}} + \Delta\mathcal{S}_{\mathrm{pu}} \\ &= \Delta\mathcal{S}_{\mathrm{tot}} + \Delta\mathcal{S}_{\mathrm{pu}}\end{aligned} \tag{3.142}$$

where $\Delta\mathcal{S}_{\mathrm{pos}}$, denoting a rigorously positive contribution in the mean, obeys the fluctuation theorem rather than $\Delta\mathcal{S}_{\mathrm{tot}}$. Whether or not such an procedure is particularly helpful is debatable, and is something we discuss in a later chapter along with assumptions on the function $\mathcal{F}_v(v)$ that cause its contribution to be considered in such a way.

3.6.2 Feedback Control

The division of entropy production advocated by Kim et al. served as a precursor to a more general recent development, namely that of feedback control [102–109]. This is where one controls the protocol in response to the observed behaviour of the system. Such a procedure is very similar to the action of Maxwell's demon, which for example, opens or closes a trapdoor based on the velocity of an observed

particle. However, in the language we have been using, it is best understood as a protocol which in turn is a, possibly stochastic, function of $\boldsymbol{x}$. Leaving aside, for now, some of the possible ambiguities, the general relations that occur in the literature are simple to define. Given, for sake of argument, that the protocol, starting at some pre-determined value $\lambda^{\mathrm{F}}(0)$, is determined by measurement so that we have $\lambda^{\mathrm{F}}(\boldsymbol{x}(t))$, we acknowledge that the protocol now effectively becomes a stochastic trajectory in its own right, $\vec{\lambda}^{\mathrm{F}}$. As such, the probability functional associated with the forward behaviour is given by the joint probability functional

$$\mathcal{P}^{\vec{\lambda}^{\mathrm{F}}}[\vec{\boldsymbol{x}}, \vec{\lambda}^{\mathrm{F}}]_{p_0^{\mathrm{F}}} = \mathcal{P}[\vec{\lambda}^{\mathrm{F}}|\vec{\boldsymbol{x}}]\mathcal{P}^{\vec{\lambda}^{\mathrm{F}}}[\vec{\boldsymbol{x}}]_{p_0^{\mathrm{F}}}. \tag{3.143}$$

Despite the stochasticity of the protocol, it is however, deemed to be environmental or perhaps under the control of some external agent so that the reverse protocol is not determined stochastically, but instead is the reversed version of the observed forward protocol. Under such constraints we thus write the joint probability functional of the reverse sequence

$$\mathcal{P}^{\vec{\lambda}^{\mathrm{R}}}[\vec{\boldsymbol{x}}^{\mathrm{R}}, \vec{\lambda}^{\mathrm{R}}]_{p_\tau^{\mathrm{F}}} = \mathcal{P}[\vec{\boldsymbol{x}}^{\mathrm{R}}|\vec{\lambda}^{\mathrm{R}}]_{p_\tau^{\mathrm{F}}}\mathcal{P}[\vec{\lambda}^{\mathrm{R}}]. \tag{3.144}$$

But since the reverse protocol is now deterministic we may write

$$\mathcal{P}^{\vec{\lambda}^{\mathrm{R}}}[\vec{\boldsymbol{x}}^{\mathrm{R}}, \vec{\lambda}^{\mathrm{R}}]_{p_\tau^{\mathrm{F}}} = \mathcal{P}^{\vec{\lambda}^{\mathrm{R}}}[\vec{\boldsymbol{x}}^{\mathrm{R}}]_{p_\tau^{\mathrm{F}}}\mathcal{P}[\vec{\lambda}^{\mathrm{F}}]. \tag{3.145}$$

When we then come to construct the total entropy production we thus find

$$\begin{aligned} k_B^{-1}\Delta\mathcal{S}_{\mathrm{tot}}[\vec{\boldsymbol{x}}, \vec{\lambda}^{\mathrm{F}}] &= \ln \mathcal{P}^{\vec{\lambda}^{\mathrm{F}}}[\vec{\boldsymbol{x}}, \vec{\lambda}^{\mathrm{F}}]_{p_0^{\mathrm{F}}} - \ln \mathcal{P}^{\vec{\lambda}^{\mathrm{R}}}[\vec{\boldsymbol{x}}^{\mathrm{R}}, \vec{\lambda}^{\mathrm{R}}]_{p_\tau^{\mathrm{F}}} \\ &= \ln \frac{\mathcal{P}^{\vec{\lambda}^{\mathrm{F}}}[\vec{\boldsymbol{x}}]_{p_0^{\mathrm{F}}}}{\mathcal{P}^{\vec{\lambda}^{\mathrm{R}}}[\vec{\boldsymbol{x}}^{\mathrm{R}}]_{p_\tau^{\mathrm{F}}}} + \ln \frac{\mathcal{P}[\vec{\lambda}^{\mathrm{F}}|\vec{\boldsymbol{x}}]}{\mathcal{P}[\vec{\lambda}^{\mathrm{F}}]}. \end{aligned} \tag{3.146}$$

Once again, a similar procedure is then used as in [100, 101], as the first term, despite being a constituent part of the functional representing irreversibility is denoted $\Delta\mathcal{S}_{\mathrm{tot}}$ and the remainder is identified as the (single shot) mutual information, $\mathcal{I}$, of the observation(s) that determined the forward protocol. As such one finds an IFT for the sum of the two as in [104]

$$\langle \exp[-k_B^{-1}(\Delta\mathcal{S}_{\mathrm{tot}} + \mathcal{I})] \rangle_{p_0^{\mathrm{F}}}^{\mathrm{F}} = 1 \tag{3.147}$$

or for equilibrium initial distributions [102]

$$\langle \exp[-((\Delta W - \Delta F)/k_B T + k_B^{-1}\mathcal{I})] \rangle_{p_0^{\mathrm{F,eq}}}^{\mathrm{F}} = 1 \tag{3.148}$$

revealing bounds on the 'failure' of the second law given the information gained from measurement. For example, the latter suggests with measurement one can obtain free energy difference, on average, with a smaller expenditure of work. Of course, the second law is not truly failing: what is now being written as the total entropy production is no longer the total entropy production of the universe. We do note that when dealing with feedback there is inherent ambiguity surrounding how that quantity can be defined which is discussed in a later chapter.

3.7 Time Reversal Symmetry in Stochastic Systems

Considering the division of entropy production into an adiabatic and non-adiabatic contribution by van den Broeck et al. [36, 98, 99] it is helpful to establish more concretely the conditions under which we can expect these two contributions to arise. Of these two contributions it is arguable that the non-adiabatic term is most intuitive with it arising explicitly when the system is driven and with its mean being controlled by changes in the probability distribution which we can immediately connect with a measure of irreversibility. Indeed if the entropy production is solely non-adiabatic the entropy production is simply given by a sum over logarithms of ratios of the canonical distribution over increments in the driving protocol $\lambda^{\mathrm{F}}(t)$ and thus one can simply invoke the arguments of Crooks [8] to form the entropy functional by way of detailed balance. The stochastic entropy, however, has a much wider range of applicability with both integral and detailed fluctuation theorems having validity even when detailed balance is broken. This wider range of applicability explicitly concerns the qualitatively separable adiabatic contribution. We can see simply from its form in Eq. (3.124) that it arises when the number of transitions between two states doesn't balance even when the distribution is not changing with time. This is the definition of broken detailed balance and is one of the defining features of a class of systems which forms non-equilibrium steady states. These systems are important in the study of non-equilibrium phenomena as they can be considered to be one of the most simple extensions beyond normal equilibrium thermodynamics allowing us to probe the nature of matter out of equilibrium. Indeed it is for steady states that the fluctuation theorems were originally identified and for which detailed fluctuation theorems are valid over any period of time. In order to understand the adiabatic contribution we must therefore have an understanding of detailed balance and how it features in the steady states of stochastic systems.

3.7.1 Steady States and Detailed Balance

In considering the behaviour of a system in the steady state one must consider the conditions required in order to achieve such a distribution, that is one which is time invariant. This is most easily described by means of a master equation which has discrete states. In order for the probability to be in a given state to be constant we must require that, on average, there are as many transitions into that state as there are leaving it. This can be written down quite simply by the condition

$$\sum_x P(x)T(x'|x) = \sum_x P(x')T(x|x'). \tag{3.149}$$

When this condition is met for all x' then the system will reach the steady state $P^{\mathrm{st}}(x')$ and the system is said to be balanced. Detailed balance however, is a much more stringent condition which requires every possible transition to be balanced in the same way the sum of all transitions was balanced above. Consequently we have for detailed balance the condition

$$P^{\mathrm{st}}(x)T(x'|x) = P^{\mathrm{st}}(x')T(x|x') \tag{3.150}$$

for all x and x'. When this condition is met every possible transition is balanced so that not only is the distribution time invariant, but there is also no flow of probability anywhere in the system. Considering now the Fokker-Planck equation for an over-damped particle in the form of a probability current J

$$\frac{\partial p(x,t)}{\partial t} = -\frac{\partial J(x,t)}{\partial x}, \tag{3.151}$$

we consequently identify the balance condition as being equivalent to

$$\frac{\partial p(x,t)}{\partial t} = 0 \qquad\qquad J(x,t) = \text{const} \tag{3.152}$$

and detailed balance as being equivalent to

$$\frac{\partial p(x,t)}{\partial t} = 0 \qquad\qquad J(x,t) = 0 \tag{3.153}$$

so that we understand that detailed balance ensures balance, but balance does not ensure detailed balance. The above is how detailed balance is usually defined for most stochastic systems, however we point out that its definition should be slightly more general and is based on the fact that some system coordinates may be odd with respect to time reversal. For example if one is describing both the position and velocity of a particle it is impossible to balance the transition from a positive position and velocity to some other positive position and velocity since normal dynamics will not permit negative positional steps to result from positive velocities. To make a physically meaningful interpretation we must reverse the sign of the odd, velocity variables and thus have the definition of detailed balance as

$$p^{\text{st}}(x,v)p(x',v',\tau|x,v,0) = p^{\text{st}}(x',-v')p(x,-v,\tau|x',-v',0). \tag{3.154}$$

In contrast, the position (and all other quantities we have discussed) do not change their sign and are thus described as being even with respect to time reversal. The most widely used notation for this time reversal is to consider the quantity $\varepsilon_i x_i$ as the time reversal of the general coordinate x_i where ε_i is $+1$ for even coordinates and -1 for odd. As such we write more generally, for a system which depends on many coordinates both odd and even $\boldsymbol{x} = x_1, x_2, x_3 \ldots$, the expression

$$p^{\text{st}}(\boldsymbol{x})p(\boldsymbol{x}',\tau|\boldsymbol{x},0) = p^{\text{st}}(\boldsymbol{\varepsilon x}')p(\boldsymbol{\varepsilon x},\tau|\boldsymbol{\varepsilon x}',0). \tag{3.155}$$

Specifically for the case of a Markov process we can make a simplification by considering $\tau \to 0$. In this case the conditional probabilities reduce to delta functions so that $p^{\text{st}}(\boldsymbol{\varepsilon x}')\delta(\boldsymbol{\varepsilon x}' - \boldsymbol{\varepsilon x}) = p^{\text{st}}(\boldsymbol{x})\delta(\boldsymbol{x} - \boldsymbol{x}')$. By the symmetry of the delta function we therefore find $p^{\text{st}}(\boldsymbol{\varepsilon x}') = p^{\text{st}}(\boldsymbol{x})$ when $\boldsymbol{x} = \boldsymbol{x}'$ meaning we can write

$$p^{\text{st}}(\boldsymbol{x})p(\boldsymbol{x}',\tau|\boldsymbol{x},0) = p^{\text{st}}(\boldsymbol{x}')p(\boldsymbol{\varepsilon x},\tau|\boldsymbol{\varepsilon x}',0). \tag{3.156}$$

In the context of Fokker-Planck equations there exist well defined conditions for identifying whether a system obeys detailed balance with them commonly referred to as 'potential conditions' [38, 46]. For the case of the general multidimensional Fokker-Planck equation

$$\frac{\partial p(\boldsymbol{x},t)}{\partial t} = -\sum_i \frac{\partial}{\partial x_i}\left[A_i(\boldsymbol{x})p(\boldsymbol{x},t)\right] + \frac{1}{2}\sum_{i,j}\frac{\partial^2}{\partial x_i \partial x_j}\left[B_{ij}(\boldsymbol{x})p(\boldsymbol{x},t)\right] \tag{3.157}$$

the conditions are

$$\varepsilon_i A_i(\boldsymbol{\varepsilon x})p^{\text{st}}(\boldsymbol{x}) = -A_i(\boldsymbol{x})p^{\text{st}}(\boldsymbol{x}) + \sum_j \frac{\partial}{\partial x_j}\left[B_{ij}p^{\text{st}}(\boldsymbol{x})\right] \tag{3.158}$$

and

$$\varepsilon_i \varepsilon_j B_{ij}(\boldsymbol{\varepsilon x}) = B_{ij}(\boldsymbol{x}). \tag{3.159}$$

For the case of even variables only, in one dimension, applicable to the simple over-damped Brownian motion, these conditions reduce to the simple single constraint

$$A(x)p^{\mathrm{st}}(x) = \frac{1}{2}\frac{\partial}{\partial x}\left[B(x)p^{\mathrm{st}}(x)\right]. \tag{3.160}$$

In contrast to the over-damped case, detailed balance, in general, is not the requirement for zero stationary current, but a more complicated requirement based on a division of the current which shall be addressed in a later chapter. Since the definition of detailed balance, as defined for even variables, is central to the division of the entropy production seen in Sect. 3.5.1, and the thermodynamics that it describes [35], it is this generalisation, such that it includes odd variables, and its consequences that we go on to develop in subsequent chapters.

We briefly mention that from the potential conditions we realise that we cannot identify time reversal invariance until we have the stationary solution which is obtained only after the necessary specification of boundary conditions. If we consider natural, that is reflective, boundary conditions at infinity it is trivial to show that the stationary solution for the one dimensional case with even variables is

$$p^{\mathrm{st}}(x) = \frac{2\mathcal{N}}{B(x)}\exp\left[\int^{x} dx' \, \frac{2A(x')}{B(x')}\right] \tag{3.161}$$

where $\mathcal{N}$ is a normalisation constant. This is entirely equivalent to the condition in Eq. (3.160) and is independent of the form of A or B. Since we can generally associate the drift A with the force on the particle and B with the diffusion parameter or temperature we can assert that this form holds whether the force upon the particle derives from a defined potential $V(x)$ or a non-conservative force $\mathcal{F}_{nc}$ or even when the temperature is spatially varying. Consequently for such a system we cannot obtain a non-zero probability current and therefore any time reversal symmetry breaking in the stationary state. This has a simple geometric cause due to the impossibility of creating stationary current on a line.

If however, we consider a system with periodic boundary conditions, for example diffusion on a circle, stationary current is possible. Taking the most trivial example consisting of diffusion on a homogeneous circle in a uniform potential subject to a constant non-conservative force the stationary solution is trivially $p^{\mathrm{st}}(x) = \text{const}$ with the constant determined by normalisation. After substitution we find that for detailed balance to hold we must require $\mathcal{F}_{nc} = 0$, that is under no direct forcing. This result is a direct consequence of having a non-conservative force. For such effects the solution is not simply integrable as there is no defined potential from which the forces arise (the work performed by the force is dependent on the number of loops of the circle the particle performs) and so the solution cannot be expressed in the form of Eq. (3.161). As such, the set up of periodic boundary conditions in addition to a non-conservative force will serve as a paradigmatic procedure for creating non-equilibrium stationary states and to explore the nature of the non-equilibrium thermodynamics.

Chapter 4

Entropy Production and its Constituent Contributions for Systems with Odd and Even Time Reversal Behaviour in Discrete Systems

As the central result of the present work we consider the form of the total entropy production and its division into relevant thermodynamic quantities for stochastic systems which concern the evolution and/or are described by coordinates and parameters which may transform differently under time reversal. Physically, quantities may be described as having a defined parity and are deemed odd or even depending on whether they change their sign or leave their sign invariant. As such, the subsequent development is essential for the consideration of stochastic behaviour in systems which describe, or are described, by velocities, torques, magnetic moments & fields and current densities to name a few.

In order to treat odd variables in stochastic systems as generally as possible we give an account using discrete dynamics described by a master equation in the knowledge that one may readily represent continuous stochastic behaviour as the limit of discrete dynamics, but not necessarily the other way round [38]. It may be argued that modelling odd variables in such a way is unnatural, an issue we believe is unimportant in general for arbitrary model systems as discussed in a later chapter, but point out, for example, that the Ising model, which deals with magnetic moments, strictly is a discrete model that concerns odd variables[1]. We also mention that the work in [110], appearing around the time the present work was reported, considered odd discrete variables, though with a rather different approach and without the subsequent consideration of the steady state thermodynamics. To proceed we consider the dynamics of a general set of variables $\boldsymbol{x} = (x_1, x_2, \ldots x_n)$ that transform differently under the time reversal operator, $\boldsymbol{\varepsilon}$, such that $\boldsymbol{\varepsilon x} = (\varepsilon_1 x_1, \varepsilon_2 x_2, \ldots \varepsilon_n x_n)$ where $\varepsilon_i x_i$ is an involution such that $\varepsilon_i \varepsilon_i x_i = x_i$, stressing that $\boldsymbol{\varepsilon}$ performs the time reversal operation on all variables and is not a vector of operators ε_i. For even variables we consider $\varepsilon_i = 1$, whilst typically for odd variables we consider

[1] Though we note that due to the presence of the odd applied magnetic field appearing in the time reversal invariant Hamiltonian, treating all quantities (including the magnetic field) as even is equivalent.

$\varepsilon_i = -1$, but note that the subsequent development, mathematically at least, is valid for any arbitrary transformation that leaves the summation over paths unchanged. We recall that the entropy production of a path of duration τ depends on two relative likelihoods of specified paths which here, in discrete space and continuous time, are described by path probability densities. The first, $p^{\mathrm{F}}[\vec{\boldsymbol{x}}]_{P_0^{\mathrm{F}}}$, is defined as the probability density of observing a forward trajectory, $\vec{\boldsymbol{x}} = \boldsymbol{x}(t)$ for $0 \leq t \leq \tau$ with a probability distribution of starting configurations, $P^{\mathrm{F}}(\boldsymbol{x}(0), 0)$, that acts as an initial condition for the appropriate master equation with matrix of transition rates, $T(\boldsymbol{x}|\boldsymbol{x}', \lambda^{\mathrm{F}}(t))$, where we consider the time dependence and functional dependence on $\boldsymbol{x}$ and $\boldsymbol{x}'$ to follow the forward protocol $\lambda^{\mathrm{F}}(t)$. The path probability density of some sequence of N transitions to configuration $\boldsymbol{x}_i$ from $\boldsymbol{x}_{i-1}$ at times t_i, such that $t_0 = 0$ and $t_{N+1} = \tau$, can then be computed as a function of transition rates and exponential waiting times according to Eq. (2.116) which we write explicitly here as

$$p^{\mathrm{F}}[\vec{\boldsymbol{x}}]_{P_0^{\mathrm{F}}} = P^{\mathrm{F}}(\boldsymbol{x}_0, 0) e^{\int_{t_0}^{t_1} dt T(\boldsymbol{x}_0|\boldsymbol{x}_0, \lambda^{\mathrm{F}}(t))} \prod_{i=1}^{N} T(\boldsymbol{x}_i|\boldsymbol{x}_{i-1}, \lambda^{\mathrm{F}}(t_i)) e^{\int_{t_i}^{t_{i+1}} dt T(\boldsymbol{x}_i|\boldsymbol{x}_i, \lambda^{\mathrm{F}}(t))}. \tag{4.1}$$

We know, from Sect. 3.1, that the form of quantities that follow the transient fluctuation theorem, with which we associate entropy production are based on the comparison of this path probability density to that for another trajectory $\vec{\boldsymbol{x}}^*$, protocol λ^*, dynamics and initial condition $P^*(\boldsymbol{x}^*(0), 0)$. For discrete systems this is written

$$\mathcal{A}[\vec{\boldsymbol{x}}] = \ln \left[p^{\mathrm{F}}[\vec{\boldsymbol{x}}]_{P_0^{\mathrm{F}}} / p^*[\vec{\boldsymbol{x}}^*]_{P_0^*} \right], \tag{4.2}$$

where here k_B is taken as unity so that such quantities can describe athermal systems noting that all quantities can be premultiplied by k_B when dealing with thermal systems. Again, this quantity will follow the various fluctuation theorems if there is a one to one mapping between $\vec{\boldsymbol{x}}$ and $\vec{\boldsymbol{x}}^*$ so that we can consider the summation over $\vec{\boldsymbol{x}}^*$ to be equivalent to that over $\vec{\boldsymbol{x}}$ and if we have ergodic consistency such that the bounds and region of the summation contains all the possible paths in both sets of dynamics. Once again we point out that in general we consider this to be the requirement $p^*[\vec{\boldsymbol{x}}^*] = 0$ for all $p^{\mathrm{F}}[\vec{\boldsymbol{x}}] = 0$ and vice versa rather than just specifying that initial distributions are nowhere zero. Finally we recall the implication $\langle \mathcal{A}[\vec{\boldsymbol{x}}] \rangle_{P_0^{\mathrm{F}}}^{\mathrm{F}} \geq 0$ by Jensen's inequality.

4.1 Expression for Total Entropy Production

We now proceed by specifying certain choices of dynamics, protocol, path and initial condition which lead to various contributions to total entropy production. We shall argue that despite this seemingly large choice in the specification of the compared path probability density, physically relevant quantities are based on only two specifications, namely a single (involutive so as to be physically meaningful in this context) transformation and a specification, or adaptation of, the dynamics, but stress that the two choices cannot be made independently. This second point is related to further specification of the ergodic consistency requirement considered earlier and is based on the fact that given a transition $\boldsymbol{x} \to \boldsymbol{x}'$ under the normal, forward dynamics, the transition $\boldsymbol{x}^* \to \boldsymbol{x}^{*\prime}$ is not, in general, possible under those same dynamics. For example, if the forward dynamics concern position and velocity and behave like Hamiltonian dynamics, if we choose $\vec{\boldsymbol{x}}^* = \vec{\boldsymbol{x}}^{\mathrm{R}}$, those normal, forward dynamics that produced $\vec{\boldsymbol{x}}$ cannot also produce $\vec{\boldsymbol{x}}^*$ since a negative positional step cannot arise whilst the momentum is positive. As such if the chosen transformation corresponds to a transition which might be forbidden under the normal dynamics, one is obliged to choose dynamics under which such a transition is possible.

Such a consideration is also necessary when defining the total entropy production. Since we consider the entropy production to be a measure of the irreversibility of the process we evaluate the reverse path probability density which is defined using the same dynamics as the forward path. As such we consider an involution which produces a path which is a solution of the normal dynamics. Such an involution is the choice $f(t) \to f^\dagger(t) \equiv \boldsymbol{\varepsilon} f(\tau - t)$ and naturally leads to the reverse path $\boldsymbol{x}^*(t) = \boldsymbol{x}^\dagger(t) = \boldsymbol{\varepsilon}\boldsymbol{x}(\tau - t)$ which, for position and velocity variables for example, is equal to the retracing of the forward sequences of positions and velocities, but with the velocities reversed. Next we argue that we need only perform this involution on the protocol and initial condition to fully and consistently specify the compared path probability density. We define the reversed protocol $\lambda^*(t) = \lambda^\dagger(t) = \varepsilon\lambda^{\mathrm{F}}(\tau - t)$ which amounts to the replaying of any external switching protocol controlling the dynamics in reverse sequence as we have already seen, but also with an instantaneous time reversal of any odd terms (for example a magnetic field). Since in such a master equation approach the dynamics are specified by transition rates, rather than explicitly on physical variables with a defined parity, the notation $T(\boldsymbol{x}'|\boldsymbol{x}, \varepsilon\lambda(t))$ is taken to represent an alternative (arbitrary, so long as the same allowed transitions in $T(\boldsymbol{x}'|\boldsymbol{x}, \lambda(t))$ are possible) set of transition rates which we interpret as the time reversed transition rates. Finally we apply this involution to the dynamically evolving solution to the master equation which appears in the forward path probability density as the initial condition of the forward dynamics to find the appropriate choice of initial condition for the reverse path. This then allows an alternative argument for the choice of boundary terms in the path probability density beyond that of appealing to a deconstruction of the Gibb's entropy. To do so we consider the involution applied to the solution of the master equation so that we have $P^*(\boldsymbol{x}^*(t), t) = P^\dagger(\boldsymbol{x}^*(t), t) = \left[P^{\mathrm{F}}(\boldsymbol{x}^*(t), t)\right]^\dagger$ which involves a transformation of the functional form of the distribution according to its time reversal parity and evaluation at time $\tau - t$ which we write $\hat{\boldsymbol{\varepsilon}} P^{\mathrm{F}}(\boldsymbol{x}^*(t), \tau - t) = P^{\mathrm{F}}(\boldsymbol{\varepsilon}\boldsymbol{x}^*(t), \tau - t)$. However, since it is the initial condition for the alternative path probability density we have $t = 0$ and $\boldsymbol{x}^*(0) = \boldsymbol{x}^\dagger(0) = \boldsymbol{\varepsilon}\boldsymbol{x}(\tau)$ so that we may write the initial condition $P^*(x^*(0), 0) = P^{\mathrm{F}}(\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}\boldsymbol{x}(\tau), \tau) = P^{\mathrm{F}}(\boldsymbol{x}(\tau), \tau)$. Explicitly, the consequence of this property is the relation $p^\dagger[\vec{\boldsymbol{x}}^\dagger]_{P_0^\dagger} = p^\dagger[\vec{\boldsymbol{x}}^\dagger]_{\hat{\boldsymbol{\varepsilon}} P_\tau^{\mathrm{F}}}$ and leads to the identification of the system entropy as defined previously. This time reversal of the distribution is not usually mentioned, presumably from the relative lack of focus on systems with odd variables, but should be included as a component of the heuristic description of entropy production as the relative likelihood of observing the reverse path. We argue that since such definitions follow from a time reversal of the path and protocol, this naturally should extend to the initial condition in the same way. Constructing the time reversed path probability density we write

$$p^\dagger[\vec{\boldsymbol{x}}^\dagger]_{P_0^\dagger} = P^\dagger(\boldsymbol{x}_0^\dagger, 0) e^{\int_{t_0}^{t_1} dt T(\boldsymbol{x}_0^\dagger|\boldsymbol{x}_0^\dagger, \lambda^\dagger(t))} \prod_{i=1}^{N} T(\boldsymbol{x}_i^\dagger|\boldsymbol{x}_{i-1}^\dagger, \lambda^\dagger(t_i)) e^{\int_{t_i}^{t_{i+1}} dt T(\boldsymbol{x}_i^\dagger|\boldsymbol{x}_i^\dagger, \lambda^\dagger(t))}. \tag{4.3}$$

We have $p^\dagger[\vec{\boldsymbol{x}}^\dagger]_{P_0^\dagger} = p^\dagger[\vec{\boldsymbol{x}}^\dagger]_{\hat{\boldsymbol{\varepsilon}} P_\tau^{\mathrm{F}}}$ and $\boldsymbol{x}_i^\dagger = \boldsymbol{\varepsilon}\boldsymbol{x}_{N-i}$ so we may rearrange

$$\begin{aligned} p^\dagger[\vec{\boldsymbol{x}}^\dagger]_{\hat{\boldsymbol{\varepsilon}} P_\tau^{\mathrm{F}}} = {} & P^{\mathrm{F}}(\boldsymbol{x}_N, \tau) e^{\int_{t_N}^{t_{N+1}} dt T(\boldsymbol{\varepsilon}\boldsymbol{x}_0|\boldsymbol{\varepsilon}\boldsymbol{x}_0, \lambda^\dagger(t))} \\ & \times \prod_{i=1}^{N} e^{\int_{t_{N-i}}^{t_{N-i+1}} dt T(\boldsymbol{\varepsilon}\boldsymbol{x}_i|\boldsymbol{\varepsilon}\boldsymbol{x}_i, \lambda^\dagger(t))} T(\boldsymbol{\varepsilon}\boldsymbol{x}_{i-1}|\boldsymbol{\varepsilon}\boldsymbol{x}_i, \lambda^\dagger(t_{N-i+1})). \end{aligned} \tag{4.4}$$

We then perform a change of variable $t \to \tau - t$ and use $\lambda^\dagger(t_i) = \varepsilon\lambda^{\mathrm{F}}(t_{N-i+1})$ such that

$$p^\dagger[\vec{\boldsymbol{x}}^\dagger]_{\hat{\boldsymbol{\varepsilon}}P_\tau^{\mathrm{F}}} = P^{\mathrm{F}}(\boldsymbol{x}_N,\tau)e^{-\int_{t_1}^{t_0} dtT(\boldsymbol{\varepsilon x}_0|\boldsymbol{\varepsilon x}_0,\varepsilon\lambda^{\mathrm{F}}(t))} \times \prod_{i=1}^{N} e^{-\int_{t_{i+1}}^{t_i} dtT(\boldsymbol{\varepsilon x}_i|\boldsymbol{\varepsilon x}_i,\varepsilon\lambda^{\mathrm{F}}(t))} T(\boldsymbol{\varepsilon x}_{i-1}|\boldsymbol{\varepsilon x}_i,\varepsilon\lambda^{\mathrm{F}}(t_i)). \tag{4.5}$$

A comparison of $p^{\mathrm{F}}[\vec{\boldsymbol{x}}]_{P_0^{\mathrm{F}}}$ and $p^\dagger[\vec{\boldsymbol{x}}^\dagger]_{\hat{\boldsymbol{\varepsilon}}P_\tau^{\mathrm{F}}}$ characterises the irreversibility of the forward path and defines the total entropy production (using units $k_B = 1$)

$$\begin{aligned}
\Delta\mathcal{S}_{\mathrm{tot}} &= \ln p^{\mathrm{F}}[\vec{\boldsymbol{x}}]_{P_0^{\mathrm{F}}} - \ln p^\dagger[\vec{\boldsymbol{x}}^\dagger]_{P_0^\dagger} \\
&= \ln\frac{P^{\mathrm{F}}(\boldsymbol{x}_0,0)}{P^{\mathrm{F}}(\boldsymbol{x}_N,\tau)} + \sum_{i=0}^{N}\ln\frac{e^{\int_{t_i}^{t_{i+1}} dt\ T(\boldsymbol{x}_i|\boldsymbol{x}_i,\lambda^{\mathrm{F}}(t))}}{e^{\int_{t_i}^{t_{i+1}} dt\ T(\boldsymbol{\varepsilon x}_i|\boldsymbol{\varepsilon x}_i,\varepsilon\lambda^{\mathrm{F}}(t))}} + \sum_{i=1}^{N}\ln\frac{T(\boldsymbol{x}_i|\boldsymbol{x}_{i-1},\lambda^{\mathrm{F}}(t_i))}{T(\boldsymbol{\varepsilon x}_{i-1}|\boldsymbol{\varepsilon x}_i,\varepsilon\lambda^{\mathrm{F}}(t_i))} \\
&= \ln\frac{P^{\mathrm{F}}(\boldsymbol{x}_0,0)}{P^{\mathrm{F}}(\boldsymbol{x}_N,\tau)} + \int_0^\tau dt\left(T(\boldsymbol{x}(t)|\boldsymbol{x}(t),\lambda^{\mathrm{F}}(t)) - T(\boldsymbol{\varepsilon x}(t)|\boldsymbol{\varepsilon x}(t),\varepsilon\lambda^{\mathrm{F}}(t))\right) \\
&\quad + \sum_{i=1}^{N}\ln\frac{T(\boldsymbol{x}_i|\boldsymbol{x}_{i-1},\lambda^{\mathrm{F}}(t_i))}{T(\boldsymbol{\varepsilon x}_{i-1}|\boldsymbol{\varepsilon x}_i,\varepsilon\lambda^{\mathrm{F}}(t_i))}.
\end{aligned} \tag{4.6}$$

Recognising the structure of its form, identifying $\mathcal{A}[\vec{\boldsymbol{x}}] = \Delta\mathcal{S}_{\mathrm{tot}}$ and $\hat{\mathcal{A}}[\vec{\boldsymbol{x}}^\dagger] = \Delta\hat{\mathcal{S}}_{\mathrm{tot}}$, we thus find it obeys [12]

$$\langle\exp\left[-\Delta\mathcal{S}_{\mathrm{tot}}\right]\rangle_{P_0^{\mathrm{F}}}^{\mathrm{F}} = 1, \tag{4.7}$$

$$p^\dagger(\Delta\hat{\mathcal{S}}_{\mathrm{tot}} = -A)_{\hat{\boldsymbol{\varepsilon}}P_\tau^{\mathrm{F}}} = e^{-A}p^{\mathrm{F}}(\Delta\mathcal{S}_{\mathrm{tot}} = A)_{P_0^{\mathrm{F}}} \tag{4.8}$$

and

$$\langle g(\mathcal{C}[\vec{\boldsymbol{x}}^\dagger])\rangle_{\hat{\boldsymbol{\varepsilon}}P_\tau^{\mathrm{F}}}^\dagger = \langle g(\mathcal{B}[\vec{\boldsymbol{x}}])e^{-\Delta\mathcal{S}_{\mathrm{tot}}}\rangle_{P_0^{\mathrm{F}}}^{\mathrm{F}} \tag{4.9}$$

recalling g is an arbitrary function and the two new functionals are related by $\mathcal{B}[\vec{\boldsymbol{x}}] = \mathcal{C}[\vec{\boldsymbol{x}}^\dagger]$. The form of $\Delta\mathcal{S}_{\mathrm{tot}}$ is more complicated than previous descriptions [25, 36] unless $\boldsymbol{\varepsilon x} = \boldsymbol{x}$ and $\varepsilon\lambda^{\mathrm{F}} = \lambda^{\mathrm{F}}$. This leads to some notable differences between this description and that in the literature. First, we note that the medium entropy change is a path functional that, for a master equation approach, is delivered continuously, not solely discontinuously at jumps as is usually expected when considering a system consisting of even variables. This also provides the first major revision to the mean entropy production rate as described by Schnakenberg [58] since we may now write

$$\begin{aligned}
\frac{d\langle\Delta\mathcal{S}_{\mathrm{tot}}\rangle_{P_0^{\mathrm{F}}}^{\mathrm{F}}}{dt} &= \sum_{\boldsymbol{x}\neq\boldsymbol{x}'} P(\boldsymbol{x},t)T(\boldsymbol{x}'|\boldsymbol{x},\lambda^{\mathrm{F}}(t))\ln\frac{P(\boldsymbol{x},t)T(\boldsymbol{x}'|\boldsymbol{x},\lambda^{\mathrm{F}}(t))}{P(\boldsymbol{x}',t)T(\boldsymbol{\varepsilon x}|\boldsymbol{\varepsilon x}',\varepsilon\lambda^{\mathrm{F}}(t))} \\
&\quad + \sum_{\boldsymbol{x}} P(\boldsymbol{x},t)\left(T(\boldsymbol{x}|\boldsymbol{x},\lambda^{\mathrm{F}}(t)) - T(\boldsymbol{\varepsilon x}|\boldsymbol{\varepsilon x},\varepsilon\lambda^{\mathrm{F}}(t))\right).
\end{aligned} \tag{4.10}$$

Secondly, since we generally suppose that under time reversal, there may be an arbitrary transformation in the time independent behaviour of the transition rates we must add another condition to those required for a DFT and related symmetries and behaviour. This is because $p^\dagger$ will only reduce to p^{F} as usual when the initial and final distributions are identical, protocol obeys $\lambda^{\mathrm{F}}(\tau - t) = \lambda^{\mathrm{F}}(t)$ as before *and* $\varepsilon\lambda^{\mathrm{F}}(t) = \lambda^{\mathrm{F}}(t)$. Conceptually, this addition means that in a given stationary state, where the total entropy production has previously been expected to obey a DFT, one will not necessarily observe a DFT unless the protocol is time reversal invariant. This behaviour ultimately stems from the fact that the nature of entropy production depends explicitly on the definition of time reversal in its

construction which depends on more than just the mathematics which describe the observed system, but rather on a set of decisions regarding how one should treat that system physically. As such the precise definition of the reverse path and the probability density of generating it is rendered somewhat ambiguous. Consequently one must be careful with the reasoning employed and the physical definitions one attaches to such quantities. This issue is expanded on in a later chapter.

4.2 The Use of the Adjoint Dynamics and Three Contributions to Total Entropy Production

Next we consider alternative specifications of p^*. In the same manner as in the consideration of Hatano-Sasa relation/non-adiabatic entropy production and Speck-Seifert relation/adiabatic entropy production we consider the adjoint dynamics which are those that lead to the same stationary state as the normal dynamics, T, but generate flux of the opposite sign in that stationary state. It can be shown [25, 36, 111] that this requires an adjoint transition rate matrix $T^{\rm ad}$ described by

$$T^{\rm ad}(\boldsymbol{x}|\boldsymbol{x}',\lambda^{\rm F}(t)) = T(\boldsymbol{x}'|\boldsymbol{x},\lambda^{\rm F}(t))\frac{P^{\rm F,st}(\boldsymbol{x},\lambda^{\rm F}(t))}{P^{\rm F,st}(\boldsymbol{x}',\lambda^{\rm F}(t))} \tag{4.11}$$

where $P^{\rm F,st}$ is the stationary probability distribution corresponding to $\lambda^{\rm F}(t)$. However, in the same way that the normal dynamics may not, in general, permit transitions $\boldsymbol{x}' \to \boldsymbol{x}$ or $\boldsymbol{\varepsilon x} \to \boldsymbol{\varepsilon x}'$, similarly the adjoint dynamics may not, in general, permit transitions $\boldsymbol{x} \to \boldsymbol{x}'$ or $\boldsymbol{\varepsilon x}' \to \boldsymbol{\varepsilon x}$. Thus we must consider the representation of the adjoint dynamics as either Eq. (4.11) or

$$T^{\rm ad}(\boldsymbol{\varepsilon x}'|\boldsymbol{\varepsilon x},\varepsilon\lambda^{\rm F}(t)) = T(\boldsymbol{\varepsilon x}|\boldsymbol{\varepsilon x}',\varepsilon\lambda^{\rm F}(t))\frac{P^{\rm F,st}(\boldsymbol{\varepsilon x}',\varepsilon\lambda^{\rm F}(t))}{P^{\rm F,st}(\boldsymbol{\varepsilon x},\varepsilon\lambda^{\rm F}(t))} \tag{4.12}$$

depending on the specific transition being considered. Explicitly, when choosing $p^*[\vec{\boldsymbol{x}}^*]$, we should not consider $p^{\rm F,ad}[\vec{\boldsymbol{x}}]_{P_0^{\rm F}}$ or $p^{\dagger,{\rm ad}}[\vec{\boldsymbol{x}}^\dagger]_{P_0^\dagger}$.

Under the adjoint dynamics, however, an appropriate involution choice is that which we have seen before, for the fluctuation theorems with only even variables, and consists of a reversal of sequence of the path, but without the instantaneous time reversal operation such that $f(t) \to f^{\rm R}(t) \equiv f(\tau - t)$. As such the path is given by $\boldsymbol{x}^*(t) = \boldsymbol{x}^{\rm R}(t) = \boldsymbol{x}(\tau - t)$. Applying the same involution yields the backwards protocol $\lambda^*(t) = \lambda^{\rm F}(\tau - t) = \lambda^{\rm R}(t)$ and the initial distribution $P^*(\boldsymbol{x}^*(0),0) = P^{\rm R,ad}(\boldsymbol{x}^{\rm R}(0),0) = P^{\rm F}(\boldsymbol{x}(\tau),\tau)$ so that $p^*[\vec{\boldsymbol{x}}^*]_{P_0^{\rm R}} = p^*[\vec{\boldsymbol{x}}^*]_{P_\tau^{\rm F}}$. The path probability density under this involution and adjoint dynamics may then be written

$$\begin{aligned} p^{\rm R,ad}[\vec{\boldsymbol{x}}^{\rm R}]_{P_0^{\rm R}} &= P^{\rm R}(\boldsymbol{x}_0^{\rm R},0)e^{\int_{t_0}^{t_1} dt T^{\rm ad}(\boldsymbol{x}_0^{\rm R}|\boldsymbol{x}_0^{\rm R},\lambda^{\rm R}(t))}\prod_{i=1}^{N} T^{\rm ad}(\boldsymbol{x}_i^{\rm R}|\boldsymbol{x}_{i-1}^{\rm R},\lambda^{\rm R}(t_i))e^{\int_{t_i}^{t_{i+1}} dt T^{\rm ad}(\boldsymbol{x}_i^{\rm R}|\boldsymbol{x}_i^{\rm R},\lambda^{\rm R}(t))} \\ &= P^{\rm F}(\boldsymbol{x}_N,\tau)e^{-\int_{t_1}^{t_0} dt T^{\rm ad}(\boldsymbol{x}_0|\boldsymbol{x}_0,\lambda^{\rm F}(t))}\prod_{i=1}^{N} e^{-\int_{t_{i+1}}^{t_i} dt T^{\rm ad}(\boldsymbol{x}_i|\boldsymbol{x}_i,\lambda^{\rm F}(t))}T^{\rm ad}(\boldsymbol{x}_{i-1}|\boldsymbol{x}_i,\lambda^{\rm F}(t_i)). \end{aligned} \tag{4.13}$$

We then construct a quantity of the form given in Eq. (4.2) and utilise Eq. (4.11) to obtain

$$\begin{aligned}\Delta\mathcal{S}_1 &= \ln p^{\mathrm{F}}[\vec{\boldsymbol{x}}]_{P_0^{\mathrm{F}}} - \ln p^{\mathrm{R,ad}}[\vec{\boldsymbol{x}}^{\mathrm{R}}]_{P_0^{\mathrm{R}}} \\ &= \ln\frac{P^{\mathrm{F}}(\boldsymbol{x}_0,0)}{P^{\mathrm{F}}(\boldsymbol{x}_N,\tau)} + \sum_{i=0}^{N}\ln\frac{e^{\int_{t_i}^{t_{i+1}}dt\ T(\boldsymbol{x}_i|\boldsymbol{x}_i,\lambda^{\mathrm{F}}(t))}}{e^{\int_{t_i}^{t_{i+1}}dt\ T^{\mathrm{ad}}(\boldsymbol{x}_i|\boldsymbol{x}_i,\lambda^{\mathrm{F}}(t))}} + \sum_{i=1}^{N}\ln\frac{T(\boldsymbol{x}_i|\boldsymbol{x}_{i-1},\lambda^{\mathrm{F}}(t_i))}{T^{\mathrm{ad}}(\boldsymbol{x}_{i-1}|\boldsymbol{x}_i,\lambda^{\mathrm{F}}(t_i))} \\ &= \ln\frac{P^{\mathrm{F}}(\boldsymbol{x}_0,0)}{P^{\mathrm{F}}(\boldsymbol{x}_N,\tau)} + \int_0^{\tau}dt\ \left(T(\boldsymbol{x}(t)|\boldsymbol{x}(t),\lambda^{\mathrm{F}}(t)) - T^{\mathrm{ad}}(\boldsymbol{x}(t)|\boldsymbol{x}(t),\lambda^{\mathrm{F}}(t))\right) \\ &\quad + \sum_{i=1}^{N}\ln\frac{P^{\mathrm{F,st}}(\boldsymbol{x}_i,\lambda^{\mathrm{F}}(t_i))}{P^{\mathrm{F,st}}(\boldsymbol{x}_{i-1},\lambda^{\mathrm{F}}(t_i))}.\end{aligned} \tag{4.14}$$

However, since by defining the adjoint dynamics we have assumed the existence of a stationary state, we may simplify this expression by an explicit consideration of balance under the adjoint dynamics as follows

$$\sum_{\boldsymbol{x}'\neq\boldsymbol{x}} P^{\mathrm{F,st}}(\boldsymbol{x},\lambda^{\mathrm{F}}(t))T^{\mathrm{ad}}(\boldsymbol{x}'|\boldsymbol{x},\lambda^{\mathrm{F}}(t)) = \sum_{\boldsymbol{x}'\neq\boldsymbol{x}} P^{\mathrm{F,st}}(\boldsymbol{x}',\lambda^{\mathrm{F}}(t))T^{\mathrm{ad}}(\boldsymbol{x}|\boldsymbol{x}',\lambda^{\mathrm{F}}(t)). \tag{4.15}$$

We may rearrange and identify

$$\begin{aligned}\sum_{\boldsymbol{x}'\neq\boldsymbol{x}} T^{\mathrm{ad}}(\boldsymbol{x}'|\boldsymbol{x},\lambda^{\mathrm{F}}(t)) &= \sum_{\boldsymbol{x}'\neq\boldsymbol{x}} \frac{P^{\mathrm{F,st}}(\boldsymbol{x}',\lambda^{\mathrm{F}}(t))}{P^{\mathrm{F,st}}(\boldsymbol{x},\lambda^{\mathrm{F}}(t))}T^{\mathrm{ad}}(\boldsymbol{x}|\boldsymbol{x}',\lambda^{\mathrm{F}}(t)) \\ \sum_{\boldsymbol{x}'\neq\boldsymbol{x}} T^{\mathrm{ad}}(\boldsymbol{x}'|\boldsymbol{x},\lambda^{\mathrm{F}}(t)) &= \sum_{\boldsymbol{x}'\neq\boldsymbol{x}} T(\boldsymbol{x}'|\boldsymbol{x},\lambda^{\mathrm{F}}(t)) \\ -T^{\mathrm{ad}}(\boldsymbol{x}|\boldsymbol{x},\lambda^{\mathrm{F}}(t)) &= -T(\boldsymbol{x}|\boldsymbol{x},\lambda^{\mathrm{F}}(t))\end{aligned} \tag{4.16}$$

allowing us to simplify

$$\begin{aligned}\Delta\mathcal{S}_1 &= \ln P^{\mathrm{F}}[\vec{\boldsymbol{x}}]_{P_0^{\mathrm{F}}} - \ln P^{\mathrm{R,ad}}[\vec{\boldsymbol{x}}^{\mathrm{R}}]_{P_0^{\mathrm{R}}} \\ &= \ln\frac{P^{\mathrm{F}}(\boldsymbol{x}_0,0)}{P^{\mathrm{F}}(\boldsymbol{x}_N,\tau)} + \sum_{i=1}^{N}\ln\frac{P^{\mathrm{F,st}}(\boldsymbol{x}_i,\lambda^{\mathrm{F}}(t_i))}{P^{\mathrm{F,st}}(\boldsymbol{x}_{i-1},\lambda^{\mathrm{F}}(t_i))}.\end{aligned} \tag{4.17}$$

Recognising the structure of its form and by identifying $\mathcal{A}[\vec{\boldsymbol{x}}] = \Delta\mathcal{S}_1$ and $\hat{\mathcal{A}}[\vec{\boldsymbol{x}}^{\mathrm{R}}] = \Delta\hat{\mathcal{S}}_1$ we find that it obeys

$$\langle\exp\left[-\Delta\mathcal{S}_1\right]\rangle^{\mathrm{F}}_{P_0^{\mathrm{F}}} = 1, \tag{4.18}$$

$$P^{\mathrm{R,ad}}(\Delta\hat{\mathcal{S}}_1 = -A)_{P_\tau^{\mathrm{F}}} = e^{-A}P^{\mathrm{F}}(\Delta\mathcal{S}_1 = A)_{P_0^{\mathrm{F}}} \tag{4.19}$$

and

$$\langle g(\mathcal{C}[\vec{\boldsymbol{x}}^{\mathrm{R}}])\rangle^{\mathrm{R,ad}}_{P_\tau^{\mathrm{R}}} = \langle g(\mathcal{B}[\vec{\boldsymbol{x}}])e^{-\Delta\mathcal{S}_1}\rangle^{\mathrm{F}}_{P_0^{\mathrm{F}}}, \tag{4.20}$$

the first of which exists in the literature as the Hatano-Sasa relation [93, 112] or IFT for the non-adiabatic entropy production [36, 98, 99] with the final two being their logical extension when the symmetry is considered more generally.

Let us now consider, once again under the adjoint dynamics, the involution choice $f(t) \to f^{\mathrm{T}}(t) \equiv \varepsilon f(t)$ which in turn leads to the path $\boldsymbol{x}^*(t) = \boldsymbol{x}^{\mathrm{T}}(t) = \boldsymbol{\varepsilon x}(t)$ (all of the considered involutions are illustrated in Fig. 4.1). Applying the involution to the protocol we obtain $\lambda^*(t) = \lambda^{\mathrm{T}}(t) = \varepsilon\lambda^{\mathrm{F}}(t)$ and initial distribution $P^*(\boldsymbol{x}^*(0),0) = P^{\mathrm{T}}(\boldsymbol{x}^{\mathrm{T}}(0),0) = \hat{\boldsymbol{\varepsilon}}P^{\mathrm{F}}(\boldsymbol{\varepsilon x}(0),0) = P^{\mathrm{F}}(\boldsymbol{x}(0),0)$ such that $p^{\mathrm{T}}[\vec{\boldsymbol{x}}^{\mathrm{T}}]_{P_0^{\mathrm{T}}} = p^{\mathrm{T}}[\vec{\boldsymbol{x}}^{\mathrm{T}}]_{\hat{\boldsymbol{\varepsilon}}P_0^{\mathrm{F}}}$.

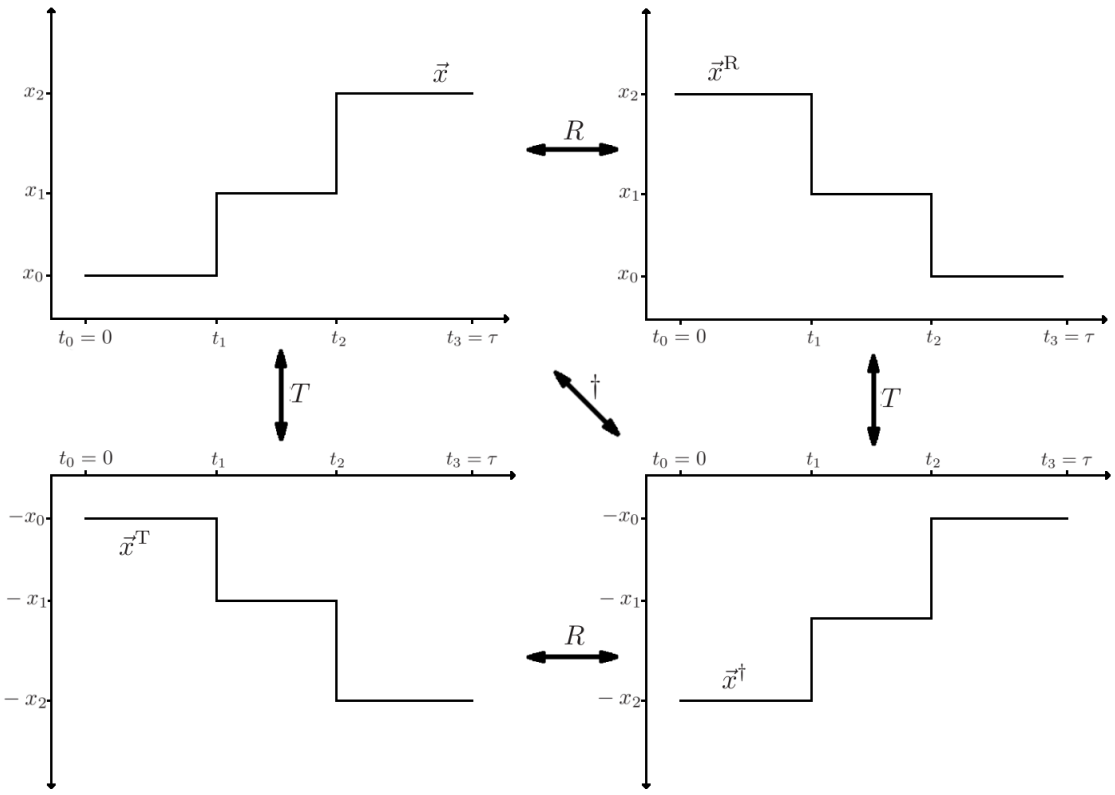

Figure 4.1: Illustration of the discretisation procedure and all possible involutions for the evolution of a sole odd variable for $N = 2$.

The path probability density for this case is therefore

$$p^{\mathrm{T,ad}}[\vec{\boldsymbol{x}}^{\mathrm{T}}]_{P_0^{\mathrm{T}}} = P^{\mathrm{T}}(\boldsymbol{x}_0^{\mathrm{T}}, 0) e^{\int_{t_0}^{t_1} dt T^{\mathrm{ad}}(\boldsymbol{x}_0^{\mathrm{T}}|\boldsymbol{x}_0^{\mathrm{T}}, \lambda^{\mathrm{T}}(t))} \prod_{i=1}^{N} T^{\mathrm{ad}}(\boldsymbol{x}_i^{\mathrm{T}}|\boldsymbol{x}_{i-1}^{\mathrm{T}}, \lambda^{\mathrm{T}}(t_i)) e^{\int_{t_i}^{t_{i+1}} dt T^{\mathrm{ad}}(\boldsymbol{x}_i^{\mathrm{T}}|\boldsymbol{x}_i^{\mathrm{T}}, \lambda^{\mathrm{T}}(t))}$$
$$= P^{\mathrm{F}}(\boldsymbol{x}_0, 0) e^{\int_{t_0}^{t_1} dt T^{\mathrm{ad}}(\boldsymbol{\varepsilon x}_0|\boldsymbol{\varepsilon x}_0, \varepsilon\lambda^{\mathrm{F}}(t))} \prod_{i=1}^{N} T^{\mathrm{ad}}(\boldsymbol{\varepsilon x}_i|\boldsymbol{\varepsilon x}_{i-1}, \varepsilon\lambda^{\mathrm{F}}(t_i)) e^{\int_{t_i}^{t_{i+1}} dt T^{\mathrm{ad}}(\boldsymbol{\varepsilon x}_i|\boldsymbol{\varepsilon x}_i, \varepsilon\lambda^{\mathrm{F}}(t))}. \tag{4.21}$$

By Eq. (4.2), along with the property $T(\boldsymbol{x}_i|\boldsymbol{x}_i, \lambda^{\mathrm{F}}(t)) = T^{\mathrm{ad}}(\boldsymbol{x}_i|\boldsymbol{x}_i, \lambda^{\mathrm{F}}(t))$ from Eq. (4.16), this then allows us to define

$$\begin{aligned}
\Delta\mathcal{S}_2 &= \ln p^{\mathrm{F}}[\vec{\boldsymbol{x}}]_{P_0^{\mathrm{F}}} - \ln p^{\mathrm{T,ad}}[\vec{\boldsymbol{x}}^{\mathrm{T}}]_{P_0^{\mathrm{T}}} \\
&= \sum_{i=0}^{N} \ln \frac{e^{\int_{t_i}^{t_{i+1}} dt T(\boldsymbol{x}_i|\boldsymbol{x}_i, \lambda^{\mathrm{F}}(t))}}{e^{\int_{t_i}^{t_{i+1}} dt T(\boldsymbol{\varepsilon x}_i|\boldsymbol{\varepsilon x}_i, \varepsilon\lambda^{\mathrm{F}}(t))}} + \sum_{i=1}^{N} \ln \frac{P^{\mathrm{F,st}}(\boldsymbol{\varepsilon x}_{i-1}, \varepsilon\lambda^{\mathrm{F}}(t_i))}{P^{\mathrm{F,st}}(\boldsymbol{\varepsilon x}_i, \varepsilon\lambda^{\mathrm{F}}(t_i))} \frac{T(\boldsymbol{x}_i|\boldsymbol{x}_{i-1}, \lambda^{\mathrm{F}}(t_i))}{T(\boldsymbol{\varepsilon x}_{i-1}|\boldsymbol{\varepsilon x}_i, \varepsilon\lambda^{\mathrm{F}}(t_i))} \\
&= \int_0^{\tau} dt \ \big(T(\boldsymbol{x}(t)|\boldsymbol{x}(t), \lambda^{\mathrm{F}}(t)) - T(\boldsymbol{\varepsilon x}(t)|\boldsymbol{\varepsilon x}(t), \varepsilon\lambda^{\mathrm{F}}(t))\big) \\
&\quad + \sum_{i=1}^{N} \ln \frac{P^{\mathrm{F,st}}(\boldsymbol{\varepsilon x}_{i-1}, \varepsilon\lambda^{\mathrm{F}}(t_i))}{P^{\mathrm{F,st}}(\boldsymbol{\varepsilon x}_i, \varepsilon\lambda^{\mathrm{F}}(t_i))} \frac{T(\boldsymbol{x}_i|\boldsymbol{x}_{i-1}, \lambda^{\mathrm{F}}(t_i))}{T(\boldsymbol{\varepsilon x}_{i-1}|\boldsymbol{\varepsilon x}_i, \varepsilon\lambda^{\mathrm{F}}(t_i))}.
\end{aligned} \tag{4.22}$$

Similarly, by recognising the structure of its form we find that it obeys

$$\langle \exp\left[-\Delta\mathcal{S}_2\right] \rangle_{P_0^{\mathrm{F}}}^{\mathrm{F}} = 1 \tag{4.23}$$

and

$$\langle g(\mathcal{C}[\vec{\boldsymbol{x}}^{\mathrm{T}}])\rangle^{\mathrm{T,ad}}_{\hat{\boldsymbol{\varepsilon}}P^{\mathrm{F}}_0} = \langle g(\mathcal{B}[\vec{\boldsymbol{x}}])e^{-\Delta\mathcal{S}_2}\rangle^{\mathrm{F}}_{P^{\mathrm{F}}_0}, \tag{4.24}$$

but note that unlike for $\Delta\mathcal{S}_{\mathrm{tot}}$ and $\Delta\mathcal{S}_1$, $\Delta\mathcal{S}_2$ is odd with respect to the involution $\vec{\boldsymbol{x}} \to \vec{\boldsymbol{x}}^{\mathrm{T}}$ owing to the lack of a system entropy term. As such we can identify $\mathcal{A}[\vec{\boldsymbol{x}}] = \Delta\mathcal{S}_2$, but also that $\hat{\mathcal{A}}[\vec{\boldsymbol{x}}^{\mathrm{T}}] = \Delta\hat{\mathcal{S}}_2 = \Delta\mathcal{S}_2^{\mathrm{T,ad}}$ and so can write, more generally,

$$p^{\mathrm{T,ad}}(\Delta\mathcal{S}_2^{\mathrm{T,ad}} = -A)_{\hat{\boldsymbol{\varepsilon}}P^{\mathrm{F}}_0} = e^{-A}p^{\mathrm{F}}(\Delta\mathcal{S}_2 = A)_{P^{\mathrm{F}}_0}. \tag{4.25}$$

Unlike $\Delta\mathcal{S}_1$, the quantity $\Delta\mathcal{S}_2$ and relevant fluctuation theorems are new in the literature. We must immediately recognise that $\Delta\mathcal{S}_{\mathrm{tot}} \neq \Delta\mathcal{S}_1 + \Delta\mathcal{S}_2$ differing by a quantity

$$\Delta\mathcal{S}_3 = \sum_{i=1}^{N} \ln \frac{P^{\mathrm{F,st}}(\boldsymbol{x}_{i-1}, \lambda^{\mathrm{F}}(t_i))P^{\mathrm{F,st}}(\boldsymbol{\varepsilon}\boldsymbol{x}_i, \varepsilon\lambda^{\mathrm{F}}(t_i))}{P^{\mathrm{F,st}}(\boldsymbol{x}_i, \lambda^{\mathrm{F}}(t_i))P^{\mathrm{F,st}}(\boldsymbol{\varepsilon}\boldsymbol{x}_{i-1}, \varepsilon\lambda^{\mathrm{F}}(t_i))} \tag{4.26}$$

such that $\Delta\mathcal{S}_{\mathrm{tot}} = \Delta\mathcal{S}_1 + \Delta\mathcal{S}_2 + \Delta\mathcal{S}_3$. If however, $\boldsymbol{\varepsilon}\boldsymbol{x} = \boldsymbol{x}$ and $\varepsilon\lambda^{\mathrm{F}}(t) = \lambda^{\mathrm{F}}(t)$ then $\Delta\mathcal{S}_3 = 0$ and $\Delta\mathcal{S}_2$ reduces to the adiabatic entropy production appearing in [36, 98, 99].

We make note here that when one considers the form of the final two fluctuation relations for $\Delta\mathcal{S}_1$ and $\Delta\mathcal{S}_2$ they are somewhat unhelpful since they explicitly concern dynamics and trajectories which are not realisable under the forward dynamics and therefore may be entirely unphysical. As such it is instructive to define what we shall call the 'reversed adjoint dynamics' which should be defined as follows

$$T^{\text{ad-rev}}(\boldsymbol{\varepsilon}\boldsymbol{x}|\boldsymbol{\varepsilon}\boldsymbol{x}', \varepsilon\lambda(t)) = T^{\mathrm{ad}}(\boldsymbol{x}|\boldsymbol{x}', \lambda(t)) \tag{4.27}$$

and

$$T^{\text{ad-rev}}(\boldsymbol{x}'|\boldsymbol{x}, \lambda(t)) = T^{\mathrm{ad}}(\boldsymbol{\varepsilon}\boldsymbol{x}'|\boldsymbol{\varepsilon}\boldsymbol{x}, \varepsilon\lambda(t)). \tag{4.28}$$

Such dynamics are physically realisable and so allow us to rewrite the final fluctuation relations in the strictly identical, but more helpful forms

$$P^{\dagger,\text{ad-rev}}(\Delta\hat{\mathcal{S}}_1 = -A)_{\hat{\boldsymbol{\varepsilon}}P^{\mathrm{F}}_\tau} = e^{-A}P^{\mathrm{F}}(\Delta\mathcal{S}_1 = A)_{P^{\mathrm{F}}_0} \tag{4.29}$$

and

$$\langle g(\mathcal{C}[\vec{\boldsymbol{x}}^{\dagger}])\rangle^{\dagger,\text{ad-rev}}_{\hat{\boldsymbol{\varepsilon}}P^{\mathrm{F}}_\tau} = \langle g(\mathcal{B}[\vec{\boldsymbol{x}}])e^{-\Delta\mathcal{S}_1}\rangle^{\mathrm{F}}_{P^{\mathrm{F}}_0} \tag{4.30}$$

for $\Delta\mathcal{S}_1$ where $\mathcal{A}[\vec{\boldsymbol{x}}] = \Delta\mathcal{S}_1$ and $\hat{\mathcal{A}}[\vec{\boldsymbol{x}}^{\dagger}] = \Delta\hat{\mathcal{S}}_1$. Similarly, we may also write

$$p^{\mathrm{F,ad-rev}}(\Delta\mathcal{S}_2^{\text{ad-rev}} = -A)_{P^{\mathrm{F}}_0} = e^{-A}p^{\mathrm{F}}(\Delta\mathcal{S}_2 = A)_{P^{\mathrm{F}}_0} \tag{4.31}$$

and

$$\begin{aligned}\langle g(\mathcal{C}[\vec{\boldsymbol{x}}])\rangle^{\mathrm{F,ad-rev}}_{P^{\mathrm{F}}_0} &= \langle g(\mathcal{B}[\vec{\boldsymbol{x}}])e^{-\Delta\mathcal{S}_2}\rangle^{\mathrm{F}}_{P^{\mathrm{F}}_0} \\ &= \langle g(\mathcal{C}[\vec{\boldsymbol{x}}])e^{-\Delta\mathcal{S}_2}\rangle^{\mathrm{F}}_{P^{\mathrm{F}}_0}\end{aligned} \tag{4.32}$$

for $\Delta\mathcal{S}_2$ where $\mathcal{A}[\vec{\boldsymbol{x}}] = \Delta\mathcal{S}_2$ and $\hat{\mathcal{A}}[\vec{\boldsymbol{x}}] = \Delta\hat{\mathcal{S}}_2 = \Delta\mathcal{S}_2^{\text{ad-rev}}$. Again, the nature of the adjoint and reversed adjoint dynamics are not of primary concern since their purpose has been to elucidate the main contributions to entropy production which we shall now discuss, however the physical interpretation and relationship between adjoint and reversed adjoint dynamics will be discussed in the later chapter on

continuous systems.

4.3 Thermodynamic Interpretation and Relation to other Fluctuation Theorems

An important consequence we must recognise is that neither

$$\begin{aligned}\Delta\mathcal{S}_{\text{tot}} - \Delta\mathcal{S}_1 &= \Delta\mathcal{S}_2 + \Delta\mathcal{S}_3 \\ &= \ln p^{\text{R,ad}}[\vec{\boldsymbol{x}}^{\text{R}}]_{P_0^{\text{F}}} - \ln p^{\dagger}[\vec{\boldsymbol{x}}^{\dagger}]_{\hat{\boldsymbol{\varepsilon}}P_0^{\text{F}}} \\ &= \ln p^{\dagger,\text{ad-rev}}[\vec{\boldsymbol{x}}^{\dagger}]_{P_0^{\text{F}}} - \ln p^{\dagger}[\vec{\boldsymbol{x}}^{\dagger}]_{\hat{\boldsymbol{\varepsilon}}P_0^{\text{F}}}\end{aligned} \tag{4.33}$$

nor

$$\begin{aligned}\Delta\mathcal{S}_{\text{tot}} - \Delta\mathcal{S}_2 &= \Delta\mathcal{S}_1 + \Delta\mathcal{S}_3 \\ &= \ln p^{\text{T,ad}}[\vec{\boldsymbol{x}}^{\text{T}}]_{\hat{\boldsymbol{\varepsilon}}P_0^{\text{F}}} - \ln p^{\dagger}[\vec{\boldsymbol{x}}^{\dagger}]_{\hat{\boldsymbol{\varepsilon}}P_\tau^{\text{F}}} \\ &= \ln p^{\text{F,ad-rev}}[\vec{\boldsymbol{x}}]_{P_0^{\text{F}}} - \ln p^{\dagger}[\vec{\boldsymbol{x}}^{\dagger}]_{\hat{\boldsymbol{\varepsilon}}P_\tau^{\text{F}}}\end{aligned} \tag{4.34}$$

can be written in the form required for Eq. (3.10) and so do not obey an IFT and do not necessarily have any bounds on the sign of their mean. The implication of such a division of the entropy production is to identify that, in general, the mean total entropy production cannot be split into two rigorously positive quantities aligned with relaxation and driving and steady non-equilibrium constraints as argued in [36, 98, 99]. For circumstances where it is suitable to associate an external temperature with the external or medium entropy change we may proceed by following the formalism of Seifert [12, 33] and write (simplifying for one state dependent temperature and reintroducing k_B since such a system is implicitly thermal)

$$\Delta\mathcal{S}_{\text{tot}} = k_B \ln \frac{P^{\text{F}}(\boldsymbol{x}(0),0)}{P^{\text{F}}(\boldsymbol{x}(\tau),\tau)} + \frac{\Delta Q}{T} = \Delta\mathcal{S}_{\text{sys}} + \frac{\Delta Q}{T}, \tag{4.35}$$

where T is the temperature of the environment. Further, under such conditions, to understand our division of entropy production we should attempt to align our quantities with those of Oono and Paniconi, such that total heat transfer to the environment, ΔQ, is the sum of the excess heat and house-keeping heat $\Delta Q = \Delta Q_{\text{ex}} + \Delta Q_{\text{hk}}$ [35]. It is in this division and the precise definitions of these quantities that we make three notable points. First, however we must understand their meaning. The house-keeping heat was introduced as the heat flow that is constantly dissipated in a non-equilibrium stationary state or the heat flow required to keep the system out of equilibrium whilst the excess heat comprised the remainder of the heat flow and characterised system behaviour on top of the steady dissipation of heat. What, however, is lacking in this definition, despite the intuitive approach, is the precise microscopic (or otherwise, that is to say mean) definition of such quantities without which two questions immediately arise. Assuming the characterisation of a stationary state is by its mean behaviour (that is to say by a stationary probability distribution), are the excess and house-keeping heats microscopic or mean heat flows? And the second, but related question is how one defines the house-keeping heat when the system is not in the stationary state. In other words, is it the value of the functional form evaluated out of the stationary state or is it the quantity which would be required to maintain the corresponding stationary state to which the system would relax if left unperturbed. For the case of even variables these questions have either been answered or are irrelevant. This is illustrated by the adiabatic and non-adiabatic entropy productions which generalise the Hatano-Sasa and Speck-Seifert relations respectively. For ex-

ample, the non-adiabatic entropy contribution, which is comprised of the excess heat, is identically zero for all individual realisations and in the mean when the system is in the stationary state. The remainder, the adiabatic entropy contribution which is comprised of the house-keeping heat, similarly vanishes, in both mean and in detail, when the corresponding stationary state of the system is the equilibrium distribution such that detailed balance holds.

However, when we look at the contributions when odd time reversal behaviour is included this picture is not as simple. $\Delta\mathcal{S}_1$ vanishes in both mean and in detail when the system is in the stationary state, which is to be expected as it has the same functional form as the non-adiabatic entropy production. This together with the adherence to an IFT would imply alignment with the excess heat. Similarly $\Delta\mathcal{S}_2$, contributes in both mean and in detail out of equilibrium and vanishes in mean and in detail when the underlying stationary state is in equilibrium. Likewise, its adherence to an IFT, and thus its positivity in the mean, suggests alignment with the house-keeping heat. This however, leaves $\Delta\mathcal{S}_3$ unaccounted for. Examining its structure we see that it must vanish for all individual realisations when the underlying dynamics produce an equilibrium stationary state owing to time reversal invariance in that stationary state, but if we consider its mean properties we see that on average it vanishes when the system is in any stationary state. This is easily seen by recognising

$$\begin{aligned}\frac{d\langle\Delta\mathcal{S}_3\rangle^{\mathrm{F}}_{P^{\mathrm{F}}_0}}{dt} &= \sum_{\boldsymbol{x},\boldsymbol{x}'} P^{\mathrm{F}}(\boldsymbol{x},t)T(\boldsymbol{x}'|\boldsymbol{x},\lambda^{\mathrm{F}}(t))\ln\frac{P^{\mathrm{F,st}}(\boldsymbol{x},\lambda^{\mathrm{F}}(t))P^{\mathrm{F,st}}(\boldsymbol{\varepsilon x}',\varepsilon\lambda^{\mathrm{F}}(t))}{P^{\mathrm{F,st}}(\boldsymbol{x}',\lambda^{\mathrm{F}}(t))P^{\mathrm{F,st}}(\boldsymbol{\varepsilon x},\varepsilon\lambda^{\mathrm{F}}(t))}\\ &= \sum_{\boldsymbol{x},\boldsymbol{x}'} P^{\mathrm{F}}(\boldsymbol{x},t)T(\boldsymbol{x}'|\boldsymbol{x},\lambda^{\mathrm{F}}(t))\left[\ln\frac{P^{\mathrm{F,st}}(\boldsymbol{x},\lambda^{\mathrm{F}}(t))}{P^{\mathrm{F,st}}(\boldsymbol{\varepsilon x},\varepsilon\lambda^{\mathrm{F}}(t))} - \ln\frac{P^{\mathrm{F,st}}(\boldsymbol{x}',\lambda^{\mathrm{F}}(t))}{P^{\mathrm{F,st}}(\boldsymbol{\varepsilon x}',\varepsilon\lambda^{\mathrm{F}}(t))}\right]\\ &= -\sum_{\boldsymbol{x}'}\frac{dP^{\mathrm{F}}(\boldsymbol{x}',t)}{dt}\ln\frac{P^{\mathrm{F,st}}(\boldsymbol{x}',\lambda^{\mathrm{F}}(t))}{P^{\mathrm{F,st}}(\boldsymbol{\varepsilon x}',\varepsilon\lambda^{\mathrm{F}}(t))}.\end{aligned} \tag{4.36}$$

Such an ambiguity is reflected in the fact that, unlike for systems with only even variables and the resultant adiabatic entropy production, there is no single quantity which uniquely vanishes when detailed balance is obeyed. That is to say the breakage of detailed balance and thus the departure from equilibrium of the stationary state could be measured by any of $\Delta\mathcal{S}_2$, $\Delta\mathcal{S}_3$ or $\Delta\mathcal{S}_2+\Delta\mathcal{S}_3$ since all vanish in equilibrium. This means we need to revisit the question of whether the excess and house-keeping heats are mean or microscopic quantities. If they are defined by their mean behaviour, $\Delta\mathcal{S}_3$ would align with $\Delta\mathcal{S}_1$ and the excess heat whilst if they are defined by their microscopic behaviour, $\Delta\mathcal{S}_3$ would align with $\Delta\mathcal{S}_2$ and the house-keeping heat.

Ultimately, since we have defined microscopic quantities which are subject to constraints upon their distribution, namely fluctuation theorems, we consider such quantities to be defined microscopically. However, we remain aware of the mean behaviour of all our quantities since many of the key properties of the fluctuation theorems, namely the second law and its generalisations, are considered in terms of mean quantities. Doing so allows us to associate the house-keeping heat with all entropy production contributions that arise from a non-equilibrium constraint that breaks detailed balance. Such a consideration (see Sect. 3.7) leads us to the following definition

$$\Delta Q_{\mathrm{hk}} = (\Delta\mathcal{S}_2+\Delta\mathcal{S}_3)T. \tag{4.37}$$

Consequently, with $\Delta\mathcal{S}_1$ vanishing for all trajectories in the stationary state we consolidate the definition of the excess heat as the heat transfer associated with an entropy flow that exactly cancels the change

in system entropy over a period in the stationary state such that

$$\Delta Q_{\rm ex} = (\Delta \mathcal{S}_1 - \Delta \mathcal{S}_{\rm sys})T. \tag{4.38}$$

Following from these definitions, the first of our three statements about the thermodynamics of such systems is that the house-keeping heat can be split into two thermodynamically meaningful quantities which map onto $\Delta\mathcal{S}_2$ and $\Delta\mathcal{S}_3$. Despite the definition of our quantities as microscopic heat contributions it is important to recognise their mean behaviour and they are named to reflect this. Since the mean rate of change of $\Delta\mathcal{S}_3$ vanishes in the stationary state we define the 'transient house-keeping heat'

$$\Delta Q_{\rm hk,T} = \Delta \mathcal{S}_3 T \tag{4.39}$$

and the 'generalised house-keeping heat'

$$\Delta Q_{\rm hk,G} = \Delta \mathcal{S}_2 T \tag{4.40}$$

such that $\Delta Q_{\rm hk} = \Delta Q_{\rm hk,T} + \Delta Q_{\rm hk,G}$. A further implication of $\langle d\Delta\mathcal{S}_3/dt\rangle^{\rm F,st}_{p_0^{\rm F,st}} = 0$, is that the generalised house-keeping heat, when averaged, has the mean properties previously attributed to the house-keeping heat. It is the generalised house-keeping heat (appropriately scaled by k_BT) which obeys an IFT, is rigorously positive in the mean and thus reflects the positive heat flow required to maintain a stationary state. As such we may rewrite Eq. (4.23) as

$$\langle \exp\left[-\Delta Q_{\rm hk,G}/k_BT\right]\rangle^{\rm F}_{P_0^{\rm F}} = 1 \tag{4.41}$$

providing the bound $\langle \Delta Q_{\rm hk,G}\rangle^{\rm F} \geq 0$ for all times, protocols and initial conditions.

Our second statement, as a corollary to the first, states that in general $\Delta Q_{\rm hk}/T = \Delta\mathcal{S}_2 + \Delta\mathcal{S}_3 = k_B \ln p^{\rm R,ad}[\vec{\boldsymbol{x}}^{\rm R}]_{P_0^{\rm F}} - k_B \ln p^{\dagger}[\vec{\boldsymbol{x}}^{\dagger}]_{\hat{\boldsymbol{\varepsilon}} P_0^{\rm F}}$, which cannot be written in the form $\ln p^{\rm F}[\vec{\boldsymbol{x}}]_{P_0^{\rm F}} - \ln p^{*}[\vec{\boldsymbol{x}}^{*}]_{P_0^{*}}$, does not obey the equality which was previously derived for over-damped dynamics [94]. This is the statement

$$\langle \exp\left[-\Delta Q_{\rm hk}/k_BT\right]\rangle^{\rm F}_{P_0^{\rm F}} \neq 1 \tag{4.42}$$

providing no bounds on $\langle \Delta Q_{\rm hk}\rangle^{\rm F}$ *except* in the stationary state when $\Delta Q_{\rm hk}/T = \Delta\mathcal{S}_{\rm tot}$ or generally when $P^{\rm F,st}(\boldsymbol{\varepsilon}\boldsymbol{x}, \varepsilon\lambda^{\rm F}(t)) = P^{\rm F,st}(\boldsymbol{x}, \lambda^{\rm F}(t))$.

Our final point relates to the Hatano-Sasa equality and second law generalisation. Given the interpretation laid out above we also find that one has in agreement with the original statement

$$\langle \exp[-k_B^{-1}\Delta\mathcal{S}_{\rm sys} - \Delta Q_{\rm ex}/k_BT]\rangle^{\rm F}_{P_0^{\rm F}} = 1 \tag{4.43}$$

which leads to the generalised second law inequality

$$\Delta S \geq -\frac{\langle \Delta Q_{\rm ex}\rangle^{\rm F}_{P_0^{\rm F}}}{T} \tag{4.44}$$

where S is the Gibbs entropy $\langle \mathcal{S}_{\rm sys}\rangle = -k_B \sum P \ln P$. A key aspect to this second law generalisation is that it concerns state functions and mean heat flows and comprises solely of transient terms that reach fixed values when the system is in the steady state. However since $\Delta\mathcal{S}_1 + \Delta\mathcal{S}_3 = \ln p^{\rm T,ad}[\vec{\boldsymbol{x}}^{\rm T}]_{\hat{\boldsymbol{\varepsilon}} P_0^{\rm F}} - \ln p^{\dagger}[\vec{\boldsymbol{x}}^{\dagger}]_{\hat{\boldsymbol{\varepsilon}} P_\tau^{\rm F}}$ cannot be written in the form $\ln p^{\rm F}[\vec{\boldsymbol{x}}]_{P_0^{\rm F}} - \ln p^{*}[\vec{\boldsymbol{x}}^{*}]_{P_0^{*}}$, the failure of its adherence to an

IFT,

$$\langle \exp[-k_B^{-1}\Delta\mathcal{S}_{\rm sys} - (\Delta Q_{\rm ex} + \Delta Q_{\rm hk,T})/k_B T]\rangle^{\rm F}_{P_0^{\rm F}} \neq 1, \tag{4.45}$$

expresses a failure to express a second law generalisations in terms of *all* the transient state variables and heat flows of the system unlike the case where only even time reversals are present. To be clear, the second law generalisation exists (Eq. (4.44)), but could not be identified by the observation of the mean transient terms in, for example, the discussion of transitions between stationary states, which one might expect since the second law generalisation involves only mean quantities.

Chapter 5

Entropy Production and its Constituent Contributions for Systems with Odd and Even Time Reversal Behaviour in Continuous Systems

We now provide a development along the lines of the previous chapter, this time however, in the context of continuous behaviour. Again we consider a general set of variables $\boldsymbol{x} = (x_1, x_2, \dots x_N)$ that may be odd or even under time reversal $\boldsymbol{\varepsilon x} = (\varepsilon_1 x_1, \varepsilon_2 x_2, \dots \varepsilon_N x_N)$. Specifically, we consider continuous Markovian dynamics described by a system of arbitrary correlated Itō stochastic differential equations (SDEs) such that the evolution of the coordinates $\boldsymbol{x}$, in vector notation such that $\boldsymbol{x}$ and $\boldsymbol{A}$ are vectors, $\boldsymbol{B}$ a matrix and $d\boldsymbol{W}$ a vector of independent uncorrelated Wiener processes, is given as

$$d\boldsymbol{x} = \boldsymbol{A}(\boldsymbol{x}, \boldsymbol{\lambda}(t))dt + \boldsymbol{B}(\boldsymbol{x}, \boldsymbol{\lambda}(t))d\boldsymbol{W}. \tag{5.1}$$

Such a description may then account for a wide range of non-equilibrium behaviour such as non-conservative forcing, state dependent diffusion and also correlated diffusion [113]. Since the evolution of $\boldsymbol{x}$ is given explicitly we take it to depend only on $\boldsymbol{x}$ and a generalised notion of the protocol, written for brevity as $\boldsymbol{\lambda}^{\mathrm{F}}(t) \equiv \boldsymbol{\lambda}(t) = (\lambda_1(t), \lambda_2(t) \dots \lambda_M(t))$, which characterises, entirely, all other functional dependence of $\boldsymbol{A}$ and $\boldsymbol{B}$ not accounted for by $\boldsymbol{x}$. Similarly we then define a time reversed protocol by acknowledging $\boldsymbol{\varepsilon\lambda}(t) = (\varepsilon_1\lambda_1(t), \varepsilon_2\lambda_2(t) \dots \varepsilon_M\lambda_M(t))$. Since we allow x_i and λ_i to be either odd or even under time reversal we can divide the deterministic dynamics into reversible and irreversible components [46] such that

$$dx_i = A_i^{\mathrm{rev}}(\boldsymbol{x}, \boldsymbol{\lambda}(t))dt + A_i^{\mathrm{ir}}(\boldsymbol{x}, \boldsymbol{\lambda}(t))dt + \sum_j B_{ij}(\boldsymbol{x}, \boldsymbol{\lambda}(t))dW_j \tag{5.2}$$

by defining

$$A_i^{\mathrm{ir}}(\boldsymbol{x}, \boldsymbol{\lambda}(t)) = \frac{1}{2}\left(A_i(\boldsymbol{x}, \boldsymbol{\lambda}(t)) + \varepsilon_i A_i(\boldsymbol{\varepsilon x}, \boldsymbol{\varepsilon\lambda}(t))\right) = \varepsilon_i A_i^{\mathrm{ir}}(\boldsymbol{\varepsilon x}, \boldsymbol{\varepsilon\lambda}(t)) \tag{5.3}$$

$$A_i^{\mathrm{rev}}(\boldsymbol{x}, \boldsymbol{\lambda}(t)) = \frac{1}{2}\left(A_i(\boldsymbol{x}, \boldsymbol{\lambda}(t)) - \varepsilon_i A_i(\boldsymbol{\varepsilon x}, \boldsymbol{\varepsilon\lambda}(t))\right) = -\varepsilon_i A_i^{\mathrm{rev}}(\boldsymbol{\varepsilon x}, \boldsymbol{\varepsilon\lambda}(t)). \tag{5.4}$$

We may describe such a system by using the appropriate Fokker-Planck equation with diffusion matrix

$$\boldsymbol{D}(\boldsymbol{x},\boldsymbol{\lambda}(t)) = \frac{1}{2}\boldsymbol{B}(\boldsymbol{x},\boldsymbol{\lambda}(t))\boldsymbol{B}(\boldsymbol{x},\boldsymbol{\lambda}(t))^{\mathrm{T}} \tag{5.5}$$

such that

$$\frac{\partial p(\boldsymbol{x},t)}{\partial t} = -\sum_i \frac{\partial}{\partial x_i}\left(A_i(\boldsymbol{x},\boldsymbol{\lambda}(t))p(\boldsymbol{x},t)\right) + \sum_{i,j}\frac{\partial^2}{\partial x_i \partial x_j}\left(D_{ij}(\boldsymbol{x},\boldsymbol{\lambda}(t))p(\boldsymbol{x},t)\right) \tag{5.6}$$

where again for brevity we consider $p(\boldsymbol{x},t) \equiv p^{\mathrm{F}}(\boldsymbol{x},t)$. We will find it helpful, subsequently, to express this Fokker-Planck equation as a continuity equation in terms of the vector probability density current $\boldsymbol{J}(\boldsymbol{x},t)$

$$\begin{aligned}\frac{\partial p(\boldsymbol{x},t)}{\partial t} &= -\nabla\cdot\boldsymbol{J}(\boldsymbol{x},t)\\ &= -\nabla\cdot\left(\boldsymbol{J}^{\mathrm{ir}}(\boldsymbol{x},t) + \boldsymbol{J}^{\mathrm{rev}}(\boldsymbol{x},t)\right)\end{aligned} \tag{5.7}$$

which we separate into irreversible and reversible components. These too take vector form such that

$$\begin{aligned}J_i^{\mathrm{ir}}(\boldsymbol{x},t) &= A_i^{\mathrm{ir}}(\boldsymbol{x},\boldsymbol{\lambda}(t))p(\boldsymbol{x},t) - \sum_j \frac{\partial}{\partial x_j}\left(D_{ij}(\boldsymbol{x},\boldsymbol{\lambda}(t))p(\boldsymbol{x},t)\right)\\ J_i^{\mathrm{rev}}(\boldsymbol{x},t) &= A_i^{\mathrm{rev}}(\boldsymbol{x},\boldsymbol{\lambda}(t))p(\boldsymbol{x},t).\end{aligned} \tag{5.8}$$

Using the same involution definitions $\boldsymbol{x}^\dagger(t) = \boldsymbol{\varepsilon}\boldsymbol{x}(\tau - t)$, $\boldsymbol{x}^{\mathrm{R}}(t) = \boldsymbol{x}(\tau - t)$ and $\boldsymbol{x}^{\mathrm{T}}(t) = \boldsymbol{\varepsilon}\boldsymbol{x}(t)$, we may construct the relevant dimensionless entropy changes. Since we now are considering systems in both continuous space and time the entropy changes now are constructed, as in Sect. 3.1, from path probability functionals. We write the total path probability functional as $\mathcal{P}[\vec{\boldsymbol{x}}]$ which may be divided into an initial probability density function, which we write $p(\boldsymbol{x}(t),t)$ taken to be the instantaneous solution to the Fokker-Planck equation, and the path probability functional which we write $\mathcal{P}[\vec{\boldsymbol{x}}|\boldsymbol{x}(0)]$. The unitless contributions to entropy production, which then become thermodynamically meaningful when multiplied by k_B, are given, for a process starting at $t = 0$ and of duration τ, as:

$$\begin{aligned}\Delta\mathcal{S}_{\mathrm{tot}} &= \ln\mathcal{P}^{\mathrm{F}}[\vec{\boldsymbol{x}}]_{p_0^{\mathrm{F}}} - \ln\mathcal{P}^{\dagger}[\vec{\boldsymbol{x}}^{\dagger}]_{\hat{\boldsymbol{\varepsilon}}p_\tau^{\mathrm{F}}}\\ &= \ln\frac{p(\boldsymbol{x}(0),0)}{p(\boldsymbol{x}(\tau),\tau)} + \ln\frac{\mathcal{P}^{\mathrm{F}}[\vec{\boldsymbol{x}}|\boldsymbol{x}(0)]}{\mathcal{P}^{\dagger}[\vec{\boldsymbol{x}}^{\dagger}|\boldsymbol{\varepsilon}\boldsymbol{x}(\tau)]},\end{aligned} \tag{5.9}$$

$$\begin{aligned}\Delta\mathcal{S}_1 &= \ln\mathcal{P}^{\mathrm{F}}[\vec{\boldsymbol{x}}]_{p_0^{\mathrm{F}}} - \ln\mathcal{P}^{\mathrm{R,ad}}[\vec{\boldsymbol{x}}^{\mathrm{R}}]_{p_\tau^{\mathrm{F}}}\\ &= \ln\frac{p(\boldsymbol{x}(0),0)}{p(\boldsymbol{x}(\tau),\tau)} + \ln\frac{\mathcal{P}^{\mathrm{F}}[\vec{\boldsymbol{x}}|\boldsymbol{x}(0)]}{\mathcal{P}^{\mathrm{R,ad}}[\vec{\boldsymbol{x}}^{\mathrm{R}}|\boldsymbol{x}(\tau)]}\end{aligned} \tag{5.10}$$

and

$$\begin{aligned}\Delta\mathcal{S}_2 &= \ln\mathcal{P}^{\mathrm{F}}[\vec{\boldsymbol{x}}]_{p_0^{\mathrm{F}}} - \ln\mathcal{P}^{\mathrm{T,ad}}[\vec{\boldsymbol{x}}^{\mathrm{T}}]_{\hat{\boldsymbol{\varepsilon}}p_0^{\mathrm{F}}}\\ &= \ln\frac{p(\boldsymbol{x}(0),0)}{p(\boldsymbol{x}(0),0)} + \ln\frac{\mathcal{P}^{\mathrm{F}}[\vec{\boldsymbol{x}}|\boldsymbol{x}(0)]}{\mathcal{P}^{\mathrm{T,ad}}[\vec{\boldsymbol{x}}^{\mathrm{T}}|\boldsymbol{\varepsilon}\boldsymbol{x}(0)]}.\end{aligned} \tag{5.11}$$

As before the involutions R, T and † are applied to path, protocol and distribution, but have had the involutions explicitly performed for the initial distribution. Similarly 'ad' designates that the dynamics are adjoint with respect to the forward dynamics. All three are expected to obey the relevant fluctuation

theorems by the nature of their form. Then, by the construction $\Delta\mathcal{S}_{\rm tot} = \Delta\mathcal{S}_1 + \Delta\mathcal{S}_2 + \Delta\mathcal{S}_3$, we have

$$\begin{aligned}\Delta\mathcal{S}_3 &= \ln \mathcal{P}^{\rm T,ad}[\vec{\boldsymbol{x}}^{\rm T}]_{\hat{\boldsymbol{\varepsilon}}p_0^{\rm F}} + \ln \mathcal{P}^{\rm R,ad}[\vec{\boldsymbol{x}}^{\rm R}]_{p_\tau^{\rm F}} - \ln \mathcal{P}^{\rm F}[\vec{\boldsymbol{x}}]_{p_0^{\rm F}} - \ln \mathcal{P}^{\dagger}[\vec{\boldsymbol{x}}^{\dagger}]_{\hat{\boldsymbol{\varepsilon}}p_\tau^{\rm F}} \\ &= \ln \frac{\mathcal{P}^{\rm R,ad}[\vec{\boldsymbol{x}}^{\rm R}|\boldsymbol{x}(\tau)]\mathcal{P}^{\rm T,ad}[\vec{\boldsymbol{x}}^{\rm T}|\boldsymbol{\varepsilon}\boldsymbol{x}(0)]}{\mathcal{P}^{\rm F}[\vec{\boldsymbol{x}}|\boldsymbol{x}(0)]\mathcal{P}^{\dagger}[\vec{\boldsymbol{x}}^{\dagger}|\boldsymbol{\varepsilon}\boldsymbol{x}(\tau)]}\end{aligned} \tag{5.12}$$

which cannot be expressed in the form of Eq. (4.2) and so does not obey an IFT. Again we may divide the entropy production into a system and medium entropy change, which for a general thermal system with a phase space or time dependent temperature can be defined in terms of a heat flow so that

$$\begin{aligned}\Delta\mathcal{S}_{\rm tot} &= k_B \ln \frac{p(\boldsymbol{x}(0),0)}{p(\boldsymbol{x}(\tau),\tau)} + \int_{t=0}^{t=\tau} d\left(\frac{Q}{T(\boldsymbol{x}(t),t)}\right) \\ &= \Delta\mathcal{S}_{\rm sys} + \Delta\mathcal{S}_{\rm med},\end{aligned} \tag{5.13}$$

where $d(Q/T)$ is the microscopic medium entropy production for one of, in principle, an arbitrarily large number of independent equilibrium heat baths all at fixed temperature with each experiencing a microscopic heat flow divisible into components $dQ = dQ_{\rm ex} + dQ_{\rm hk} = dQ_{\rm ex} + dQ_{\rm hk,G} + dQ_{\rm hk,T}$ as before.

At this point it should be mentioned that unlike for discrete stochastic systems, the situations where the integrals over $d\vec{\boldsymbol{x}}$ and $d\vec{\boldsymbol{x}}^*$ are the same and ergodic consistency is achieved (such that one arrives at fluctuation theorems) are not as straightforward to define. One, however, can invoke the so-called Girsanov theorem which states, for such continuous systems, under what conditions the probability measures of the forward and alternative processes are equivalent [38, 114] which in turn guarantees the above two conditions. In one dimension we require that the noise strengths experienced by the forward and alternative paths at times that contribute to the same point in the functional $\mathcal{A}[\vec{\boldsymbol{x}}]$ are identical. Examining the form of the path integrals used in the formulation of the fluctuation theorems we understand that all these properties coincide when the relevant pre-exponential factors in the short time propagators, which depend on the noise strength and form the path integral measure, are invariant under the path and protocol transformation. In higher dimensions the condition translates to the determinant of the diffusion matrix appearing in the relevant short time propagator (Eq. (2.142)) being invariant under the path and protocol transformation. Sufficient conditions for such a property involve enforcing the parity dependent behaviour of the diffusion matrix to be $D_{ij}(\boldsymbol{\varepsilon}\boldsymbol{x}, \boldsymbol{\varepsilon}\boldsymbol{\lambda}(t)) = \varepsilon_i\varepsilon_j D_{ij}(\boldsymbol{x}, \boldsymbol{\lambda}(t))$ and for the path and the protocol to undergo the same transformation (that is we cannot allow $\boldsymbol{x}^*(t) = \boldsymbol{x}(t)$ if $\boldsymbol{\lambda}^*(t) = \boldsymbol{\lambda}(\tau - t)$, for example), but clearly these conditions can be relaxed should there be additive noise. It is therefore important to note that all the above quantities, under the assumption $D_{ij}(\boldsymbol{\varepsilon}\boldsymbol{x}, \boldsymbol{\varepsilon}\boldsymbol{\lambda}(t)) = \varepsilon_i\varepsilon_j D_{ij}(\boldsymbol{x}, \boldsymbol{\lambda}(t))$, are defined in such a way that ensures this.

Since we are describing the dynamics using SDEs it is sensible to seek a description of a small increment in each entropy production contribution given an increment in the underlying variables $\boldsymbol{x}' - \boldsymbol{x} = \boldsymbol{x}(t + dt) - \boldsymbol{x}(t)$ in a small time dt. By considering appropriate involutions applied to the protocol and increments of the paths we identify, from Eqs. (5.9), (5.10), (5.11) and (5.12), in terms of the short term

propagators for the forward and adjoint dynamics,

$$d\Delta\mathcal{S}_{\rm tot} = -d(\ln(p)) + \ln\frac{p(\boldsymbol{x}', t+dt|\boldsymbol{x}, t; \boldsymbol{\lambda}(t))}{p(\varepsilon\boldsymbol{x}, t+dt|\varepsilon\boldsymbol{x}', t; \varepsilon\boldsymbol{\lambda}(t))} + O(dt^{3/2}) \tag{5.14}$$

$$d\Delta\mathcal{S}_{1} = -d(\ln(p)) + \ln\frac{p(\boldsymbol{x}', t+dt|\boldsymbol{x}, t; \boldsymbol{\lambda}(t))}{p^{\rm ad}(\boldsymbol{x}, t+dt|\boldsymbol{x}', t; \boldsymbol{\lambda}(t))} + O(dt^{3/2}) \tag{5.15}$$

$$d\Delta\mathcal{S}_{2} = \ln\frac{p(\boldsymbol{x}', t+dt|\boldsymbol{x}, t; \boldsymbol{\lambda}(t))}{p^{\rm ad}(\varepsilon\boldsymbol{x}', t+dt|\varepsilon\boldsymbol{x}, t; \varepsilon\boldsymbol{\lambda}(t))} + O(dt^{3/2}) \tag{5.16}$$

$$d\Delta\mathcal{S}_{3} = \ln\frac{p^{\rm ad}(\boldsymbol{x}, t+dt|\boldsymbol{x}', t; \boldsymbol{\lambda}(t))p^{\rm ad}(\varepsilon\boldsymbol{x}', t+dt|\varepsilon\boldsymbol{x}, t; \varepsilon\boldsymbol{\lambda}(t))}{p(\boldsymbol{x}', t+dt|\boldsymbol{x}, t; \boldsymbol{\lambda}(t))p(\varepsilon\boldsymbol{x}, t+dt|\varepsilon\boldsymbol{x}', t; \varepsilon\boldsymbol{\lambda}(t))} + O(dt^{3/2}) \tag{5.17}$$

where we ignore $O(dt^{3/2})$ terms owing to the stochastic nature of $\boldsymbol{x}$. We point out the validity of approximating the path probability functionals with short time propagators, despite them including pre-exponential normalisation terms which usually form the path integral measure, since we are only considering cases where these normalising terms must cancel in line with Girsanov's theorem. We also note the abbreviation $d(\ln(p)) = \ln(p(\boldsymbol{x}(t+dt), t+dt)/p(\boldsymbol{x}(t), t))$ and thereby establish the procedure for finding, if not the explicit form of, the SDEs that describe the total entropy production and its three contributions.

5.1 Constructing an SDE for Total Entropy Production

To proceed we utilise the short time Green's function or 'short time propagator' [47] given in Sect. 2.3.2 so that for displacement $d\boldsymbol{x} = \boldsymbol{x}' - \boldsymbol{x}$, and corresponding displacements in each dimension $dx_i = x_i' - x_i$, in a time dt we have the conditional probability density of the form

$$\begin{aligned} p(\boldsymbol{x}', t+dt|\boldsymbol{x}, t; \boldsymbol{\lambda}(t)) = &\sqrt{\frac{1}{(4\pi dt)^N\,|\boldsymbol{D}(\boldsymbol{r}, \boldsymbol{\lambda}(t))|}} \\ &\times\exp\left[\sum_{i,j} -\frac{1}{4dt}\left(dx_i - A_i(\boldsymbol{r}, \boldsymbol{\lambda}(t))dt + \sum_m 2a\frac{\partial D_{im}(\boldsymbol{r}, \boldsymbol{\lambda}(t))}{\partial r_m}dt\right)D^{-1}_{ij}(\boldsymbol{r}, \boldsymbol{\lambda}(t))\right. \\ &\left.\times\left(dx_j - A_j(\boldsymbol{r}, \boldsymbol{\lambda}(t))dt + \sum_n 2a\frac{\partial D_{jk}(\boldsymbol{r}, \boldsymbol{\lambda}(t))}{\partial r_k}dt\right) - \sum_i adt\frac{\partial A_i(\boldsymbol{r}, \boldsymbol{\lambda}(t))}{\partial r_i} + \sum_{i,j} a^2 dt\frac{\partial^2 D_{ij}(\boldsymbol{r}, \boldsymbol{\lambda}(t))}{\partial r_i \partial r_j}\right] \end{aligned} \tag{5.18}$$

where a is the parameter ranging from 0 to 1 which defines the evaluation point of certain terms in the propagator $\boldsymbol{r} = a\boldsymbol{x}' + (1-a)\boldsymbol{x}$ and $r_i = ax_i' + (1-a)x_i$ for which all choices are correct to first order in dt and lead to the same Fokker-Planck equation.

We consider the incremental contribution to total entropy production comprising the conditional probabilities, namely the medium entropy change in the formalism of Seifert and so write

$$d\Delta\mathcal{S}_{\rm med} = \ln\frac{p(\boldsymbol{x}', t+dt|\boldsymbol{x}, t; \boldsymbol{\lambda}(t))}{p(\varepsilon\boldsymbol{x}, t+dt|\varepsilon\boldsymbol{x}', t; \varepsilon\boldsymbol{\lambda}(t))}. \tag{5.19}$$

By employing the appropriate reverse short time propagator

$$p(\boldsymbol{\varepsilon x},t+dt|\boldsymbol{\varepsilon x}',t;\boldsymbol{\varepsilon\lambda}(t))=\sqrt{\frac{1}{(4\pi dt)^N\left|\boldsymbol{D}(\boldsymbol{\varepsilon r}',\boldsymbol{\varepsilon\lambda}(t))\right|}}$$
$$\exp\left[-\sum_{i,j}\frac{1}{4dt}\left(-\varepsilon_i dx_i-A_i(\boldsymbol{\varepsilon r}',\boldsymbol{\varepsilon\lambda}(t))dt+\sum_m 2b\frac{\partial D_{im}(\boldsymbol{\varepsilon r}',\boldsymbol{\varepsilon\lambda}(t))}{\partial(\varepsilon_m r'_m)}dt\right)D^{-1}_{ij}(\boldsymbol{\varepsilon r}',\boldsymbol{\varepsilon\lambda}(t))\right.$$
$$\times\left(-\varepsilon_j dx_j-A_j(\boldsymbol{\varepsilon r}',\boldsymbol{\varepsilon\lambda}(t))dt+\sum_n 2b\frac{\partial D_{jn}(\boldsymbol{\varepsilon r}',\boldsymbol{\varepsilon\lambda}(t))}{\partial(\varepsilon_n r'_n)}dt\right)$$
$$\left.-\sum_i bdt\frac{\partial A_i(\boldsymbol{\varepsilon r}',\boldsymbol{\varepsilon\lambda}(t))}{\partial(\varepsilon_i r'_i)}+\sum_{i,j}b^2dt\frac{\partial^2 D_{ij}(\boldsymbol{\varepsilon r}',\boldsymbol{\varepsilon\lambda}(t))}{\partial(\varepsilon_i r'_i)\partial(\varepsilon_j r'_j)}\right] \tag{5.20}$$

where b ranges from 0 to 1 and corresponds to the evaluation point such that $\boldsymbol{r}'=b\boldsymbol{x}+(1-b)\boldsymbol{x}'$ and $r'_i=bx_i+(1-b)x'_i$. Using Eqs. (5.3) and (5.4) along with the assumption $D_{ij}(\boldsymbol{\varepsilon x},\boldsymbol{\varepsilon\lambda}(t))=\varepsilon_i\varepsilon_j D_{ij}(\boldsymbol{x},\boldsymbol{\lambda}(t))$ we may write

$$p(\boldsymbol{\varepsilon x},t+dt|\boldsymbol{\varepsilon x}',t;\boldsymbol{\varepsilon\lambda}(t))=\sqrt{\frac{1}{(4\pi dt)^N\left|\boldsymbol{D}(\boldsymbol{r}',\boldsymbol{\lambda}(t))\right|}}$$
$$\exp\left[-\sum_{i,j}\frac{1}{4dt}\left(-dx_i-(A^{\mathrm{ir}}_i(\boldsymbol{r}',\boldsymbol{\lambda}(t))-A^{\mathrm{rev}}(\boldsymbol{r}',\boldsymbol{\lambda}(t)))dt+\sum_m 2b\frac{\partial D_{im}(\boldsymbol{r}',\boldsymbol{\lambda}(t))}{\partial(r'_m)}dt\right)D^{-1}_{ij}(\boldsymbol{r}',\boldsymbol{\lambda}(t))\right.$$
$$\times\left(-dx_j-(A^{\mathrm{ir}}_j(\boldsymbol{r}',\boldsymbol{\lambda}(t))-A^{\mathrm{rev}}_j(\boldsymbol{r}',\boldsymbol{\lambda}(t)))dt+\sum_n 2b\frac{\partial D_{jn}(\boldsymbol{r}',\boldsymbol{\lambda}(t))}{\partial(r'_n)}dt\right)$$
$$\left.-\sum_i bdt\left(\frac{\partial A^{\mathrm{ir}}_i(\boldsymbol{r}',\boldsymbol{\lambda}(t))}{\partial(r'_i)}-\frac{\partial A^{\mathrm{rev}}_i(\boldsymbol{r}',\boldsymbol{\lambda}(t))}{\partial(r'_i)}\right)+\sum_{i,j}b^2dt\frac{\partial^2 D_{ij}(\boldsymbol{r}',\boldsymbol{\lambda}(t))}{\partial r'_i\partial r'_j}\right]. \tag{5.21}$$

One can then proceed by evaluating the ratio of the propagators recasting them so that they utilise the same evaluation point in order to define a consistent SDE. This is achieved, for an expansion in Itō form, by using the heuristic rules (Sect. 2.2.2)

$$f(\boldsymbol{r})dt=f(\boldsymbol{x})dt+O(dt^{3/2}) \tag{5.22}$$

$$f(\boldsymbol{r}')dt=f(\boldsymbol{x})dt+O(dt^{3/2}) \tag{5.23}$$

$$f(\boldsymbol{r})dx_i=f(\boldsymbol{x})dx_i+\sum_m 2aD_{im}(\boldsymbol{x})\frac{\partial f(\boldsymbol{x})}{\partial x_m}dt+O(dt^{3/2}) \tag{5.24}$$

$$f(\boldsymbol{r}')dx_i=f(\boldsymbol{x})dx_i+\sum_m 2(1-b)D_{im}(\boldsymbol{x})\frac{\partial f(\boldsymbol{x})}{\partial x_m}dt+O(dt^{3/2}). \tag{5.25}$$

At this point we stress that the forward propagator leads to the correct path probability independently of the choice a and the reverse propagator leads to the correct path probability independently of the choice b, however *when there is multiplicative noise* in the dynamics the increment in medium entropy change does depend on the choice of a and b. This is a manifestation of the ambiguity in stochastic calculus arising from the unbounded variation in $\boldsymbol{x}(t)$ and corresponding lack of smoothness on any timescale. As such, despite considering an infinitesimal time step, the different evaluation points exploit differences in the correlation of the terms in the propagator and $\boldsymbol{x}$ and $\boldsymbol{x}'$ that persists on all timescales. Crucially, we

find that it is overcome by evaluating the point $\boldsymbol{r}'$ at the same point $\boldsymbol{r}$ equivalent to the choice $b = 1-a$ whereupon the dependence on a and b disappears. One may think of this in a number of ways. One may consider that there should be a jump process which exists as the limit of such a continuous process where the incremental medium entropy change is evaluated in numerator and denominator for a single consistent functional dependence (i.e. a single choice $\boldsymbol{r}' = \boldsymbol{r}$) or that the path involution $\boldsymbol{x}^\dagger = \boldsymbol{\varepsilon x}(\tau - t)$ persists on a sub-infinitesimal scale necessitating the choice $b = 1 - a$. To clarify, in such an approach choices may include Stratonovich-like evaluation ($a = b = 1/2$) for both propagators in Eq. (5.19), a choice which is sometimes implicitly used by authors [33, 111] within integrated Onsager-Machlup approaches (with some exceptions which would break down should multiplicative noise be utilised [75]). Strictly, however, this does not preclude other choices of a and b in the construction of the fundamental SDEs such as, for example, an Itō prescription ($a = 0$) in the forward propagator and a Hänggi-Klimontovich ($b = 1$) in the backwards propagator and indeed with all independent choices of a and b being valid if the noise is additive. In summary, all the evaluation points lead to the correct *path probability* when supplemented with the correct multiplication scheme, but that if one has multiplicative noise, the correct representation of the *entropy production* requires the more exact relation between the evaluation points.

Choosing for simplicity $a = b = 1/2$, such that all multiplications of the form $f(\boldsymbol{r}) * dx_i$ arising from an explicit expansion of the short time propagator should be interpreted according to Stratonovich rules $f(\boldsymbol{x}) \circ dx_i$, we retain terms to first order in dt and utilise the abbreviation $f(\boldsymbol{x}) \equiv f(\boldsymbol{x}, \boldsymbol{\lambda}(t))$ to obtain

$$\begin{aligned} d\Delta\mathcal{S}_{\text{med}} = &\sum_i -\frac{\partial A^{\text{rev}}(\boldsymbol{x})}{\partial x_i}dt + \sum_{i,j}\frac{D^{-1}_{ij}(\boldsymbol{x})}{2}\left(A^{\text{ir}}_i(\boldsymbol{x}) \circ dx_j + A^{\text{ir}}_j(\boldsymbol{x}) \circ dx_i\right) \\ &- \frac{D^{-1}_{ij}(\boldsymbol{x})}{2}\left(A^{\text{rev}}_i(\boldsymbol{x})A^{\text{ir}}_j(\boldsymbol{x}) + A^{\text{rev}}_j(\boldsymbol{x})A^{\text{ir}}_i(\boldsymbol{x})\right)dt \\ &- \frac{D^{-1}_{ij}(\boldsymbol{x})}{2}\left(\left(\sum_n \frac{\partial D_{jn}(\boldsymbol{x})}{\partial x_n}\right) \circ dx_i + \left(\sum_m \frac{\partial D_{im}(\boldsymbol{x})}{\partial x_m}\right) \circ dx_j\right) \\ &+ \frac{D^{-1}_{ij}(\boldsymbol{x})}{2}\left(A^{\text{rev}}_j(\boldsymbol{x})\left(\sum_m \frac{\partial D_{im}(\boldsymbol{x})}{\partial x_m}\right) + A^{\text{rev}}_i(\boldsymbol{x})\left(\sum_n \frac{\partial D_{jn}(\boldsymbol{x})}{\partial x_n}\right)\right)dt. \end{aligned} \tag{5.26}$$

This in turn leads to

$$\begin{aligned} d\Delta\mathcal{S}_{\text{tot}} = &-\frac{1}{p(\boldsymbol{x})}\frac{\partial p(\boldsymbol{x})}{\partial t}dt - \sum_i \frac{1}{p(\boldsymbol{x})}\frac{\partial p(\boldsymbol{x})}{\partial x_i} \circ dx_i \\ &- \sum_i \frac{\partial A^{\text{rev}}(\boldsymbol{x})}{\partial x_i}dt + \sum_{i,j}\frac{D^{-1}_{ij}(\boldsymbol{x})}{2}\left(A^{\text{ir}}_i(\boldsymbol{x}) \circ dx_j + A^{\text{ir}}_j(\boldsymbol{x}) \circ dx_i\right) \\ &- \frac{D^{-1}_{ij}(\boldsymbol{x})}{2}\left(A^{\text{rev}}_i(\boldsymbol{x})A^{\text{ir}}_j(\boldsymbol{x}) + A^{\text{rev}}_j(\boldsymbol{x})A^{\text{ir}}_i(\boldsymbol{x})\right)dt \\ &- \frac{D^{-1}_{ij}(\boldsymbol{x})}{2}\left(\left(\sum_n \frac{\partial D_{jn}(\boldsymbol{x})}{\partial x_n}\right) \circ dx_i + \left(\sum_m \frac{\partial D_{im}(\boldsymbol{x})}{\partial x_m}\right) \circ dx_j\right) \\ &+ \frac{D^{-1}_{ij}(\boldsymbol{x})}{2}\left(A^{\text{rev}}_j(\boldsymbol{x})\left(\sum_m \frac{\partial D_{im}(\boldsymbol{x})}{\partial x_m}\right) + A^{\text{rev}}_i(\boldsymbol{x})\left(\sum_n \frac{\partial D_{jn}(\boldsymbol{x})}{\partial x_n}\right)\right)dt. \end{aligned} \tag{5.27}$$

This is a very general and robust definition of the entropy production for continuous stochastic behaviour and can be thought of as a generalisation of the pioneering approach in [12] wherein the equation of motion for entropy essentially describes $d\Delta\mathcal{S}_{\text{tot}}$ for a specific system with additive noise leading to uncorrelated diffusion, even variables and dynamics ($\boldsymbol{\varepsilon x} = \boldsymbol{x}$ and $\boldsymbol{\varepsilon\lambda}(t) = \boldsymbol{\lambda}(t)$) and implicitly using

Stratonovich rules.

We point out that such a construction allows us to consider purely deterministic coordinates ($D_{ij}(\boldsymbol{x}) = 0 \ \forall j$) as would apply, for example, to the case of spatial coordinates within a full phase space Langevin description. In such coordinates $D_{ij}(\boldsymbol{x})$ is assumed constant and taken to zero. The remaining terms then clearly diverge unless we demand $A_i^{\mathrm{ir}}(\boldsymbol{x}) = 0$ since in these instances, for the reverse path to be a solution to the forward dynamics the motion must be purely reversible. This condition simply amounts to the requirement that the reverse path exists. There is, however, a contribution to the medium entropy change that persists due to the dynamics of these coordinates, since deterministic phase space density distributions can evolve in time, as described by the deterministic limit of the medium entropy formulated from probability *densities* in continuous space, even if the phase space volume, analogous to probabilities, cannot. The contribution to the medium entropy change due to the deterministic behaviour of these coordinates is

$$\Delta\mathcal{S}_{\mathrm{med,det}} = -\frac{\partial A_i^{\mathrm{rev}}(\boldsymbol{x})}{\partial x_i}dt, \tag{5.28}$$

a result that provides an insight into the similarities and differences between stochastic and deterministic measures of irreversibility: it is demonstrably equal to the phase space contraction found in non-linear dynamical systems, which is associated with the heat transfer to the environment brought about by thermostatting terms in such approaches. This leads to a quantity that is positive in the mean for deterministic systems: the dissipation function [4]. We point out, however that *total entropy production*, as defined here for stochastic systems, is zero for deterministic dynamics. This is because the change in the system entropy would be equal and opposite to the change in medium entropy, technically since it involves probability densities at the start and end of the process. In contrast the dissipation function can provide a measure of irreversibility because it involves a comparison of trajectories originating *from the same* starting distribution. This contrast is to be expected as the total entropy production, as defined for the systems we consider, arises from explicit irreversibility in the dynamics, which deterministic, reversible equations do not provide.

5.2 Constructing the Instantaneous Average Entropy Production Rate

Frequently the average entropy production rate is argued to be proportional to the mean probability flux squared, as derived, for example, by taking the time derivative of the Gibbs entropy of a system, and identifying an evidently positive contribution as the total entropy production rate and the remainder as the (negative) medium entropy production rate [73, 98, 115]. We prefer however, to derive the average contributions directly from the SDEs so that we can avoid arbitrarily identifying a positive contribution with a quantity expected to obey an IFT: strictly there is no guarantee such a division is unique, as another description shows [116]. To do so is straightforward and requires us to find the average increment in $\Delta\mathcal{S}_{\mathrm{tot}}$ by means of the integral

$$\langle d\Delta\mathcal{S}_{\mathrm{tot}}\rangle_{p_0^{\mathrm{F}}}^{\mathrm{F}} = \int d\boldsymbol{x}\int d\boldsymbol{x}'\ p(\boldsymbol{x},t)p(\boldsymbol{x}',t+dt|\boldsymbol{x},t)d\Delta\mathcal{S}_{\mathrm{tot}}. \tag{5.29}$$

The benefit of such a formulation is that we may characterise $d\Delta\mathcal{S}_{\mathrm{tot}}$ using an Itō SDE based on the underlying relations $dx_i = A_i dt + B_i dW_i$ and then use the martingale property of the Itō stochastic integral $\langle B_i dW_i\rangle = 0$ since B_i is non-anticipating, such that we can simplify the integral in Eq. (5.29)

by writing

$$\langle d\Delta\mathcal{S}_{\rm tot}\rangle^{\rm F}_{p_0^{\rm F}} = \int d\boldsymbol{x}\, p(\boldsymbol{x}) \langle d\Delta\mathcal{S}_{\rm tot}|\boldsymbol{x}\rangle \tag{5.30}$$

and evaluating the conditional average $\langle d\Delta\mathcal{S}_{\rm tot}|\boldsymbol{x}\rangle$ by simply replacing all occurrences of dx_i with $(A_i^{\rm ir} + A_i^{\rm rev})dt$ in $d\Delta\mathcal{S}_{\rm tot}$. To do this, however, we must first convert the SDE for $\Delta\mathcal{S}_{\rm tot}$ into Itō form for which we use the conversion formula (Sect. 2.2.2)

$$f(x) \circ dx_i = f(x)dx_i + \sum_k D_{ik}\frac{\partial f(x)}{\partial x_k}dt. \tag{5.31}$$

For the increment in medium entropy change this gives us

$$\begin{aligned}
d\Delta\mathcal{S}_{\rm med} = &\sum_i -\frac{\partial A_i^{\rm rev}(\boldsymbol{x})}{\partial x_i}dt + \sum_{i,j}\Bigg\{\frac{D_{ij}^{-1}(\boldsymbol{x})}{2}\left(A_i^{\rm ir}(\boldsymbol{x})dx_j + A_j^{\rm ir}(\boldsymbol{x})dx_i\right)\\
&-\frac{D_{ij}^{-1}(\boldsymbol{x})}{2}\left(\left(\sum_n \frac{\partial D_{jn}(\boldsymbol{x})}{\partial x_n}\right)dx_i + \left(\sum_m \frac{\partial D_{im}(\boldsymbol{x})}{\partial x_m}\right)dx_j\right)\\
&-\frac{D_{ij}^{-1}(\boldsymbol{x})}{2}\left(A_i^{\rm rev}(\boldsymbol{x})A_j^{\rm ir}(\boldsymbol{x}) + A_j^{\rm rev}(\boldsymbol{x})A_i^{\rm ir}(\boldsymbol{x})\right)dt\\
&+\frac{D_{ij}^{-1}(\boldsymbol{x})}{2}\left(A_j^{\rm rev}(\boldsymbol{x})\left(\sum_m \frac{\partial D_{im}(\boldsymbol{x})}{\partial x_m}\right) + A_i^{\rm rev}(\boldsymbol{x})\left(\sum_n \frac{\partial D_{jn}(\boldsymbol{x})}{\partial x_n}\right)\right)dt\\
&+\frac{1}{2}\sum_k\Bigg[D_{ik}(\boldsymbol{x})\frac{\partial}{\partial x_k}\left(D_{ij}^{-1}(\boldsymbol{x})A_j^{\rm ir}(\boldsymbol{x})\right) + D_{jk}(\boldsymbol{x})\frac{\partial}{\partial x_k}\left(D_{ij}^{-1}(\boldsymbol{x})A_i^{\rm ir}(\boldsymbol{x})\right)\\
&-D_{ik}(\boldsymbol{x})\frac{\partial}{\partial x_k}\left(D_{ij}^{-1}(\boldsymbol{x})\left(\sum_n \frac{\partial D_{jn}(\boldsymbol{x})}{\partial x_n}\right)\right) - D_{jk}(\boldsymbol{x})\frac{\partial}{\partial x_k}\left(D_{ij}^{-1}(\boldsymbol{x})\left(\sum_m \frac{\partial D_{im}(\boldsymbol{x})}{\partial x_m}\right)\right)\Bigg]dt\Bigg\}.
\end{aligned} \tag{5.32}$$

We may then average over all realisations using the procedure described above to obtain

$$\begin{aligned}
\langle d\Delta\mathcal{S}_{\rm med}\rangle^{\rm F}_{p_0^{\rm F}} = &\int\sum_i -p(\boldsymbol{x})\frac{\partial A_i^{\rm rev}(\boldsymbol{x})}{\partial x_i}dt + \sum_{i,j} p(\boldsymbol{x})\Bigg\{\frac{D_{ij}^{-1}(\boldsymbol{x})}{2}\left(A_i^{\rm ir}(\boldsymbol{x})A_j^{\rm ir}(\boldsymbol{x}) + A_j^{\rm ir}(\boldsymbol{x})A_i^{\rm ir}(\boldsymbol{x})\right)dt\\
&-\frac{D_{ij}^{-1}(\boldsymbol{x})}{2}\left(\left(\sum_n \frac{\partial D_{jn}(\boldsymbol{x})}{\partial x_n}\right)A_i^{\rm ir}(\boldsymbol{x}) + \left(\sum_m \frac{\partial D_{im}(\boldsymbol{x})}{\partial x_m}\right)A_j^{\rm ir}(\boldsymbol{x})\right)dt\\
&+\frac{1}{2}\sum_k\Bigg[D_{ik}(\boldsymbol{x})\frac{\partial}{\partial x_k}\left(D_{ij}^{-1}(\boldsymbol{x})A_j^{\rm ir}(\boldsymbol{x})\right) + D_{jk}(\boldsymbol{x})\frac{\partial}{\partial x_k}\left(D_{ij}^{-1}(\boldsymbol{x})A_i^{\rm ir}(\boldsymbol{x})\right)\\
&-D_{ik}(\boldsymbol{x})\frac{\partial}{\partial x_k}\left(D_{ij}^{-1}(\boldsymbol{x})\left(\sum_n \frac{\partial D_{jn}(\boldsymbol{x})}{\partial x_n}\right)\right) - D_{jk}(\boldsymbol{x})\frac{\partial}{\partial x_k}\left(D_{ij}^{-1}(\boldsymbol{x})\left(\sum_m \frac{\partial D_{im}(\boldsymbol{x})}{\partial x_m}\right)\right)\Bigg]dt\Bigg\}.
\end{aligned} \tag{5.33}$$

Integrating by parts, dropping boundary terms, applying the product rule and rearranging we get

$$\begin{aligned}\langle d\Delta\mathcal{S}_{\mathrm{med}}\rangle = \int \sum_i -p(\boldsymbol{x})\frac{\partial A_i^{\mathrm{rev}}(\boldsymbol{x})}{\partial x_i}dt + \sum_{i,j} p(\boldsymbol{x})\Bigg\{ \frac{D_{ij}^{-1}(\boldsymbol{x})}{2}\left(A_i^{\mathrm{ir}}(\boldsymbol{x})A_j^{\mathrm{ir}}(\boldsymbol{x}) + A_j^{\mathrm{ir}}(\boldsymbol{x})A_i^{\mathrm{ir}}(\boldsymbol{x})\right)dt \\ -\frac{D_{ij}^{-1}(\boldsymbol{x})}{2}\left(\left(\sum_n \frac{\partial D_{jn}(\boldsymbol{x})}{\partial x_n}\right)A_i^{\mathrm{ir}}(\boldsymbol{x}) + \left(\sum_m \frac{\partial D_{im}(\boldsymbol{x})}{\partial x_m}\right)A_j^{\mathrm{ir}}(\boldsymbol{x})\right)dt\Bigg\} \\ -\frac{D_{ij}^{-1}(\boldsymbol{x})}{2}\Bigg\{A_j^{\mathrm{ir}}(\boldsymbol{x})\left(\sum_k D_{ik}(\boldsymbol{x})\frac{\partial p(\boldsymbol{x})}{\partial x_k}\right) + A_j^{\mathrm{ir}}(\boldsymbol{x})p(\boldsymbol{x})\left(\sum_k \frac{\partial D_{ik}(\boldsymbol{x})}{\partial x_k}\right) \\ +A_i^{\mathrm{ir}}(\boldsymbol{x})\left(\sum_k D_{jk}(\boldsymbol{x})\frac{\partial p(\boldsymbol{x})}{\partial x_k}\right) + A_i^{\mathrm{ir}}(\boldsymbol{x})p(\boldsymbol{x})\left(\sum_k \frac{\partial D_{jk}(\boldsymbol{x})}{\partial x_k}\right) \\ -\left(\sum_k D_{ik}(\boldsymbol{x})\frac{\partial p(\boldsymbol{x})}{\partial x_k}\right)\left(\sum_n \frac{\partial D_{jn}(\boldsymbol{x})}{\partial x_n}\right) - p(\boldsymbol{x})\left(\sum_k \frac{\partial D_{ik}(\boldsymbol{x})}{\partial x_k}\right)\left(\sum_n \frac{\partial D_{jn}(\boldsymbol{x})}{\partial x_n}\right) \\ -\left(\sum_k D_{jk}(\boldsymbol{x})\frac{\partial p(\boldsymbol{x})}{\partial x_k}\right)\left(\sum_m \frac{\partial D_{im}(\boldsymbol{x})}{\partial x_m}\right) - p(\boldsymbol{x})\left(\sum_k \frac{\partial D_{jk}(\boldsymbol{x})}{\partial x_k}\right)\left(\sum_m \frac{\partial D_{im}(\boldsymbol{x})}{\partial x_m}\right)\Bigg\}dt. \quad (5.34)\end{aligned}$$

By similar means the system entropy can be written, after substitution of the original Fokker-Planck equation,

$$\begin{aligned}d\Delta\mathcal{S}_{\mathrm{sys}} &= -\frac{1}{p(\boldsymbol{x})}\frac{\partial p(\boldsymbol{x})}{\partial t}dt - \sum_i \frac{1}{p(\boldsymbol{x})}\frac{\partial p(\boldsymbol{x})}{\partial x_i}\circ dx_i \\ &= -\frac{1}{p(\boldsymbol{x})}\left(-\sum_i \frac{\partial(A_i(\boldsymbol{x})p(\boldsymbol{x}))}{\partial x_i} + \sum_{i,j}\frac{\partial^2(D_{ij}(\boldsymbol{x})p(\boldsymbol{x}))}{\partial x_i\partial x_j}\right)dt - \sum_i \frac{1}{p(\boldsymbol{x})}\frac{\partial p(\boldsymbol{x})}{\partial x_i}dx_i \\ &\quad - \sum_{i,j}\frac{D_{ij}(\boldsymbol{x})}{p(\boldsymbol{x})}\left(\frac{\partial^2 p(\boldsymbol{x})}{\partial x_i\partial x_j} - \frac{1}{p(\boldsymbol{x})}\frac{\partial p(\boldsymbol{x})}{\partial x_i}\frac{\partial p(\boldsymbol{x})}{\partial x_j}\right)dt \quad (5.35)\end{aligned}$$

so by averaging and integration by parts and dropping or cancelling boundary terms we have

$$\begin{aligned}\langle d\Delta\mathcal{S}_{\mathrm{sys}}\rangle_{p_0^{\mathrm{F}}}^{\mathrm{F}} &= \int d\boldsymbol{x}\ -\left(-\sum_i \frac{\partial((A_i^{\mathrm{ir}}(\boldsymbol{x}) + A_i^{\mathrm{rev}}(\boldsymbol{x}))p(\boldsymbol{x}))}{\partial x_i} + \sum_{i,j}\frac{\partial^2(D_{ij}(\boldsymbol{x})p(\boldsymbol{x}))}{\partial x_i\partial x_j}\right)dt \\ &\quad - \sum_i \frac{\partial p(\boldsymbol{x})}{\partial x_i}(A_i^{\mathrm{ir}}(\boldsymbol{x}) + A_i^{\mathrm{rev}}(\boldsymbol{x}))dt - \sum_{i,j} D_{ij}(\boldsymbol{x})\left(\frac{\partial^2 p(\boldsymbol{x})}{\partial x_i\partial x_j} - \frac{1}{p(\boldsymbol{x})}\frac{\partial p(\boldsymbol{x})}{\partial x_i}\frac{\partial p(\boldsymbol{x})}{\partial x_j}\right)dt \\ &= \int d\boldsymbol{x}\ -\sum_i A_i^{\mathrm{ir}}(\boldsymbol{x})\frac{\partial p(\boldsymbol{x})}{\partial x_i}dt + \sum_i p(\boldsymbol{x})\frac{\partial A_i^{\mathrm{rev}}(\boldsymbol{x})}{\partial x_i}dt + \sum_{i,j}\frac{\partial D_{ij}(\boldsymbol{x})}{\partial x_i}\frac{\partial p(\boldsymbol{x})}{\partial x_j}dt \\ &\quad + \sum_{i,j}\frac{D_{ij}(\boldsymbol{x})}{p(\boldsymbol{x})}\frac{\partial p(\boldsymbol{x})}{\partial x_i}\frac{\partial p(\boldsymbol{x})}{\partial x_j}dt. \quad (5.36)\end{aligned}$$

To proceed we use the identity $\sum_k D_{ik}^{-1}D_{kj} = \delta_{ij}$ noting both are symmetric matrices in order to re-write

$$\begin{aligned}&\sum_{i,j} D_{ij}\frac{\partial p}{\partial x_i}\frac{\partial p}{\partial x_j} = \sum_{m,n} D_{mn}\frac{\partial p}{\partial x_m}\frac{\partial p}{\partial x_n} = \sum_{m,n}\frac{\partial p}{\partial x_m}\frac{\partial p}{\partial x_n}\left(\sum_i D_{im}\delta_{in}\right) \\ &= \sum_{m,n}\frac{\partial p}{\partial x_m}\frac{\partial p}{\partial x_n}\left(\sum_i D_{im}\left(\sum_j D_{ij}^{-1}D_{jn}\right)\right) = \sum_{i,j} D_{ij}^{-1}\left(\sum_m D_{im}\frac{\partial p}{\partial x_m}\right)\left(\sum_n D_{jn}\frac{\partial p}{\partial x_n}\right), \quad (5.37)\end{aligned}$$

$$\sum_i A_i^{\rm ir}\frac{\partial p}{\partial x_i}=\sum_{i,m}A_i^{\rm ir}\frac{\partial p}{\partial x_m}\delta_{im}=\sum_{i,m}A_i^{\rm ir}\frac{\partial p}{\partial x_m}\left(\sum_j D^{-1}_{ij}D_{jm}\right)$$
$$=\sum_{i,j}A_i^{\rm ir}D^{-1}_{ij}\left(\sum_m D_{jm}\frac{\partial p}{\partial x_m}\right)\equiv\sum_{i,j}A_j^{\rm ir}D^{-1}_{ij}\left(\sum_n D_{in}\frac{\partial p}{\partial x_n}\right) \tag{5.38}$$

and

$$\sum_{i,j}\frac{\partial D_{ij}}{\partial x_i}\frac{\partial p}{\partial x_j}=\sum_{m,n}\frac{\partial D_{mn}}{\partial x_m}\frac{\partial p}{\partial x_n}=\sum_{m,n}\frac{\partial p}{\partial x_n}\left(\sum_i\frac{\partial D_{mi}}{\partial x_m}\delta_{in}\right)$$
$$=\sum_{m,n}\frac{\partial p}{\partial x_n}\left(\sum_i\frac{\partial D_{mi}}{\partial x_m}\left(\sum_j D^{-1}_{i,j}D_{jn}\right)\right)=\sum_{i,j}D^{-1}_{ij}\left(\sum_n D_{jn}\frac{\partial p}{\partial x_n}\right)\left(\sum_m\frac{\partial D_{mi}}{\partial x_m}\right). \tag{5.39}$$

Combining the two contributions, recognising the equivalence of the dummy summation indices we find

$$\frac{d\langle\Delta\mathcal{S}_{\rm tot}\rangle^{\rm F}_{p_0^{\rm F}}}{dt}=\int d\boldsymbol{x}\sum_{i,j}\frac{D^{-1}_{ij}(\boldsymbol{x})}{p(\boldsymbol{x})}\left(A_i^{\rm ir}(\boldsymbol{x})p(\boldsymbol{x})-\sum_m D_{mi}(\boldsymbol{x})\frac{\partial p(\boldsymbol{x})}{\partial x_m}-\sum_m p(\boldsymbol{x})\frac{\partial D_{mi}(\boldsymbol{x})}{\partial x_m}\right)$$
$$\times\left(A_j^{\rm ir}(\boldsymbol{x})p(\boldsymbol{x})-\sum_n D_{nj}(\boldsymbol{x})\frac{\partial p(\boldsymbol{x})}{\partial x_n}-\sum_n p(\boldsymbol{x})\frac{\partial D_{nj}(\boldsymbol{x})}{\partial x_n}\right)$$
$$=\sum_{i,j}\int d\boldsymbol{x}\,\frac{D^{-1}_{ij}(\boldsymbol{x})}{p(\boldsymbol{x})}J_i^{\rm ir}(\boldsymbol{x})J_j^{\rm ir}(\boldsymbol{x})$$
$$=\int d\boldsymbol{x}\,p(\boldsymbol{x},t)\left[\frac{\boldsymbol{J}^{\rm ir}(\boldsymbol{x},t)}{p(\boldsymbol{x},t)}\right]^{\rm T}\boldsymbol{D}^{-1}(\boldsymbol{x},\boldsymbol{\lambda}(t))\left[\frac{\boldsymbol{J}^{\rm ir}(\boldsymbol{x},t)}{p(\boldsymbol{x},t)}\right] \tag{5.40}$$

which is rigorously non-negative because $\boldsymbol{D}$ is semi-positive definite and, unlike in previous definitions [98], contributes when there is a non-zero *irreversible* flux. We also briefly mention that this result is compatible with the potential conditions from Sect. 3.7 and reveals that the condition for detailed balance, in general, is the requirement of zero irreversible flux.

5.3 Expressions for $\Delta\mathcal{S}_1$, $\Delta\mathcal{S}_2$ and $\Delta\mathcal{S}_3$

In order to consider a division of the entropy production as considered previously we are once again required to construct path probabilities using the so-called adjoint dynamics. We stress these dynamics may not be physically realisable: for example they may require negative positional steps to result from positive velocities (as indicated by the paths $\vec{\boldsymbol{x}}^{\rm R}$ and $\vec{\boldsymbol{x}}^{\rm T}$), but this is of no concern since they are only introduced for the mathematical construction of the entropy contributions. We consider an arbitrary stationary distribution of a given system which may be written in terms of a non-equilibrium potential, $\phi(\boldsymbol{x},\boldsymbol{\lambda}(t))$, such that

$$p^{\rm st}(\boldsymbol{x},\boldsymbol{\lambda}(t))=\exp[-\phi(\boldsymbol{x},\boldsymbol{\lambda}(t))], \tag{5.41}$$

where again $p^{\rm st}(\boldsymbol{x},\boldsymbol{\lambda}(t))\equiv p^{\rm F,st}(\boldsymbol{x},\boldsymbol{\lambda}^{\rm F}(t))$, and assert that the adjoint dynamics are those that result in the same stationary distribution, but have a stationary flux of opposite sign. As such we require

$$\frac{\partial p^{\rm st}(\boldsymbol{x},\boldsymbol{\lambda}(t))}{\partial t}=-\nabla\cdot\boldsymbol{J}^{\rm st}(\boldsymbol{x},\boldsymbol{\lambda}(t))=\nabla\cdot\boldsymbol{J}^{\rm st,ad}(\boldsymbol{x},\boldsymbol{\lambda}(t))=0 \tag{5.42}$$

with

$$J_i^{\mathrm{st,ad}}(\boldsymbol{x},\boldsymbol{\lambda}(t)) = -J_i^{\mathrm{st}}(\boldsymbol{x},\boldsymbol{\lambda}(t)). \tag{5.43}$$

In order to characterise the adjoint dynamics we construct the adjoint flux according to

$$\begin{aligned}
J_i^{\mathrm{st,ad}}(\boldsymbol{x},\boldsymbol{\lambda}(t)) &= A_i^{\mathrm{ad}}(\boldsymbol{x},\boldsymbol{\lambda}(t))p^{\mathrm{st}}(\boldsymbol{x},\boldsymbol{\lambda}(t)) - \sum_m \frac{\partial}{\partial x_m}\left(D_{im}(\boldsymbol{x},\boldsymbol{\lambda}(t))p^{\mathrm{st}}(\boldsymbol{x},\boldsymbol{\lambda}(t))\right) \\
&= A_i^{\mathrm{ad}}(\boldsymbol{x},\boldsymbol{\lambda}(t))e^{-\phi(\boldsymbol{x},\boldsymbol{\lambda}(t))} - \sum_m \frac{\partial}{\partial x_m}\left(D_{im}(\boldsymbol{x},\boldsymbol{\lambda}(t))e^{-\phi(\boldsymbol{x},\boldsymbol{\lambda}(t))}\right) \\
&= \left(A_i^{\mathrm{ad}}(\boldsymbol{x},\boldsymbol{\lambda}(t)) - \sum_m \frac{\partial D_{im}(\boldsymbol{x},\boldsymbol{\lambda}(t))}{\partial x_m} + \sum_m D_{im}(\boldsymbol{x},\boldsymbol{\lambda}(t))\frac{\partial \phi(\boldsymbol{x},\boldsymbol{\lambda}(t))}{\partial x_m}\right) e^{-\phi(\boldsymbol{x},\boldsymbol{\lambda}(t))} \\
&= -\left(A_i(\boldsymbol{x},\boldsymbol{\lambda}(t)) - \sum_m \frac{\partial D_{im}(\boldsymbol{x},\boldsymbol{\lambda}(t))}{\partial x_m} + \sum_m D_{im}(\boldsymbol{x},\boldsymbol{\lambda}(t))\frac{\partial \phi(\boldsymbol{x},\boldsymbol{\lambda}(t))}{\partial x_m}\right) e^{-\phi(\boldsymbol{x},\boldsymbol{\lambda}(t))}.
\end{aligned} \tag{5.44}$$

Consequently we have the requirement

$$A_i^{\mathrm{ad}}(\boldsymbol{x},\boldsymbol{\lambda}(t)) = -A_i(\boldsymbol{x},\boldsymbol{\lambda}(t)) + \sum_m 2\frac{\partial D_{im}(\boldsymbol{x},\boldsymbol{\lambda}(t))}{\partial x_m} - \sum_m 2D_{im}(\boldsymbol{x},\boldsymbol{\lambda}(t))\frac{\partial \phi(\boldsymbol{x},\boldsymbol{\lambda}(t))}{\partial x_m}. \tag{5.45}$$

We also note that this construction ensures the validity of the fluctuation theorems since the noise strength is left unchanged. Let us now consider the quantity

$$d\Delta\mathcal{S}_{\mathrm{ex}} = \ln \frac{p(\boldsymbol{x}', t+dt|\boldsymbol{x},t;\boldsymbol{\lambda}(t))}{p^{\mathrm{ad}}(\boldsymbol{x},t+dt|\boldsymbol{x}',t;\boldsymbol{\lambda}(t))}, \tag{5.46}$$

where $\Delta\mathcal{S}_{\mathrm{ex}} = \Delta Q_{\mathrm{ex}}/k_BT$, comprising the excess heat transfer from Eq. (3.113), which we have previously asserted constitutes part of the incremental contribution to the quantity $\Delta\mathcal{S}_1$ based on relations in Eqs. (5.10) and its short time representation. We evaluate Eq. (5.46), taking the transition probability density in the numerator from Eq. (5.18) and, for convenience, choosing $a = 1/2$. We can represent the transition probability density appearing in the denominator through a similar construction, but using a substitution for the adjoint drift term from Eq. (5.45), together with the complementary evaluation point choice $b = 1 - a = 1/2$ such that

$$\begin{aligned}
&p^{\mathrm{ad}}(\boldsymbol{x},t+dt|\boldsymbol{x}',t;\boldsymbol{\lambda}(t)) = \sqrt{\frac{1}{(4\pi dt)^N\left|\boldsymbol{D}(\boldsymbol{r},\boldsymbol{\lambda}(t))\right|}} \\
&\exp\left[\sum_{i,j} -\frac{1}{4dt}\left(-dx_i + \left(A_i(\boldsymbol{r},\boldsymbol{\lambda}(t)) - \sum_m \frac{\partial D_{im}(\boldsymbol{r},\boldsymbol{\lambda}(t))}{\partial r_m} + \sum_m 2D_{im}(\boldsymbol{r},\boldsymbol{\lambda}(t))\frac{\partial\phi(\boldsymbol{r},\boldsymbol{\lambda}(t))}{\partial r_m}\right)dt\right)\right. \\
&\quad \times D_{ij}^{-1}(\boldsymbol{r},\boldsymbol{\lambda}(t))\left(-dx_j + \left(A_j(\boldsymbol{r},\boldsymbol{\lambda}(t)) - \sum_n \frac{\partial D_{jn}(\boldsymbol{r},\boldsymbol{\lambda}(t))}{\partial r_n} + \sum_n 2D_{jn}(\boldsymbol{r},\boldsymbol{\lambda}(t))\frac{\partial\phi(\boldsymbol{r},\boldsymbol{\lambda}(t))}{\partial r_n}\right)dt\right) \\
&\left.+\sum_i \frac{dt}{2}\frac{\partial}{\partial r_i}\left(A_i(\boldsymbol{r},\boldsymbol{\lambda}(t)) - \sum_m 2\frac{\partial D_{im}(\boldsymbol{r},\boldsymbol{\lambda}(t))}{\partial r_m} + \sum_m 2D_{im}(\boldsymbol{r},\boldsymbol{\lambda}(t))\frac{\partial\phi(\boldsymbol{r})}{\partial r_m}\right) + \sum_{i,j}\frac{dt}{4}\frac{\partial^2 D_{ij}(\boldsymbol{r},\boldsymbol{\lambda}(t))}{\partial r_i\partial r_j}\right].
\end{aligned} \tag{5.47}$$

Since in both cases we have chosen evaluation at $a = b = 1/2$ we note that multiplication follows Stratonovich rules so that we have $f(\boldsymbol{r})*dx_i = f(\boldsymbol{x})\circ dx_i$. Considering the ratio of these two propagators,

again expressing $f(\boldsymbol{x}, \boldsymbol{\lambda}(t)) \equiv f(\boldsymbol{x})$, we find

$$
\begin{aligned}
&d\Delta\mathcal{S}_{\rm ex} = \ln \frac{p(\boldsymbol{x}', t+dt|\boldsymbol{x}, t)}{p^{\rm ad}(\boldsymbol{x}, t+dt|\boldsymbol{x}', t)} = \\
&\sum_i -\frac{\partial A_i(\boldsymbol{x})}{\partial x_i} dt + \sum_{i,j} \frac{\partial^2 D_{ij}(\boldsymbol{x})}{\partial x_i \partial x_j} dt - \sum_i \frac{\partial}{\partial x_i} \left(D_{im}(\boldsymbol{x}) \frac{\partial \phi(\boldsymbol{x})}{\partial x_m} \right) dt \\
&- \sum_{i,j} \frac{D^{-1}_{ij}(\boldsymbol{x})}{2} \left[\left(\sum_m D_{im}(\boldsymbol{x}) \frac{\partial \phi(\boldsymbol{x})}{\partial x_m} \right) \circ dx_j + \left(\sum_n D_{jn}(\boldsymbol{x}) \frac{\partial \phi(\boldsymbol{x})}{\partial x_n} \right) \circ dx_i \right. \\
&+ \left(\sum_m D_{im}(\boldsymbol{x}) \frac{\partial \phi(\boldsymbol{x})}{\partial x_m} \right) \left(\sum_n \frac{\partial D_{jn}(\boldsymbol{x})}{\partial x_n} \right) dt + \left(\sum_n D_{jn}(\boldsymbol{x}) \frac{\partial \phi(\boldsymbol{x})}{\partial x_n} \right) \left(\sum_m \frac{\partial D_{im}(\boldsymbol{x})}{\partial x_m} \right) dt \\
&\left. - A_j(\boldsymbol{x}) \left(\sum_m D_{im}(\boldsymbol{x}) \frac{\partial \phi(\boldsymbol{x})}{\partial x_m} \right) dt - A_i(\boldsymbol{x}) \left(\sum_n D_{jn}(\boldsymbol{x}) \frac{\partial \phi(\boldsymbol{x})}{\partial x_n} \right) dt \right] \\
&+ \sum_{i,j} D^{-1}_{ij}(\boldsymbol{x}) \left(\sum_m D_{im}(\boldsymbol{x}) \frac{\partial \phi(\boldsymbol{x})}{\partial x_m} \right) \left(\sum_n D_{jn}(\boldsymbol{x}) \frac{\partial \phi(\boldsymbol{x})}{\partial x_n} \right) dt
\end{aligned}
\tag{5.48}
$$

and in Itō form

$$
\begin{aligned}
d\Delta\mathcal{S}_{\rm ex} = &\ln \frac{p(\boldsymbol{x}', t+dt|\boldsymbol{x}, t; \boldsymbol{\lambda}(t))}{p^{\rm ad}(\boldsymbol{x}, t+dt|\boldsymbol{x}', t; \boldsymbol{\lambda}(t))} = \sum_i -\frac{\partial A_i(\boldsymbol{x})}{\partial x_i} dt + \sum_{i,j} \frac{\partial^2 D_{ij}(\boldsymbol{x})}{\partial x_i \partial x_j} dt - \sum_i \frac{\partial}{\partial x_i} \left(D_{im}(\boldsymbol{x}) \frac{\partial \phi}{\partial x_m} \right) dt \\
&- \sum_{i,j} \frac{D^{-1}_{ij}(\boldsymbol{x})}{2} \left[\left(\sum_m D_{im}(\boldsymbol{x}) \frac{\partial \phi}{\partial x_m} \right) dx_j + \left(\sum_n D_{jn}(\boldsymbol{x}) \frac{\partial \phi(\boldsymbol{x})}{\partial x_n} \right) dx_i \right. \\
&+ \left(\sum_m D_{im}(\boldsymbol{x}) \frac{\partial \phi(\boldsymbol{x})}{\partial x_m} \right) \left(\sum_n \frac{\partial D_{jn}(\boldsymbol{x})}{\partial x_n} \right) dt + \left(\sum_n D_{jn}(\boldsymbol{x}) \frac{\partial \phi(\boldsymbol{x})}{\partial x_n} \right) \left(\sum_m \frac{\partial D_{im}(\boldsymbol{x})}{\partial x_m} \right) dt \\
&\left. - A_j(\boldsymbol{x}) \left(\sum_m D_{im}(\boldsymbol{x}) \frac{\partial \phi(\boldsymbol{x})}{\partial x_m} \right) dt - A_i(\boldsymbol{x}) \left(\sum_n D_{jn}(\boldsymbol{x}) \frac{\partial \phi}{\partial x_n} \right) dt \right] \\
&+ \sum_{i,j} D^{-1}_{ij}(\boldsymbol{x}) \left(\sum_m D_{im}(\boldsymbol{x}) \frac{\partial \phi(\boldsymbol{x})}{\partial x_m} \right) \left(\sum_n D_{jn}(\boldsymbol{x}) \frac{\partial \phi(\boldsymbol{x})}{\partial x_n} \right) dt \\
&- \sum_{i,j,k} \frac{D_{jk}(\boldsymbol{x})}{2} \frac{\partial}{\partial x_k} \left(\left(\sum_m D_{im}(\boldsymbol{x}) \frac{\partial \phi(\boldsymbol{x})}{\partial x_m} \right) D^{-1}_{ij}(\boldsymbol{x}) \right) dt \\
&- \sum_{i,j,k} \frac{D_{ik}(\boldsymbol{x})}{2} \frac{\partial}{\partial x_k} \left(\left(\sum_n D_{jn}(\boldsymbol{x}) \frac{\partial \phi(\boldsymbol{x})}{\partial x_n} \right) D^{-1}_{ij}(\boldsymbol{x}) \right) dt.
\end{aligned}
\tag{5.49}
$$

However, we also have the stationarity condition

$$
\begin{aligned}
\nabla \cdot \boldsymbol{J}^{\rm st}(\boldsymbol{x}) &= \sum_i \frac{\partial}{\partial x_i} \left(e^{-\phi(\boldsymbol{x})} \left(A_i(\boldsymbol{x}) - \sum_m \frac{\partial D_{im}(\boldsymbol{x})}{\partial x_m} + \sum_m D_{im}(\boldsymbol{x}) \frac{\partial \phi(\boldsymbol{x})}{\partial x_m} \right) \right) \\
&= \sum_i \left(-A_i(\boldsymbol{x}) \frac{\partial \phi(\boldsymbol{x})}{\partial x_i} - \sum_m D_{im}(\boldsymbol{x}) \frac{\partial \phi(\boldsymbol{x})}{\partial x_i} \frac{\partial \phi(\boldsymbol{x})}{\partial x_m} + \frac{\partial A_i(\boldsymbol{x})}{\partial x_i} \right. \\
&\qquad \left. - \sum_m \frac{\partial^2 D_{im}(\boldsymbol{x})}{\partial x_i \partial x_m} + \sum_m D_{im}(\boldsymbol{x}) \frac{\partial^2 \phi(\boldsymbol{x})}{\partial x_i \partial x_m} + \sum_m 2 \frac{\partial D_{im}(\boldsymbol{x})}{\partial x_i} \frac{\partial \phi(\boldsymbol{x})}{\partial x_m} \right) e^{-\phi(\boldsymbol{x})} \\
&= 0
\end{aligned}
\tag{5.50}
$$

and so by utilising $\sum_k D^{-1}_{ik} D_{jk} = \delta_{ij}$ repeatedly we arrive at

$$\ln \frac{p(\boldsymbol{x}', t+dt|\boldsymbol{x}, t; \boldsymbol{\lambda}(t))}{p^{\mathrm{ad}}(\boldsymbol{x}, t+dt|\boldsymbol{x}', t; \boldsymbol{\lambda}(t))} = \sum_i -\frac{\partial \phi(\boldsymbol{x})}{\partial x_i} \circ dx_i \tag{5.51}$$

which justifies the usual characterisation of the adjoint dynamics [25, 36, 111] for use in continuous dynamics when written

$$\frac{p(\boldsymbol{x}', t+dt|\boldsymbol{x}, t; \boldsymbol{\lambda}(t))}{p^{\mathrm{ad}}(\boldsymbol{x}, t+dt|\boldsymbol{x}', t; \boldsymbol{\lambda}(t))} = \frac{p^{\mathrm{st}}(\boldsymbol{x}', \boldsymbol{\lambda}(t))}{p^{\mathrm{st}}(\boldsymbol{x}, \boldsymbol{\lambda}(t))} \tag{5.52}$$

and produces the SDE for the entropy production contribution $\Delta \mathcal{S}_1$

$$d\Delta \mathcal{S}_1 = -\frac{1}{p(\boldsymbol{x})} \frac{\partial p(\boldsymbol{x})}{\partial t} dt - \sum_i \frac{1}{p(\boldsymbol{x})} \frac{\partial p(\boldsymbol{x})}{\partial x_i} \circ dx_i - \sum_i \frac{\partial \phi(\boldsymbol{x})}{\partial x_i} \circ dx_i. \tag{5.53}$$

To find the mean production rate we must utilise the unsimplified Itō form along with the expansion of terms

$$\sum_{i,m} D_{im}(\boldsymbol{x}) \frac{\partial p(\boldsymbol{x})}{\partial x_i} \frac{\partial \phi(\boldsymbol{x})}{\partial x_m} = \sum_{i,j} D^{-1}_{ij}(\boldsymbol{x}) \left(\sum_m D_{im}(\boldsymbol{x}) \frac{\partial \phi(\boldsymbol{x})}{\partial x_m} \right) \left(\sum_n D_{jn}(\boldsymbol{x}) \frac{\partial p(\boldsymbol{x})}{\partial x_n} \right). \tag{5.54}$$

Such an application, along with the inclusion of the system entropy in appropriate form, integrating by parts and dropping boundary terms where appropriate and equating terms with dummy indices gives

$$\begin{aligned}
\frac{d\langle \Delta \mathcal{S}_1 \rangle^{\mathrm{F}}_{p^{\mathrm{F}}_0}}{dt} &= \sum_{i,j} \int d\boldsymbol{x} \left[\frac{D^{-1}_{ij}(\boldsymbol{x})}{p(\boldsymbol{x})} \left(\sum_m D_{im}(\boldsymbol{x}) \frac{\partial p(\boldsymbol{x})}{\partial x_m} \right) \left(\sum_n D_{jn}(\boldsymbol{x}) \frac{\partial p(\boldsymbol{x})}{\partial x_n} \right) \right. \\
&\quad + p(\boldsymbol{x}) D^{-1}_{ij}(\boldsymbol{x}) \left(\sum_m D_{im}(\boldsymbol{x}) \frac{\partial \phi(\boldsymbol{x})}{\partial x_m} \right) \left(\sum_n D_{jn}(\boldsymbol{x}) \frac{\partial \phi(\boldsymbol{x})}{\partial x_n} \right) \\
&\quad + D^{-1}_{ij}(\boldsymbol{x}) \left(\sum_m D_{im}(\boldsymbol{x}) \frac{\partial p(\boldsymbol{x})}{\partial x_m} \right) \left(\sum_n D_{jn}(\boldsymbol{x}) \frac{\partial \phi(\boldsymbol{x})}{\partial x_n} \right) \\
&\quad \left. + D^{-1}_{ij}(\boldsymbol{x}) \left(\sum_n D_{jn}(\boldsymbol{x}) \frac{\partial p(\boldsymbol{x})}{\partial x_n} \right) \left(\sum_m D_{im}(\boldsymbol{x}) \frac{\partial \phi(\boldsymbol{x})}{\partial x_m} \right) \right] \\
&= \sum_{i,j} \int d\boldsymbol{x}\, p(\boldsymbol{x}) D^{-1}_{ij}(\boldsymbol{x}) \left(-\sum_m \frac{D_{im}(\boldsymbol{x})}{p(\boldsymbol{x})} \frac{\partial p(\boldsymbol{x})}{\partial x_m} - \sum_m D_{im}(\boldsymbol{x}) \frac{\partial \phi(\boldsymbol{x})}{\partial x_m} \right) \\
&\qquad \times \left(-\sum_n \frac{D_{jn}(\boldsymbol{x})}{p(\boldsymbol{x})} \frac{\partial p(\boldsymbol{x})}{\partial x_n} - \sum_n D_{jn}(\boldsymbol{x}) \frac{\partial \phi(\boldsymbol{x})}{\partial x_n} \right) \\
&= \int d\boldsymbol{x}\, p(\boldsymbol{x}, t) \left[\frac{\boldsymbol{J}^{\mathrm{ir}}(\boldsymbol{x}, t)}{p(\boldsymbol{x}, t)} - \frac{\boldsymbol{J}^{\mathrm{ir,st}}(\boldsymbol{x}, \boldsymbol{\lambda}(t))}{p^{\mathrm{st}}(\boldsymbol{x}, \boldsymbol{\lambda}(t))} \right]^{\mathrm{T}} \boldsymbol{D}^{-1}(\boldsymbol{x}, \boldsymbol{\lambda}(t)) \left[\frac{\boldsymbol{J}^{\mathrm{ir}}(\boldsymbol{x}, t)}{p(\boldsymbol{x}, t)} - \frac{\boldsymbol{J}^{\mathrm{ir,st}}(\boldsymbol{x}, \boldsymbol{\lambda}(t))}{p^{\mathrm{st}}(\boldsymbol{x}, \boldsymbol{\lambda}(t))} \right].
\end{aligned} \tag{5.55}$$

Since the above expression is in fact independent of $\boldsymbol{A}(\boldsymbol{x}, \boldsymbol{\lambda}(t))$, it can be expressed with the irreversible current $\boldsymbol{J}^{\mathrm{ir}}$ replaced by total current $\boldsymbol{J}$ and so maps precisely onto the non-adiabatic entropy production appearing in [98] and thus can be expressed as

$$\frac{d\langle \Delta \mathcal{S}_1 \rangle^{\mathrm{F}}_{p^{\mathrm{F}}_0}}{dt} = -\int d\boldsymbol{x}\, \frac{\partial p(\boldsymbol{x}, t)}{\partial t} \ln \frac{p(\boldsymbol{x}, t)}{p^{\mathrm{st}}(\boldsymbol{x}, \boldsymbol{\lambda}(t))} \tag{5.56}$$

as highlighted by the authors of [98]. We emphasise however, that Eq. (5.56) is to be considered alongside the accompanying SDE in Eq. (5.53), from which it has been derived directly, rather than by a division of an observed positive contribution to the mean rate of change of Gibbs entropy into presumed unique transient and stationary terms; something of particular importance given that, in the mean, $\Delta\mathcal{S}_3$ is expected to behave transiently yet is separate to the positive transient contribution $\Delta\mathcal{S}_1$.

We may now by similar means consider an increment in $\Delta\mathcal{S}_2$ as follows:

$$d\Delta\mathcal{S}_2 = \ln\frac{p(\boldsymbol{x}', t+dt|\boldsymbol{x}, t; \boldsymbol{\lambda}(t))}{p^{\mathrm{ad}}(\boldsymbol{\varepsilon x}', t+dt|\boldsymbol{\varepsilon x}, t; \boldsymbol{\varepsilon\lambda}(t))}. \tag{5.57}$$

In this case the construction of the denominator follows slightly different rules since, unlike $\Delta\mathcal{S}_{\mathrm{tot}}$ and $\Delta\mathcal{S}_1$, the alternative path, $\vec{\boldsymbol{x}}^{\mathrm{T}}$, is based on a time reversal of the instantaneous coordinates, but otherwise follows the sequence of the forward path. As such instead of utilising Eq. (5.25) one should use

$$f(\boldsymbol{r}')dx_i = f(\boldsymbol{x})dx_i + \sum_m 2bD_{im}(\boldsymbol{x})\frac{\partial f(\boldsymbol{x})}{\partial x_m}dt + O(dt^{3/2}). \tag{5.58}$$

This in turn leads to the appropriate choice $b = a$ once again reflecting the correct application of the involution on a sub-infinitesimal scale so that $\boldsymbol{r}' = \boldsymbol{r}$. For continuity, we may once again choose $a = b = 1/2$ with Stratonovich multiplication rules: we represent the transition probability appearing in the numerator through Eq. (5.18), and the denominator by similar means using the drift term given in Eq. (5.45) and the path choice $\boldsymbol{x}^{\mathrm{T}}(t) = \boldsymbol{\varepsilon x}(t)$, such that

$$\begin{aligned}
p^{\mathrm{ad}}(\boldsymbol{\varepsilon x}', t+dt|\boldsymbol{\varepsilon x}, t; \boldsymbol{\varepsilon\lambda}(t)) = \prod_i \sqrt{\frac{1}{(4\pi dt)^N\,|\boldsymbol{D}(\boldsymbol{\varepsilon r}, \boldsymbol{\varepsilon\lambda}(t))|}}\exp\left[\sum_{i,j} -\frac{D^{-1}_{ij}(\boldsymbol{\varepsilon r}, \boldsymbol{\varepsilon\lambda}(t))}{4dt}\right.\\
\times\left(\varepsilon_i dx_i + \left(A_i(\boldsymbol{\varepsilon r}, \boldsymbol{\varepsilon\lambda}(t)) - \sum_m \frac{\partial D_{im}(\boldsymbol{\varepsilon r}, \boldsymbol{\varepsilon\lambda}(t))}{\partial(\varepsilon_m r_m)} + \sum_m 2D_{im}(\boldsymbol{\varepsilon r}, \boldsymbol{\varepsilon\lambda}(t))\frac{\partial\phi(\boldsymbol{\varepsilon r})}{\partial(\varepsilon_m r_m)}\right)dt\right)\\
\times\left(\varepsilon_j dx_j + \left(A_j(\boldsymbol{\varepsilon r}, \boldsymbol{\varepsilon\lambda}(t)) - \sum_n \frac{\partial D_{jn}(\boldsymbol{\varepsilon r}, \boldsymbol{\varepsilon\lambda}(t))}{\partial(\varepsilon_n r_n)} + \sum_n 2D_{jn}(\boldsymbol{\varepsilon r}, \boldsymbol{\varepsilon\lambda}(t))\frac{\partial\phi(\boldsymbol{\varepsilon r}, \boldsymbol{\varepsilon\lambda}(t))}{\partial(\varepsilon_n r_n)}\right)dt\right)\\
+\sum_i \frac{dt}{2}\frac{\partial}{\partial\varepsilon_i r_i}\left(A_i(\boldsymbol{\varepsilon r}, \boldsymbol{\varepsilon\lambda}(t)) - \sum_m 2\frac{\partial D_{im}(\boldsymbol{\varepsilon r}, \boldsymbol{\varepsilon\lambda}(t))}{\partial(\varepsilon_m r_m)} + \sum_m 2D_{im}(\boldsymbol{\varepsilon r}, \boldsymbol{\varepsilon\lambda}(t))\frac{\partial\phi(\boldsymbol{\varepsilon r}, \boldsymbol{\varepsilon\lambda}(t))}{\partial(\varepsilon_m r_m)}\right)\\
\left.+\sum_{i,j}\frac{dt}{4}\frac{\partial^2 D_{ij}(\boldsymbol{\varepsilon r}, \boldsymbol{\varepsilon\lambda}(t))}{\partial(\varepsilon_i r_i)\partial(\varepsilon_j r_j)}\right].
\end{aligned} \tag{5.59}$$

We can utilise the usual transformation rules and assumptions for A^{ir}, A^{rev} and D_{ij} and express $\partial\phi(\boldsymbol{\varepsilon r})/\partial(\varepsilon_i r_i) \equiv \phi'_i(\boldsymbol{\varepsilon r})$ (along with $\partial^2\phi(\boldsymbol{\varepsilon r})/\partial(\varepsilon_i r_i)\partial(\varepsilon_j r_j) \equiv \phi''_{ij}(\boldsymbol{\varepsilon x})$) such that we can write the

propagator

$$p^{\mathrm{ad}}(\boldsymbol{\varepsilon x}', t+dt|\boldsymbol{\varepsilon x}, t; \boldsymbol{\varepsilon\lambda}(t)) = \prod_i \sqrt{\frac{1}{(4\pi dt)^N \left|\boldsymbol{D}(\boldsymbol{r}, \boldsymbol{\lambda}(t))\right|}} \exp\left[\sum_{i,j} -\frac{D_{ij}^{-1}(\boldsymbol{r}, \boldsymbol{\lambda}(t))}{4dt}\right.$$
$$\times \left(dx_i + \left(A_i^{\mathrm{ir}}(\boldsymbol{r}, \boldsymbol{\lambda}(t)) - A_i^{\mathrm{rev}}(\boldsymbol{r}, \boldsymbol{\lambda}(t)) - \sum_m \frac{\partial D_{im}(\boldsymbol{r}, \boldsymbol{\lambda}(t))}{\partial r_m} + \sum_m 2\varepsilon_m D_{im}(\boldsymbol{r}, \boldsymbol{\lambda}(t)) \frac{\partial \phi(\boldsymbol{\varepsilon r}, \boldsymbol{\varepsilon\lambda}(t))}{\partial(\varepsilon_m r_m)}\right) dt\right)$$
$$\times \left(dx_j + \left(A_j^{\mathrm{ir}}(\boldsymbol{r}, \boldsymbol{\lambda}(t)) - A_j^{\mathrm{rev}}(\boldsymbol{r}, \boldsymbol{\lambda}(t)) - \sum_n \frac{\partial D_{jn}(\boldsymbol{r}, \boldsymbol{\lambda}(t))}{\partial r_n} + \sum_n 2\varepsilon_n D_{jn}(\boldsymbol{r}, \boldsymbol{\lambda}(t)) \frac{\partial \phi(\boldsymbol{\varepsilon r}, \boldsymbol{\varepsilon\lambda}(t))}{\partial(\varepsilon_n r_n)}\right) dt\right)$$
$$+ \sum_i \frac{dt}{2} \frac{\partial}{\partial r_i} \left(A_i^{\mathrm{ir}}(\boldsymbol{r}, \boldsymbol{\lambda}(t)) - A_i^{\mathrm{rev}}(\boldsymbol{r}, \boldsymbol{\lambda}(t)) - \sum_m 2\frac{\partial D_{im}(\boldsymbol{r}, \boldsymbol{\lambda}(t))}{\partial r_m} + \sum_m 2\varepsilon_m D_{im}(\boldsymbol{r}, \boldsymbol{\lambda}(t)) \frac{\partial \phi(\boldsymbol{\varepsilon r}, \boldsymbol{\lambda}(t))}{\partial(\varepsilon_m r_m)}\right)$$
$$\left. + \sum_{i,j} \frac{dt}{4} \frac{\partial^2 D_{ij}(\boldsymbol{r}, \boldsymbol{\lambda}(t))}{\partial r_i \partial r_j}\right]. \tag{5.60}$$

Utilising this propagator and the stationarity condition evaluated with the time reversed variables

$$\begin{aligned}
\nabla \cdot \boldsymbol{J}^{\mathrm{st}}(\boldsymbol{\varepsilon x}) &= \sum_i \frac{\partial}{\partial(\varepsilon_i x_i)} \left(e^{-\phi(\boldsymbol{\varepsilon x})} \left(A_i(\boldsymbol{\varepsilon x}) - \sum_m \frac{\partial D_{im}(\boldsymbol{\varepsilon x})}{\partial(\varepsilon_m x_m)} + \sum_m D_{im}(\boldsymbol{\varepsilon x}) \frac{\partial \phi(\boldsymbol{\varepsilon x})}{\partial(\varepsilon_m x_m)}\right)\right) \\
&= \sum_i \left(-A_i(\boldsymbol{\varepsilon x}) \frac{\partial \phi(\boldsymbol{\varepsilon x})}{\partial(\varepsilon_i x_i)} - \sum_m D_{im}(\boldsymbol{\varepsilon x}) \frac{\partial \phi(\boldsymbol{\varepsilon x})}{\partial(\varepsilon_i x_i)} \frac{\partial \phi(\boldsymbol{\varepsilon x})}{\partial(\varepsilon_m x_m)} + \frac{\partial A_i(\boldsymbol{\varepsilon x})}{\partial(\varepsilon_i x_i)}\right. \\
&\quad - \sum_m \frac{\partial^2 D_{im}(\boldsymbol{\varepsilon x})}{\partial(\varepsilon_i x_i)\partial(\varepsilon_m x_m)} + \sum_m D_{im}(\boldsymbol{\varepsilon x}) \frac{\partial^2 \phi(\boldsymbol{\varepsilon x})}{\partial(\varepsilon_i x_i)\partial(\varepsilon_m x_m)} + \left.\sum_m 2\frac{\partial D_{im}(\boldsymbol{\varepsilon x})}{\partial(\varepsilon_i x_i)} \frac{\partial \phi(\boldsymbol{\varepsilon x})}{\partial(\varepsilon_m x_m)}\right) e^{-\phi(\boldsymbol{\varepsilon x})} \\
&= 0
\end{aligned} \tag{5.61}$$

and the usual transformation properties of $\boldsymbol{A}$ and $\boldsymbol{D}$ leads to the Stratonovich SDE

$$\begin{aligned}
d\Delta\mathcal{S}_2 &= \sum_i \varepsilon_i \phi_i'(\boldsymbol{\varepsilon x}) \circ dx_i - \sum_i \frac{\partial A^{\mathrm{rev}}(\boldsymbol{x})}{\partial x_i} dt + \sum_{i,j} \frac{D_{ij}^{-1}(\boldsymbol{x})}{2} \left(A_i^{\mathrm{ir}}(\boldsymbol{x}) \circ dx_j + A_j^{\mathrm{ir}}(\boldsymbol{x}) \circ dx_i\right) \\
&\quad - \frac{D_{ij}^{-1}(\boldsymbol{x})}{2} \left(A_i^{\mathrm{rev}}(\boldsymbol{x}) A_j^{\mathrm{ir}}(\boldsymbol{x}) + A_j^{\mathrm{rev}}(\boldsymbol{x}) A_i^{\mathrm{ir}}(\boldsymbol{x})\right) dt \\
&\quad - \frac{D_{ij}^{-1}(\boldsymbol{x})}{2} \left(\left(\sum_n \frac{\partial D_{jn}(\boldsymbol{x})}{\partial x_n}\right) \circ dx_j + \left(\sum_m \frac{\partial D_{im}(\boldsymbol{x})}{\partial x_m}\right) \circ dx_i\right) \\
&\quad + \frac{D_{ij}^{-1}(\boldsymbol{x})}{2} \left(A_j^{\mathrm{rev}}(\boldsymbol{x}) \left(\sum_m \frac{\partial D_{im}(\boldsymbol{x})}{\partial x_m}\right) + A_i^{\mathrm{rev}}(\boldsymbol{x}) \left(\sum_n \frac{\partial D_{jn}(\boldsymbol{x})}{\partial x_n}\right)\right) dt.
\end{aligned} \tag{5.62}$$

Full conversion to Itō form, averaging, performing integration by parts, applying the product rule, recognising the equivalence of dummy indices, identifying

$$\sum_i A_i^{\mathrm{ir}}(\boldsymbol{x}) \frac{\partial p(\boldsymbol{x})}{\partial x_i} = \sum_{i,j} D_{ij}^{-1}(\boldsymbol{x}) A_i^{\mathrm{ir}}(\boldsymbol{x}) \left(\sum_n D_{jn}(\boldsymbol{x}) \frac{\partial p(\boldsymbol{x})}{\partial x_n}\right), \tag{5.63}$$

$$\sum_{i,m} \frac{\partial p(\boldsymbol{x})}{\partial x_i} \frac{\partial D_{im}(\boldsymbol{x})}{\partial x_m} = \sum_{i,j} D_{ij}^{-1}(\boldsymbol{x}) \left(\sum_n D_{jn}(\boldsymbol{x}) \frac{\partial p(\boldsymbol{x})}{\partial x_n}\right) \left(\sum_m \frac{\partial D_{im}(\boldsymbol{x})}{\partial x_m}\right), \tag{5.64}$$

$$\sum_{i,m}\frac{\partial p(\boldsymbol{x})}{\partial x_i}\left(\sum_m \varepsilon_m D_{im}(\boldsymbol{x})\phi'_m(\boldsymbol{\varepsilon x})\right)=\sum_{i,j}D^{-1}_{ij}(\boldsymbol{x})\left(\sum_m D_{im}(\boldsymbol{x})\frac{\partial p(\boldsymbol{x})}{\partial x_m}\right)\left(\sum_n \varepsilon_n D_{jn}(\boldsymbol{x})\phi'_n(\boldsymbol{\varepsilon x})\right) \tag{5.65}$$

and the transformation property $D^{-1}_{ij}(\boldsymbol{\varepsilon x})=\varepsilon_i\varepsilon_j D^{-1}_{ij}(\boldsymbol{x})$ arising from the identity $\sum_k D_{ik}(\boldsymbol{\varepsilon x})D^{-1}_{jk}(\boldsymbol{\varepsilon x})=\sum_k D_{ik}(\boldsymbol{x})D^{-1}_{jk}(\boldsymbol{x})=\delta_{ij}$ then leads us to

$$\begin{aligned}
\frac{d\langle\Delta\mathcal{S}_2\rangle^{\mathrm{F}}_{p^{\mathrm{F}}_0}}{dt}&=\sum_{i,j}\int d\boldsymbol{x}\, p(\boldsymbol{x})D^{-1}_{ij}(\boldsymbol{x})\left(A^{\mathrm{ir}}_i(\boldsymbol{x})-\left(\sum_m\frac{\partial D_{im}(\boldsymbol{x})}{\partial x_m}\right)+\left(\sum_m\varepsilon_m D_{im}(\boldsymbol{x})\phi'_m(\boldsymbol{\varepsilon x})\right)\right)\\
&\times\left(A^{\mathrm{ir}}_j(\boldsymbol{x})-\left(\sum_n\frac{\partial D_{jn}(\boldsymbol{x})}{\partial x_n}\right)+\left(\sum_n\varepsilon_n D_{jn}(\boldsymbol{x})\phi'_n(\boldsymbol{\varepsilon x})\right)\right)\\
&=\sum_{i,j}\int d\boldsymbol{x}\, p(\boldsymbol{x})D^{-1}_{ij}(\boldsymbol{x})\left(\varepsilon_i A^{\mathrm{ir}}_i(\boldsymbol{\varepsilon x})-\varepsilon_i\left(\sum_m\frac{\partial D_{im}(\boldsymbol{\varepsilon x})}{\partial(\varepsilon_m x_m)}\right)+\varepsilon_i\left(\sum_m D_{im}(\boldsymbol{\varepsilon x})\phi'_m(\boldsymbol{\varepsilon x})\right)\right)\\
&\times\left(\varepsilon_j A^{\mathrm{ir}}_j(\boldsymbol{\varepsilon x})-\varepsilon_j\left(\sum_n\frac{\partial D_{jn}(\boldsymbol{\varepsilon x})}{\partial(\varepsilon_n x_n)}\right)+\varepsilon_j\left(\sum_n D_{jn}(\boldsymbol{\varepsilon x})\phi'_n(\boldsymbol{\varepsilon x})\right)\right)\\
&=\sum_{i,j}\int d\boldsymbol{x}\, p(\boldsymbol{x})\varepsilon_i\varepsilon_j D^{-1}_{ij}(\boldsymbol{x})\left(A^{\mathrm{ir}}_i(\boldsymbol{\varepsilon x})-\left(\sum_m\frac{\partial D_{im}(\boldsymbol{\varepsilon x})}{\partial(\varepsilon_m x_m)}\right)+\left(\sum_m D_{im}(\boldsymbol{\varepsilon x})\phi'_m(\boldsymbol{\varepsilon x})\right)\right)\\
&\times\left(A^{\mathrm{ir}}_j(\boldsymbol{\varepsilon x})-\left(\sum_n\frac{\partial D_{jn}(\boldsymbol{\varepsilon x})}{\partial(\varepsilon_n x_n)}\right)+\left(\sum_n D_{jn}(\boldsymbol{\varepsilon x})\phi'_n(\boldsymbol{\varepsilon x})\right)\right)\\
&=\sum_{i,j}\int d\boldsymbol{x}\, p(\boldsymbol{x})\left(\frac{J^{\mathrm{ir,st}}_i(\boldsymbol{\varepsilon x})}{p^{\mathrm{st}}(\boldsymbol{\varepsilon x})}\right)\left(\frac{J^{\mathrm{ir,st}}_j(\boldsymbol{\varepsilon x})}{p^{\mathrm{st}}(\boldsymbol{\varepsilon x})}\right)\varepsilon_i\varepsilon_j D^{-1}_{ij}(\boldsymbol{x})\\
&=\sum_{i,j}\int d\boldsymbol{x}\, p(\boldsymbol{x})\left(\frac{J^{\mathrm{ir,st}}_i(\boldsymbol{\varepsilon x})}{p^{\mathrm{st}}(\boldsymbol{\varepsilon x})}\right)\left(\frac{J^{\mathrm{ir,st}}_j(\boldsymbol{\varepsilon x})}{p^{\mathrm{st}}(\boldsymbol{\varepsilon x})}\right)D^{-1}_{ij}(\boldsymbol{\varepsilon x})\\
&=\int d\boldsymbol{x}\, p(\boldsymbol{x},t)\left[\frac{\boldsymbol{J}^{\mathrm{ir,st}}(\boldsymbol{\varepsilon x},\boldsymbol{\varepsilon\lambda}(t))}{p^{\mathrm{st}}(\boldsymbol{\varepsilon x},\boldsymbol{\varepsilon\lambda}(t))}\right]^{\mathrm{T}}\boldsymbol{D}^{-1}(\boldsymbol{\varepsilon x},\boldsymbol{\varepsilon\lambda}(t))\left[\frac{\boldsymbol{J}^{\mathrm{ir,st}}(\boldsymbol{\varepsilon x},\boldsymbol{\varepsilon\lambda}(t))}{p^{\mathrm{st}}(\boldsymbol{\varepsilon x},\boldsymbol{\varepsilon\lambda}(t))}\right].
\end{aligned} \tag{5.66}$$

Such a form illustrates the positivity requirement of $\Delta\mathcal{S}_2$ in the mean, resulting from its adherence to an IFT, and again Eq. (5.66) is to be considered alongside the complementary SDE in Eq. (5.62). Since it is based on an integral over the stationary irreversible flux, $d\langle\Delta\mathcal{S}_2\rangle^{\mathrm{F}}_{p^{\mathrm{F}}_0}/dt$ describes a contribution to entropy production which arises from an absence of detailed balance and is non-zero both in and out of stationarity given a non-zero stationary irreversible flux. This quantity is to be contrasted with the adiabatic entropy production in [98] which we may now consider to be a special case when there are only even variables in the dynamics. We point out again the importance of the direct derivation of this result from the SDE in this formalism since, unlike for $\Delta\mathcal{S}_{\mathrm{tot}}$ and $\Delta\mathcal{S}_1$, it is not just a generalisation of the expressions in [98], with the total current exchanged with the irreversible current, as it also includes a time reversal of the coordinates. That is to say, a division of the irreversible flux into terms with structure based solely on $p^{\mathrm{st}}(\boldsymbol{x})$ as in [98] would not have obviously led to the above expression.

To complete the description of all three contributions to entropy production we now consider an increment in $\Delta\mathcal{S}_3$. By using the definition in Eq. (5.12)

$$d\Delta\mathcal{S}_3=\ln\frac{p^{\mathrm{ad}}(\boldsymbol{\varepsilon x}',t+dt|\boldsymbol{\varepsilon x},t;\boldsymbol{\varepsilon\lambda}(t))p^{\mathrm{ad}}(\boldsymbol{x},t+dt|\boldsymbol{x}',t;\boldsymbol{\lambda}(t))}{p(\boldsymbol{x}',t+dt|\boldsymbol{x},t;\boldsymbol{\lambda}(t))p(\boldsymbol{\varepsilon x},t+dt|\boldsymbol{\varepsilon x}',t;\boldsymbol{\varepsilon\lambda}(t))}. \tag{5.67}$$

At this point one may find the SDE by comparison of those for $\Delta\mathcal{S}_{\mathrm{tot}}$, $\Delta\mathcal{S}_1$ and $\Delta\mathcal{S}_2$, but note the same

result emerges from a consideration of the propagators along with the two stationarity conditions. The result we find is

$$\begin{aligned} d\Delta\mathcal{S}_3 &= \sum_i \phi_i'(\boldsymbol{x}) \circ dx_i - \varepsilon_i \phi_i'(\boldsymbol{\varepsilon x}) \circ dx_i \\ &= \sum_i \ln \frac{\exp\left[-\phi(\boldsymbol{x})\right]}{\exp\left[-\phi(\boldsymbol{x'})\right]} \frac{\exp\left[-\phi(\boldsymbol{\varepsilon x'})\right]}{\exp\left[-\phi(\boldsymbol{\varepsilon x})\right]}, \end{aligned} \tag{5.68}$$

which maps onto the same quantity in the master equation approach. We can then construct the average contribution by converting to Itō form and performing the path integral such that

$$\begin{aligned} \langle d\Delta\mathcal{S}_3 \rangle_{p_0^{\mathrm{F}}}^{\mathrm{F}} = &\int d\boldsymbol{x} \sum_i p(\boldsymbol{x}) A_i(\boldsymbol{x}) (\phi_i'(\boldsymbol{x}) - \varepsilon_i \phi_i'(\boldsymbol{\varepsilon x})) dt \\ &+ \sum_{i,j} p(\boldsymbol{x}) D_{ij}(\boldsymbol{x}) (\phi_{ij}''(\boldsymbol{x}) - \varepsilon_i \varepsilon_j \phi_{ij}''(\boldsymbol{\varepsilon x})) dt \end{aligned} \tag{5.69}$$

and proceed to manipulate by integrating by parts, assuming the probability density and current vanish or cancel at boundaries, such that

$$\begin{aligned} \langle d\Delta\mathcal{S}_3 \rangle_{p_0^{\mathrm{F}}}^{\mathrm{F}} &= \int d\boldsymbol{x} \sum_i p(\boldsymbol{x}) A_i(\boldsymbol{x}) (\phi_i'(\boldsymbol{x}) - \varepsilon_i \phi_i'(\boldsymbol{\varepsilon x})) dt \\ &\qquad - \int d\boldsymbol{x} \sum_{i,j} \frac{\partial}{\partial x_j} \left(p(\boldsymbol{x}) D_{ij}(\boldsymbol{x})\right) (\phi_i'(\boldsymbol{x}) - \varepsilon_i \phi_i'(\boldsymbol{\varepsilon x})) dt \\ &= \sum_i \int d\boldsymbol{x} \ (\phi_i'(\boldsymbol{x}) - \varepsilon_i \phi_i'(\boldsymbol{\varepsilon x})) \left(A_i(\boldsymbol{x}) p(\boldsymbol{x}) - \sum_j \frac{\partial}{\partial x_j} \left(p(\boldsymbol{x}) D_{ij}(\boldsymbol{x})\right) \right) dt \\ &= \sum_i \int d\boldsymbol{x} \ (\phi_i'(\boldsymbol{x}) - \varepsilon_i \phi_i'(\boldsymbol{\varepsilon x})) \, J_i(\boldsymbol{x}) dt = - \sum_i \int d\boldsymbol{x} \ (\phi(\boldsymbol{x}) - \phi(\boldsymbol{\varepsilon x})) \frac{\partial J_i(\boldsymbol{x})}{\partial x_i} dt \\ &= - \int d\boldsymbol{x} \ (\phi(\boldsymbol{x}) - \phi(\boldsymbol{\varepsilon x})) \left(\sum_i \frac{\partial J_i(\boldsymbol{x})}{\partial x_i} \right) dt = - \int d\boldsymbol{x} \ (\phi(\boldsymbol{x}) - \phi(\boldsymbol{\varepsilon x})) \left(\nabla \cdot \boldsymbol{J}(\boldsymbol{x})\right) dt. \end{aligned} \tag{5.70}$$

By substituting the original Fokker-Planck equation we may also write this as

$$\begin{aligned} \frac{d\langle \Delta\mathcal{S}_3 \rangle_{p_0^{\mathrm{F}}}^{\mathrm{F}}}{dt} &= \frac{d\langle \Delta Q_{\mathrm{hk,T}}/k_B T \rangle_{p_0^{\mathrm{F}}}^{\mathrm{F}}}{dt} \\ &= \int d\boldsymbol{x} \, \frac{\partial p(\boldsymbol{x})}{\partial t} \left(\phi(\boldsymbol{x}) - \phi(\boldsymbol{\varepsilon x})\right) \\ &= - \int d\boldsymbol{x} \, \frac{\partial p(\boldsymbol{x}, t)}{\partial t} \ln \frac{p^{\mathrm{st}}(\boldsymbol{x}, \boldsymbol{\lambda}(t))}{p^{\mathrm{st}}(\boldsymbol{\varepsilon x}, \boldsymbol{\varepsilon \lambda}(t))}. \end{aligned} \tag{5.71}$$

This has a form similar to Eq. (5.56) and is clearly a contribution to the mean total entropy production rate that behaves transiently in a manner similar to $\Delta\mathcal{S}_1$. The quantity $\Delta\mathcal{S}_1$ appears in the Hatano-Sasa relation which describes the entropy production associated with a transition between different stationary states. However, in light of Eq. (5.71) we suggest that $\Delta\mathcal{S}_1$, and thus the Hatano-Sasa relation and non-adiabatic entropy production, when viewed as a contribution in the mean, do not represent the entire entropy production associated with transitions between stationary states (or more generally relaxation) since, in the mean, we can construct a new quantity which comprises all contributions which are non-zero

only during relaxation, by combining Eqs. (5.56) and (5.71) giving

$$\frac{d\langle \Delta\mathcal{S}_1 + \Delta\mathcal{S}_3\rangle^{\mathrm{F}}_{p_0^{\mathrm{F}}}}{dt} = \frac{d\langle k_B^{-1}\Delta\mathcal{S}_{\mathrm{sys}} + (\Delta Q_{\mathrm{ex}} + \Delta Q_{\mathrm{hk,T}})/k_BT\rangle^{\mathrm{F}}_{p_0^{\mathrm{F}}}}{dt} = -\int d\boldsymbol{x}\, \frac{\partial p(\boldsymbol{x},t)}{\partial t} \ln \frac{p(\boldsymbol{x},t)}{p^{\mathrm{st}}(\boldsymbol{\varepsilon x}, \boldsymbol{\varepsilon\lambda}(t))}. \tag{5.72}$$

This describes a contribution to the mean entropy production rate which occurs when the system is out of stationarity, but it does not obey an IFT and thus has no guarantee of positivity. The existence of such a quantity highlights our statement of the level on which the generalisation of the second law is to be observed. If one, perhaps erroneously, were to identify a constantly dissipated heat with a rigorously positive mean entropy contribution maintaining the non-equilibrium character due to a lack of detailed balance (as is expected with even variables) and then expected the remaining, transient, mean quantity to adhere to a generalised form of the second law for transitions between steady states, such an expectation would not, necessarily, be met.

Lastly we mention that given the forms of Eqs. (5.40), (5.56), (5.66) and (5.71) it appears there is an inconsistency. Since the contributions $d\langle\Delta\mathcal{S}_1\rangle^{\mathrm{F}}_{p_0^{\mathrm{F}}}/dt$ and $d\langle\Delta\mathcal{S}_3\rangle^{\mathrm{F}}_{p_0^{\mathrm{F}}}/dt$ vanish in the stationary state, $d\langle\Delta\mathcal{S}_{\mathrm{tot}}\rangle^{\mathrm{F}}_{p_0^{\mathrm{F}}}/dt$ and $d\langle\Delta\mathcal{S}_2\rangle^{\mathrm{F}}_{p_0^{\mathrm{F}}}/dt$ must be equal, but have manifestly different forms. However, we find, by comparing the appropriate averages of these two contributions in the stationary state (or all four in general) with application of the zero stationary current condition evaluated at $\boldsymbol{x}$ and $\boldsymbol{\varepsilon x}$, them to be equivalent. This allows us to note that the stationary distribution is the distribution, which when averaging the *odd part* of the conditional mean contribution to $d\langle\Delta\mathcal{S}_2\rangle^{\mathrm{F}}_{p_0^{\mathrm{F}}}/dt$, produces a vanishing result. Such a property is assured in general terms (though may be challenging to prove generally if the non-equilibrium potential is not known or is given in terms of an expansion), but is given context by the demonstration of results that point to such claims for an example system in a subsequent chapter.

5.4 Adjoint and Reversed Adjoint Dynamics

In an earlier chapter it was shown that the adjoint dynamics, which typically are unphysical (for example they may produce negative positional steps from positive velocities), may be written in terms of another set of dynamics, which are physical, that we have termed the reversed adjoint dynamics. These reversed adjoint dynamics were then described by transition rates

$$T^{\text{ad-rev}}(\boldsymbol{\varepsilon x}|\boldsymbol{\varepsilon x}', \boldsymbol{\varepsilon\lambda}(t)) = T^{\mathrm{ad}}(\boldsymbol{x}|\boldsymbol{x}', \boldsymbol{\lambda}(t)). \tag{5.73}$$

We now explore some of the properties of such dynamics. We assert that a fundamental property of such dynamics is that the stationary distribution they lead to is the time reverse of that reached by both the forward and adjoint dynamics. Using the master equation approach this is simple to demonstrate. Starting with the balance relation in the forward dynamics

$$\sum_{\boldsymbol{x}\neq\boldsymbol{x}'} P^{\mathrm{F,st}}(\boldsymbol{x}', \boldsymbol{\lambda}(t))T(\boldsymbol{x}|\boldsymbol{x}', \boldsymbol{\lambda}(t)) = \sum_{\boldsymbol{x}\neq\boldsymbol{x}'} P^{\mathrm{F,st}}(\boldsymbol{x}, \boldsymbol{\lambda}(t))T(\boldsymbol{x}'|\boldsymbol{x}, \boldsymbol{\lambda}(t)), \tag{5.74}$$

the equivalent expression for the reversed adjoint dynamics reads

$$\sum_{\boldsymbol{x}\neq\boldsymbol{x}'} P^{\mathrm{F,ad\text{-}rev,st}}(\boldsymbol{x}', \boldsymbol{\lambda}(t))T^{\text{ad-rev}}(\boldsymbol{x}|\boldsymbol{x}', \boldsymbol{\lambda}(t)) = \sum_{\boldsymbol{x}\neq\boldsymbol{x}'} P^{\mathrm{F,ad\text{-}rev,st}}(\boldsymbol{x}, \boldsymbol{\lambda}(t))T^{\text{ad-rev}}(\boldsymbol{x}'|\boldsymbol{x}, \boldsymbol{\lambda}(t)). \tag{5.75}$$

However, since these expressions comprise a sum over all states we may write the balance relation for the forward dynamics

$$\sum_{\boldsymbol{x}\neq\boldsymbol{x}'} P^{\mathrm{F,st}}(\boldsymbol{\varepsilon x}',\boldsymbol{\varepsilon\lambda}(t))T(\boldsymbol{\varepsilon x}|\boldsymbol{\varepsilon x}',\boldsymbol{\varepsilon\lambda}(t)) = \sum_{\boldsymbol{x}\neq\boldsymbol{x}'} P^{\mathrm{F,st}}(\boldsymbol{\varepsilon x},\boldsymbol{\varepsilon\lambda}(t))T(\boldsymbol{\varepsilon x}'|\boldsymbol{\varepsilon x},\boldsymbol{\varepsilon\lambda}(t)). \tag{5.76}$$

Substituting in with the definition of the adjoint and reversed adjoint transition rates then gives

$$\sum_{\boldsymbol{x}\neq\boldsymbol{x}'} P^{\mathrm{F,st}}(\boldsymbol{\varepsilon x},\boldsymbol{\varepsilon\lambda}(t))T^{\text{ad-rev}}(\boldsymbol{x}'|\boldsymbol{x},\boldsymbol{\lambda}(t)) = \sum_{\boldsymbol{x}\neq\boldsymbol{x}'} P^{\mathrm{F,st}}(\boldsymbol{\varepsilon x}',\boldsymbol{\varepsilon\lambda}(t))T^{\text{ad-rev}}(\boldsymbol{x}|\boldsymbol{x}',\boldsymbol{\lambda}(t)). \tag{5.77}$$

Comparison with Eq. (5.75) then demonstrates $P^{\mathrm{F,ad\text{-}rev,st}}(\boldsymbol{x},\boldsymbol{\lambda}(t)) = P^{\mathrm{F,st}}(\boldsymbol{\varepsilon x},\boldsymbol{\varepsilon\lambda}(t))$. To consider the reversed adjoint dynamics in continuous systems we need to examine the adjoint propagator, noting the choice $a = 1/2$,

$$\begin{aligned}
p^{\mathrm{ad}}(\boldsymbol{x},t+dt|\boldsymbol{x}',t;\boldsymbol{\lambda}(t)) &= \sqrt{\frac{1}{(4\pi dt)^N\,|\boldsymbol{D}(\boldsymbol{r},\boldsymbol{\lambda}(t))|}} \\
&\exp\left[\sum_{i,j} -\frac{1}{4dt}\left(-dx_i + (A_i(\boldsymbol{r},\boldsymbol{\lambda}(t)) - \sum_m \frac{\partial D_{im}(\boldsymbol{r},\boldsymbol{\lambda}(t))}{\partial r_m} + \sum_m 2D_{im}(\boldsymbol{r},\boldsymbol{\lambda}(t))\frac{\partial\phi(\boldsymbol{r},\boldsymbol{\lambda}(t))}{\partial r_m}dt\right)\right. \\
&\times D^{-1}_{ij}(\boldsymbol{r},\boldsymbol{\lambda}(t))\left(-dx_j + (A_j(\boldsymbol{r},\boldsymbol{\lambda}(t)) - \sum_n \frac{\partial D_{jn}(\boldsymbol{r},\boldsymbol{\lambda}(t))}{\partial r_n} + \sum_n 2D_{jn}(\boldsymbol{r},\boldsymbol{\lambda}(t))\frac{\partial\phi(\boldsymbol{r},\boldsymbol{\lambda}(t))}{\partial r_n}dt\right) \\
&\left.+\sum_i \frac{dt}{2}\frac{\partial}{\partial r_i}\left(A_i(\boldsymbol{r},\boldsymbol{\lambda}(t)) - \sum_m 2\frac{\partial D_{im}(\boldsymbol{r},\boldsymbol{\lambda}(t))}{\partial r_m} + \sum_m 2D_{im}(\boldsymbol{r},\boldsymbol{\lambda}(t))\frac{\partial\phi(\boldsymbol{r})}{\partial r_m}\right) + \sum_{i,j}\frac{dt}{4}\frac{\partial^2 D_{ij}(\boldsymbol{r},\boldsymbol{\lambda}(t))}{\partial r_i\partial r_j}\right].
\end{aligned} \tag{5.78}$$

If we time reverse all coordinates and the protocol we find that in general we obtain the continuous analogue of Eq. (5.73) if we have

$$A_i^{\text{ad-rev}}(\boldsymbol{x},\boldsymbol{\lambda}(t)) = -A_i^{\mathrm{ir}}(\boldsymbol{x},\boldsymbol{\lambda}(t)) + A_i^{\mathrm{rev}}(\boldsymbol{x},\boldsymbol{\lambda}(t)) + \sum_m 2\frac{\partial D_{im}(\boldsymbol{x},\boldsymbol{\lambda}(t))}{\partial x_m} - \sum_m 2D_{im}(\boldsymbol{x},\boldsymbol{\lambda}(t))\frac{\partial\phi(\boldsymbol{\varepsilon x},\boldsymbol{\varepsilon\lambda}(t))}{\partial x_m}. \tag{5.79}$$

By using Eqs. (5.3) and (5.4) we can then find the reversible and irreversible components of this drift term. Recalling $D_{ij}(\boldsymbol{\varepsilon x},\boldsymbol{\varepsilon\lambda}(t)) = \varepsilon_i\varepsilon_j D_{ij}(\boldsymbol{x},\boldsymbol{\lambda}(t))$ and that $\phi'_i(\boldsymbol{\varepsilon x}) = \partial\phi(\boldsymbol{\varepsilon x})/\partial(\varepsilon_i x_i)$ we identify

$$\begin{aligned}
A_i^{\text{ad-rev,ir}}(\boldsymbol{x},\boldsymbol{\lambda}(t)) &= -A_i^{\mathrm{ir}}(\boldsymbol{x},\boldsymbol{\lambda}(t)) + \sum_m 2\frac{\partial D_{im}(\boldsymbol{x},\boldsymbol{\lambda}(t))}{\partial x_m} \\
&\quad - \sum_m D_{im}(\boldsymbol{x},\boldsymbol{\lambda}(t))\left(\phi'_m(\boldsymbol{x},\boldsymbol{\lambda}(t)) + \varepsilon_m\phi'_m(\boldsymbol{\varepsilon x},\boldsymbol{\varepsilon\lambda}(t))\right)
\end{aligned} \tag{5.80}$$

$$A_i^{\text{ad-rev,rev}}(\boldsymbol{x},\boldsymbol{\lambda}(t)) = A_i^{\mathrm{rev}}(\boldsymbol{x},\boldsymbol{\lambda}(t)) + \sum_m D_{im}(\boldsymbol{x},\boldsymbol{\lambda}(t))\left(\phi'_m(\boldsymbol{x},\boldsymbol{\lambda}(t)) - \varepsilon_m\phi'_m(\boldsymbol{\varepsilon x},\boldsymbol{\varepsilon\lambda}(t))\right). \tag{5.81}$$

Constructing the irreversible and reversible stationary current under the reversed adjoint dynamics by utilising these drift terms and a non-equilibrium potential, $\phi^{\text{ad-rev}}(\boldsymbol{x},\boldsymbol{\lambda}(t)) = \phi(\boldsymbol{\varepsilon x},\boldsymbol{\varepsilon\lambda}(t))$, we have

$$J_i^{\text{ad-rev,ir,st}}(\boldsymbol{x},\boldsymbol{\lambda}(t)) = \left(A_i^{\text{ad-rev,ir}}(\boldsymbol{x},\boldsymbol{\lambda}(t)) - \sum_m \frac{\partial D_{im}(\boldsymbol{x},\boldsymbol{\lambda}(t))}{\partial x_m} + \sum_m D_{im}(\boldsymbol{x},\boldsymbol{\lambda}(t))\frac{\partial \phi(\boldsymbol{\varepsilon x},\boldsymbol{\varepsilon\lambda}(t))}{\partial x_m}\right) e^{-\phi(\boldsymbol{\varepsilon x},\boldsymbol{\varepsilon\lambda}(t))} \tag{5.82}$$

$$J_i^{\text{ad-rev,rev,st}}(\boldsymbol{x},\boldsymbol{\lambda}(t)) = A_i^{\text{ad-rev,rev}}(\boldsymbol{x},\boldsymbol{\lambda}(t))e^{-\phi(\boldsymbol{\varepsilon x},\boldsymbol{\varepsilon\lambda}(t))}. \tag{5.83}$$

Finally by comparison with

$$J_i^{\text{ir,st}}(\boldsymbol{x},\boldsymbol{\lambda}(t)) = \left(A_i^{\text{ir}}(\boldsymbol{x},\boldsymbol{\lambda}(t)) - \sum_m \frac{\partial D_{im}(\boldsymbol{x},\boldsymbol{\lambda}(t))}{\partial x_m} + \sum_m D_{im}(\boldsymbol{x},\boldsymbol{\lambda}(t))\frac{\partial \phi(\boldsymbol{x},\boldsymbol{\lambda}(t))}{\partial x_m}\right) e^{-\phi(\boldsymbol{x},\boldsymbol{\lambda}(t))} \tag{5.84}$$

$$J_i^{\text{rev,st}}(\boldsymbol{x},\boldsymbol{\lambda}(t)) = A_i^{\text{rev}}(\boldsymbol{x},\boldsymbol{\lambda}(t))e^{-\phi(\boldsymbol{x},\boldsymbol{\lambda}(t))} \tag{5.85}$$

we find, by rearranging for A_i^{ir}, substituting into $A_i^{\text{ad-rev,ir}}$ and then $J_i^{\text{ad-rev,ir,st}}$ and likewise for the reversible counterparts, that such dynamics lead to a system with the time reversed stationary state and adapted stationary currents given by

$$J_i^{\text{ad-rev,ir,st}}(\boldsymbol{x},\boldsymbol{\lambda}(t)) = -J_i^{\text{ir,st}}(\boldsymbol{x},\boldsymbol{\lambda}(t))e^{\phi(\boldsymbol{x},\boldsymbol{\lambda}(t))-\phi(\boldsymbol{\varepsilon x},\boldsymbol{\varepsilon\lambda}(t))} \tag{5.86}$$

$$\begin{aligned} J_i^{\text{ad-rev,rev,st}}(\boldsymbol{x},\boldsymbol{\lambda}(t)) = &+J_i^{\text{rev,st}}(\boldsymbol{x},\boldsymbol{\lambda}(t))e^{\phi(\boldsymbol{x},\boldsymbol{\lambda}(t))-\phi(\boldsymbol{\varepsilon x},\boldsymbol{\varepsilon\lambda}(t))} \\ &+\sum_m D_{im}(\boldsymbol{x},\boldsymbol{\lambda}(t))\left(\phi'_m(\boldsymbol{x},\boldsymbol{\lambda}(t)) - \varepsilon_m\phi'_m(\boldsymbol{\varepsilon x},\boldsymbol{\varepsilon\lambda}(t))\right)e^{-\phi(\boldsymbol{\varepsilon x},\boldsymbol{\varepsilon\lambda}(t))}. \end{aligned} \tag{5.87}$$

Of particular note is that this implies $J_i^{\text{ad-rev,ir,st}}/p^{\text{ad-rev,st}} = -J_i^{\text{ir,st}}/p^{\text{st}}$. This can be seen as the same symmetry, in part, used to define the adjoint dynamics, $J_i^{\text{ad,ir,st}}/p^{\text{ad,st}} = -J_i^{\text{ir,st}}/p^{\text{st}}$, where of course the fact that $p^{\text{ad,st}} = p^{\text{st}}$ means it reduces to a simple reflection of the current. Whilst the difference, manifest in the reflected stationary state, leads to a complicated conversion of the reversible current ensuring the dynamics are physical, the symmetry in the irreversible current means that the reversed adjoint dynamics have the same mean entropy production rate in the stationary state as the forward process.

One may argue that, being physical, these dynamics are more appropriate in the construction of entropy contributions since using them allows construction only using involutions that produce allowed trajectories under the forward dynamics such that

$$\Delta\mathcal{S}_1 = \ln\mathcal{P}^{\text{F}}[\vec{\boldsymbol{x}}]_{p_0^{\text{F}}} - \ln\mathcal{P}^{\dagger,\text{ad-rev}}[\vec{\boldsymbol{x}}^\dagger]_{\hat{\boldsymbol{\varepsilon}}p_\tau^{\text{F}}} \tag{5.88}$$

$$\Delta\mathcal{S}_2 = \ln\mathcal{P}^{\text{F}}[\vec{\boldsymbol{x}}]_{p_0^{\text{F}}} - \ln\mathcal{P}^{\text{F,ad-rev}}[\vec{\boldsymbol{x}}]_{p_0^{\text{F}}} \tag{5.89}$$

which then obviously lead to the relations at the end of the previous chapter. Since these definitions are entirely equivalent, we gain further insight into the altered behaviour upon inclusion of odd variables. When we consider alternative dynamics we cannot specify any which have the same stationary state, yet only reverse the irreversible, entropy producing, current that is possible when considering systems that consist only of even variables. Instead we must additionally reverse the reversible current (leading to the adjoint dynamics) or find a complicated dynamics that produces the time reversed stationary state with the same entropy production. Doing either necessarily introduces additional terms on top of those

that arise in the original transform used in the construction of the adiabatic and non-adiabatic entropy productions. This can be further elucidated by the recognition that Eqs. (4.33) and (4.34) yield

$$\Delta\mathcal{S}_2 + \Delta\mathcal{S}_3 = \Delta\mathcal{S}_2^{\dagger,\text{ad-rev}} \tag{5.90}$$

and

$$\begin{aligned}\Delta\mathcal{S}_1 + \Delta\mathcal{S}_3 &= \Delta\mathcal{S}_{\text{sys}} + \Delta\mathcal{S}_{\text{ex}} + \Delta\mathcal{S}_3 \\ &= \Delta\mathcal{S}_{\text{sys}} + \Delta\mathcal{S}_{\text{ex}}^{\text{ad-rev}}\end{aligned} \tag{5.91}$$

to be contrasted with the equivalent from the adiabatic and non-adiabatic entropy productions which transform according to

$$\Delta\mathcal{S}_{\text{a}} = \Delta\mathcal{S}_{\text{a}}^{\text{R,ad}} \tag{5.92}$$

and

$$\begin{aligned}\Delta\mathcal{S}_{\text{na}} &= \Delta\mathcal{S}_{\text{sys}} + \Delta\mathcal{S}_{\text{ex}} \\ &= \Delta\mathcal{S}_{\text{sys}} + \Delta\mathcal{S}_{\text{ex}}^{\text{ad}}.\end{aligned} \tag{5.93}$$

5.5 Non-Itō Interpretations of the Underlying SDEs

The preceding development specifically utilised Itō SDEs in their development, but one may have dynamics in which a specific alternative interpretation is utilised. Of course, in these situations the SDEs will have equivalent Itō forms and so the above definitions are sufficient for a description of the entropy, but it is instructive to identify any change in any contributions, specifically the medium entropy contribution, leaving its evolution in terms of increments dx and dt in order to motivate a physical understanding. To do so we understand that any alternative interpretation of the stochastic integral will lead to an effective additional drift term, written as a correction to the A_i term

$$+\alpha \sum_m \sum_k \frac{\partial B_{ik}(\boldsymbol{x})}{\partial x_m} B_{mk}(\boldsymbol{x}), \tag{5.94}$$

where α defines the stochastic integral recalling Itō corresponds to $\alpha = 0$, Stratonovich to $\alpha = 0.5$ and Hänggi-Klimontovich to $\alpha = 1$. We recognise that if we demand $D_{ij}(\boldsymbol{\varepsilon x}) = \varepsilon_i \varepsilon_j D_{ij}(\boldsymbol{x})$ we thus require

$$\sum_k B_{ik}(\boldsymbol{\varepsilon x}) B_{jk}(\boldsymbol{\varepsilon x}) = \varepsilon_i \varepsilon_j \sum_k B_{ik}(\boldsymbol{x}) B_{jk}(\boldsymbol{x}). \tag{5.95}$$

Consequently inserting such transformation properties into the modified drift term in Eq. (5.94) shows that it is necessarily an addition to the irreversible drift. So to alter the results for a modified stochastic integral we need to write

$$A_i^{\text{ir}}(\boldsymbol{x}) \to A_i^{\text{ir}}(\boldsymbol{x}) + \alpha \sum_m \sum_k \frac{\partial B_{ik}(\boldsymbol{x})}{\partial x_m} B_{mk}(\boldsymbol{x}) \tag{5.96}$$

leaving all other terms unchanged. Of course the increment itself dx_i will be different, but it does not change the form of the medium entropy change contribution. Since, in general the modification is in

terms of the noise strength terms $B_{ij}(\boldsymbol{x})$ we first rewrite Eq. (5.26) partly in these terms so that

$$
\begin{aligned}
d\Delta\mathcal{S}_{\mathrm{med}} = &\sum_i -\frac{\partial A^{\mathrm{rev}}(\boldsymbol{x})}{\partial x_i}dt + \sum_{i,j}\frac{D^{-1}_{ij}(\boldsymbol{x})}{2}\left(A^{\mathrm{ir}}_i(\boldsymbol{x})\circ dx_j + A^{\mathrm{ir}}_j(\boldsymbol{x})\circ dx_i\right)\\
&- \frac{D^{-1}_{ij}(\boldsymbol{x})}{2}\Bigg[\left(A^{\mathrm{rev}}_i(\boldsymbol{x})A^{\mathrm{ir}}_j(\boldsymbol{x}) + A^{\mathrm{rev}}_j(\boldsymbol{x})A^{\mathrm{ir}}_i(\boldsymbol{x})\right)dt\\
&+ \left(\left(\sum_n \frac{\partial}{\partial x_n}\left(\sum_k \frac{1}{2}B_{jk}(\boldsymbol{x})B_{nk}(\boldsymbol{x})\right)\right)\circ dx_i + \left(\sum_m \frac{\partial}{\partial x_m}\left(\sum_k \frac{1}{2}B_{ik}(\boldsymbol{x})B_{mk}(\boldsymbol{x})\right)\right)\circ dx_j\right)\\
&- \left(A^{\mathrm{rev}}_j(\boldsymbol{x})\left(\sum_m \frac{\partial}{\partial x_m}\left(\sum_k \frac{1}{2}B_{ik}(\boldsymbol{x})B_{mk}(\boldsymbol{x})\right)\right) + A^{\mathrm{rev}}_i(\boldsymbol{x})\left(\sum_n \frac{\partial}{\partial x_n}\left(\sum_k \frac{1}{2}B_{jk}(\boldsymbol{x})B_{nk}(\boldsymbol{x})\right)\right)\right)dt\Bigg].
\end{aligned}
\tag{5.97}
$$

Making the substitution of Eq. (5.96) then gives

$$
\begin{aligned}
d\Delta\mathcal{S}_{\mathrm{med}} = &\sum_i -\frac{\partial A^{\mathrm{rev}}(\boldsymbol{x})}{\partial x_i}dt + \sum_{i,j}\frac{D^{-1}_{ij}(\boldsymbol{x})}{2}\left(A^{\mathrm{ir}}_i(\boldsymbol{x})\circ dx_j + A^{\mathrm{ir}}_j(\boldsymbol{x})\circ dx_i\right)\\
&- \frac{D^{-1}_{ij}(\boldsymbol{x})}{4}\Bigg[2\left(A^{\mathrm{rev}}_i(\boldsymbol{x})A^{\mathrm{ir}}_j(\boldsymbol{x}) + A^{\mathrm{rev}}_j(\boldsymbol{x})A^{\mathrm{ir}}_i(\boldsymbol{x})\right)dt\\
&- \left((2\alpha-1)\left(\sum_{m,k}\frac{\partial B_{ik}(\boldsymbol{x})}{\partial x_m}B_{mk}(\boldsymbol{x})\right) - \left(\sum_{m,k}B_{ik}(\boldsymbol{x})\frac{\partial B_{mk}(\boldsymbol{x})}{\partial x_m}\right)\right)(\circ dx_j - A^{\mathrm{rev}}_j(\boldsymbol{x})dt)\\
&- \left((2\alpha-1)\left(\sum_{n,k}\frac{\partial B_{jk}(\boldsymbol{x})}{\partial x_n}B_{nk}(\boldsymbol{x})\right) - \left(\sum_{n,k}B_{jk}(\boldsymbol{x})\frac{\partial B_{nk}(\boldsymbol{x})}{\partial x_n}\right)\right)(\circ dx_i - A^{\mathrm{rev}}_i(\boldsymbol{x})dt)\Bigg]
\end{aligned}
\tag{5.98}
$$

which is perhaps not particularly illuminating. However, if we restrict our consideration to systems which have only uncorrelated diffusion such that $B_{ij}(\boldsymbol{x}) = B_{ij}(\boldsymbol{x})\delta_{ij}$ this expression reduces to

$$
\begin{aligned}
d\Delta\mathcal{S}_{\mathrm{med}} = &\sum_i -\frac{\partial A^{\mathrm{rev}}(\boldsymbol{x})}{\partial x_i}dt + \frac{A^{\mathrm{ir}}_i(\boldsymbol{x})}{D_i(\boldsymbol{x})}\circ dx_i - \frac{A^{\mathrm{rev}}_i(\boldsymbol{x})A^{\mathrm{ir}}_i(\boldsymbol{x})}{D_i(\boldsymbol{x})}dt\\
&+ \frac{(\alpha-1)}{D_i(\boldsymbol{x})}\frac{\partial D_i(\boldsymbol{x})}{\partial x_i}(\circ dx_i - A^{\mathrm{rev}}_i(\boldsymbol{x})dt)
\end{aligned}
\tag{5.99}
$$

showing that as we progressively interpret the stochastic integral from Itō to Hänggi-Klimontovich the explicit multiplicative noise term contributes less with it being completely absent from the expression for the latter.

5.6 Division of the Medium Entropy Change

At this point we compare our results with, for example, the approach by Kim et al. [100, 101] where they considered the medium entropy change to be only composed of terms that comprised an energy transfer. In contrast, we have built our definition of entropy based firmly on the principle of irreversibility and that, in a model, such a quantity should represent the total entropy production of the universe. As such our medium entropy change contribution is not simply a heat flow to an idealised heat bath, but a contribution arising from any behaviour arising from possible choices of $\boldsymbol{A}$ and $\boldsymbol{B}$. That having been

said, we can examine the terms that appear in our medium entropy change and align them with the quantities included in other contributions such as the 'pumped' entropy of [100, 101]. To do so we write the contributions, for uncorrelated diffusion, as

$$d\Delta\mathcal{S}_{\text{med}} = \sum_i \underbrace{-\frac{\partial A^{\text{rev}}(\boldsymbol{x})}{\partial x_i}dt}_{\text{Deterministic thermostatting}} + \underbrace{\frac{A_i^{\text{ir}}(\boldsymbol{x})}{D_i(\boldsymbol{x})} \circ dx_i - \frac{A_i^{\text{rev}}(\boldsymbol{x})A_i^{\text{ir}}(\boldsymbol{x})}{D_i(\boldsymbol{x})}dt}_{\text{Hamiltonian}} + \underbrace{\frac{(\alpha-1)}{D_i(\boldsymbol{x})}\frac{\partial D_i(\boldsymbol{x})}{\partial x_i}(\circ dx_i - A_i^{\text{rev}}(\boldsymbol{x})dt)}_{\text{Noise induced}}. \tag{5.100}$$

First, we have identified two terms that represent a heat flow based on a change in Hamiltonian if considered for the usual over or under-damped Langevin descriptions that amount to the medium entropy change considered in stochastic thermodynamics [33], but note that more complex dynamics may lead to terms not identifiable as a heat flow, but having a similar origin based on their form. Second, we have labelled the term that persists in the deterministic limit as a deterministic thermostatting term. This demonstrably contributes when the *reversible* dynamics are non-linear. The pumped entropy in [100, 101] is essentially equivalent to this term, but we stress that the parity of such a term is crucial in identifying it as a distinct contribution, something that was implicitly assumed, but not mentioned in the definition of the pumped entropy. Finally, we label the remainder a noise induced contribution that arises only when there is multiplicative noise. Its relative contribution can then be 'tuned' by an adaptation of the stochastic integration scheme assumed in the original SDEs and can aid as a physical criterion with which to resolve the Itō-Stratonovich dilemma. Such a case where this might apply is for non-linear Brownian motion [117, 118] of the form

$$\begin{aligned} dx &= v dt \\ dv &= -\gamma(v)v dt + \frac{\mathcal{F}(x)}{m}dt + \sqrt{\frac{k_B T\gamma(v)}{m}} * dW. \end{aligned} \tag{5.101}$$

Assuming $\gamma(v)$ is an even function of v such that there is no deterministic thermostatting term one finds that the medium entropy change contains a contribution that equals the heat transfer scaled by the inverse temperature as expected from stochastic energetics, but also a noise induced term since the noise is multiplicative. If, in a model, the environment consists solely of an idealised heat bath for which the only physical result is a heat transfer term then the choice of α which achieves this is $\alpha = 1$ which corresponds to a Hänggi-Klimontovich or post-point discretisation scheme.

5.7 Ambiguity in Time Reversal

So far the development of the entropy production and its constituent components in this and the previous chapter have very deliberately left the nature of the system in question in as general terms as possible. This is motivated, naturally, by a desire for generality so as to be applicable to a range of possible physical (or otherwise) situations. However, we draw attention to the fact that whilst the model system is typically specified by $\boldsymbol{A}$ and $\boldsymbol{B}$ (or transition rates T), there is an additional specification involved in the construction of entropy production, namely the nature of time reversal manifest in the choice $\boldsymbol{\varepsilon x}$ and $\boldsymbol{\varepsilon\lambda}^{\text{F}}$. Clearly, the precise way we interpret such a choice can have radically different effects on what we consider to be the thermodynamics of the system. One may reason that this means we can constrain such a choice based on physical behaviour; for example when the environment is entirely thermal, one

should expect the medium entropy change to be in the form of a recognisable energy transfer as heat. This, of course, is reasonable, but only raises the question of what one should do when one cannot identify the environment in this way. This issue is compounded further if one considers that, as for feedback control, the nature of how the sequence of the reverse protocol is determined is a choice that must be reasoned. Given such a choice in these situations we argue that the nature of time reversal becomes something of a delicate question rooted, in part, in where one places the boundary of the system.

As a starting point we consider the evolution of both the system and the protocol in the scenario considered in this chapter. For the sake of argument, we imagine the protocol can be written as some deterministically evolving variable defined by a differential equation, but note this simply serves to illustrate rather than be exhaustive or rigorous. Here we have a stochastic set of system variables and some deterministic protocol

$$\begin{aligned} d\boldsymbol{x}(t) &= \boldsymbol{A}(\boldsymbol{x}(t), \boldsymbol{\lambda}(t))dt + \boldsymbol{B}(\boldsymbol{x}(t), \boldsymbol{\lambda}(t))d\boldsymbol{W} \\ d\boldsymbol{\lambda}(t) &= \boldsymbol{C}(\boldsymbol{\lambda}(t), t)dt. \end{aligned} \tag{5.102}$$

What was crucial in the development of the entropy production was that the specific evolution of the protocol was independent of the evolution of the system. We consider the irreversibility of the *system* by time reversing the protocol in the dynamics for $\boldsymbol{x}$, but we do not require or consider reversibility in the evolution of the protocol. Explicitly, whilst we might be able to write $\boldsymbol{\lambda}$ as an evolving variable, it is entirely independent of all system variables and therefore it is natural to consider it not to be a dynamical variable and so we don't include, for example, an entropy production contribution based on the structure of the $\boldsymbol{C}$ term. This implies a definite boundary for the system which evolves according to the a priori specified evolution of the protocol meaning the system does not include whatever environmental feature in reality, such as an external agent, has determined the protocol. This however, raises some questions. If we do not consider the evolution of the protocol to be a set of dynamics, such that it is independent of the system, should it obey the usual time reversal involutions if there is no obvious physical constraint that necessitates it? Even more probing is how one should interpret protocols that use feedback and evolve according to, for example,

$$d\boldsymbol{\lambda}(\boldsymbol{x}, t) = \boldsymbol{C}(\boldsymbol{x}(t), \boldsymbol{\lambda}(t), t)dt \tag{5.103}$$

or, as a rudimentary example of feedback control based on imprecise measurement,

$$d\boldsymbol{\lambda}(\boldsymbol{x}, t) = \boldsymbol{C}(\boldsymbol{x}(t), \boldsymbol{\lambda}(t), t)dt + \boldsymbol{D}(\boldsymbol{x}(t), \boldsymbol{\lambda}(t), t)d\boldsymbol{W}. \tag{5.104}$$

In such cases, can one really claim that the protocol is not, in some way, a representation of some dynamical variable that should be included in the system? Further, if we don't consider the evolution of the protocol as a set of dynamics, how should one define the adapted probabilistic evolution of $\boldsymbol{x}$ in both forward and reverse processes? An answer is provided by the definitions commonly used in the literature [102, 104] and introduced in Sect. 3.6.2. One considers the probabilistic behaviour of the joint process. This decomposes to a path probability based on an effective, independent, protocol and a conditional probability of observing that protocol. Then the reverse protocol is considered as the deterministic reverse of the forward protocol. This has two notable features. First it explicitly avoids the time reversal and thus the 'energetics' of the protocol. Second, and possibly more important is that whilst $\boldsymbol{\lambda}(\boldsymbol{x})$ is a (possibly stochastic itself) function of a stochastic variable necessitating the forward path probability to be considered a joint probability because of this, the distribution representing the stochastic behaviour of the system, and thus the system entropy, is only considered in terms of $\boldsymbol{x}$. This is

a curious dichotomy associated with feedback control, the protocol is in some sense considered to be both stochastic and deterministic in that whilst there will naturally be a distribution of possible outcomes $\boldsymbol{\lambda}(\boldsymbol{x}(t),t)$ there is no 'system' entropy production associated with it. It is to some extent both intimately interacting with, yet not part of the system. If one considers feedback control to be synonymous with Maxwell's demon such a specification, and the consequence that it leads to a thermodynamic information contribution, seem appropriate if possibly physically dubious: the protocol represents an agent whose behaviour depends upon the system, this renders it a stochastic variable, yet it is entirely energetically independent of it, reflected by the absence of its time reversal and its lack of inclusion in the change in system entropy. As such one could characterise such a set up as lacking a defined boundary between system and environment. Such a suggestion also implies that assuming less ideal (or indeed impossible) conditions one might define the time reversal differently. For example, perhaps the system entropy should be based on the joint probability of the protocol and what was previously characterised as the system and time reversal in the protocol should be included; this simply amounts to a wider definition of the thermodynamic system to include the measurement device. Alternatives, however could also be argued. One might suggest that one requires time reversal in the protocol, but that it should not be included in the system entropy. This would lead to a reversed total path probability functional of the form

$$\mathcal{P}^{\vec{\boldsymbol{\lambda}}^\dagger}[\vec{\boldsymbol{x}}^\dagger,\vec{\boldsymbol{\lambda}}^\dagger]_{\hat{\boldsymbol{\varepsilon}}P_\tau} = \mathcal{P}[\vec{\boldsymbol{\lambda}}^\dagger|\vec{\boldsymbol{x}}^\dagger]\mathcal{P}^{\vec{\boldsymbol{\lambda}}^\dagger}[\vec{\boldsymbol{x}}^\dagger]_{\hat{\boldsymbol{\varepsilon}}P_\tau} \tag{5.105}$$

thus providing a definition of the total entropy production of the universe as

$$\begin{aligned} k_B^{-1}\Delta\mathcal{S}_{\mathrm{tot}}[\vec{\boldsymbol{x}},\vec{\boldsymbol{\lambda}}^{\mathrm{F}}] &= \ln\mathcal{P}^{\vec{\boldsymbol{\lambda}}^{\mathrm{F}}}[\vec{\boldsymbol{x}},\vec{\boldsymbol{\lambda}}^{\mathrm{F}}]_{p_0^{\mathrm{F}}} - \ln\mathcal{P}^{\vec{\boldsymbol{\lambda}}^\dagger}[\vec{\boldsymbol{x}}^\dagger,\vec{\boldsymbol{\lambda}}^\dagger]_{\hat{\boldsymbol{\varepsilon}}p_\tau^{\mathrm{F}}} \\ &= \ln\frac{\mathcal{P}^{\vec{\boldsymbol{\lambda}}^{\mathrm{F}}}[\vec{\boldsymbol{x}}]_{p_0^{\mathrm{F}}}}{\mathcal{P}^{\vec{\boldsymbol{\lambda}}^\dagger}[\vec{\boldsymbol{x}}^\dagger]_{\hat{\boldsymbol{\varepsilon}}p_\tau^{\mathrm{F}}}} + \ln\frac{\mathcal{P}[\vec{\boldsymbol{\lambda}}^{\mathrm{F}}|\vec{\boldsymbol{x}}]}{\mathcal{P}[\vec{\boldsymbol{\lambda}}^\dagger|\vec{\boldsymbol{x}}^\dagger]}. \end{aligned} \tag{5.106}$$

The final term then might represent the entropy production due to measurement which contributes like an additional medium entropy change contribution in much the same way as deterministic thermostatting term might. Such a description isn't, necessarily, any less valid than that which arrives at a mutual information contribution, and tellingly one cannot a priori prove one to be wrong: they both represent the total entropy of the universe, both obey fluctuation theorems and are both rigorously positive in the mean. However, their difference does imply something quite different about how time reversal occurs and the implication about who or what is measuring or interacting with the system and what that system is defined as.

Equally important, though slightly more straightforward, is the ambiguity that is introduced without feedback control if one needs to determine whether or not a protocol is deemed to be odd. Since that choice, for example in continuous systems, determines A_i^{ir} and A_i^{rev} it in turn defines the entropy production. For example, if one considers a protocol that contains magnetic fields included in a Hamiltonian, one might naturally consider it odd in order for the medium entropy change contribution to represent an energy change, however if a non-conservative force arises due to, for example, the angular momentum of a stirrer or paddle stirring a fluid one must make a decision about whether the nature of the force is relevant in the time reversal. This is similar in nature to the discussion of the stochastic protocol; perhaps if one knows something of the motion of the external agent one needs to consider it in the time reversal, and perhaps not if viewed on a more microscopic level, but this might in turn imply whether the dynamics of these protocol terms should be included in the description of a wider system and whether we have appropriately identified the boundary between system and environment. If one considers all

possible permutations of how one interprets the nature of the protocol one is forced to conclude there is no one correct procedure; like the choice about what causes feedback in a system, it depends on how you interpret your system, the nature of the protocol and perhaps your place as an observer or modeller. In summary we suggest that despite the apparent power of viewing entropy production as a quantity in the context of fluctuation theorems there is vast potential for it to be misused, ultimately since there are as many ways to define the entropy production as there are ways to define alternative reverse paths and all possess properties that make them plausible candidates since they naturally obey the fluctuation theorems. We therefore stress that such a quantity, which if accounted for properly is always the total entropy production of the universe, not of just the system and the physical medium it is in (a distinction which if not made can lead to a potentially infinite, but ultimately unnecessary, series of 'generalisations' of the fluctuation theorems, see for example [101]), is only appropriate if the time reversed path and the boundary of the system is physically sensible which requires clear physical justification. This is not always something of a priority in the literature. For example, experimental proof [119] of an extended fluctuation theorem for feedback control relating dissipative work and information, though impressive, should not be surprising as long as the dynamics in the model are close enough to the real dynamics because the fluctuation theorems are *necessarily* obeyed because of their form, see also [114]. Alternatively, whilst such a result demonstrates an apparent equivalence between entropy and information, it can only ever do so based on the definition of entropy production (or work and free energy) implicitly assumed by the definition of time reversal being used. What it cannot confirm is that such a theoretical procedure is correct, or rather imparts any particularly relevant information over alternative definitions, since another definition of the entropy production of the universe, if using the same dynamics which agree with experiment, will also be confirmed as both rely on the same symmetry in their proof.

Chapter 6

Illustrative Results from Specific Systems

The fluctuation theorems have been examined extensively since their conception with experimental, theoretical and numerical studies having been performed on a great variety of systems. The subjects of such studies have included harmonic potentials [70, 71, 120–125], charged particles in the presence of magnetic fields [126–132], electrical circuits [133–136], colloidal particles in explicitly non-harmonic potentials [137], simple non-equilibrium steady states on periodic geometries [138–141] along with a range of discrete models [142–147] meaning the behaviour of entropy, work and heat distributions are well known for the classes of system treated here. As such we consider systems that illustrate our main result, namely that of an alternative division of entropy production based on the inclusion of odd variables. For continuous systems, the model that we shall utilise will be that of an under-damped Langevin particle where we note the fluctuations of quantities like work and heat have previously been studied [80, 148–152]. Such a description allows a microscopic treatment of entropy production to be applied to situations where use of the simpler, over-damped, Langevin equation would not reproduce the pertinent physical behaviour or would coarse grain away some of the refinements of the model. The first examples use the paradigmatic case of a non-conservatively forced particle on a ring in order to examine the three contributions to entropy production and then consider some more sophisticated versions of such a model, including the introduction of dry friction and non-conservative forcing for relativistic Brownian motion. Next we consider heat conduction in one dimension due to a spatially inhomogeneous temperature field where we demonstrate the validity and indeed necessity of using odd system variables for such a system along with illustrating more complicated dependence and behaviour in the contribution to $\langle \Delta \mathcal{S}_2 \rangle^{\mathrm{F}}_{p_0^{\mathrm{F}}}$ in particular. Finally, some discrete models of full phase space are explored to illustrate the forms of entropy production in such a context along with highlighting the nuances of how entropy production is affected by the coarse-graining implied by the model.

6.1 Particle Driven by a Non-conservative Force

Our first example of a system with odd variables is that of the full phase space Langevin equation where we consider diffusion of a particle on a ring driven by a spatially independent non-conservative force and

spatially independent (additive) noise such that

$$\begin{aligned} dx &= v dt \\ dv &= -\gamma v dt + \frac{\mathcal{F}(t)}{m} dt + \sqrt{\frac{2k_B T\gamma}{m}} dW \end{aligned} \tag{6.1}$$

thus giving terms $A_x^{\mathrm{ir}} = 0$, $A_x^{\mathrm{rev}} = v$, $A_v^{\mathrm{ir}} = -\gamma v$, $A_v^{\mathrm{rev}} = \mathcal{F}(t)/m$, $D_x = 0$ and $D_v = k_B T\gamma/m$. The key feature of the model is that for any non-zero value of $\mathcal{F}(t)$ there will exist a stationary solution which is Gaussian in v (with a non-zero mean such that it is asymmetric in the odd variable about the origin) and uniform in x due to the symmetry of the problem thus allowing stationarity, yet a constant non-zero particle flux in x. Further, any relaxation from a given stationary state caused by changes to the non-conservative force (or any other system parameter) will then also result in a uniform distribution in x for all time by the translational symmetry. As such we may proceed by considering the marginalised velocity distribution when starting from a stationary state. Exploiting the fact that the initial Gaussian solution will remain Gaussian for any $\mathcal{F}(t)$, we can parametrise a transient solution to the Fokker-Planck equation

$$p^{\mathrm{F}}(x, v, t) \propto \sqrt{\frac{m}{2\pi k_B T}} \exp\left[-\frac{m(v - \langle v \rangle)^2}{2k_B T}\right] \tag{6.2}$$

with

$$\frac{d\langle v \rangle}{dt} = \left(\frac{\mathcal{F}}{m} - \gamma \langle v \rangle\right) \tag{6.3}$$

such that

$$\langle v \rangle^{\mathrm{st}} = \frac{\mathcal{F}}{m\gamma}. \tag{6.4}$$

A scenario where closed form solutions exist for all contributions to entropy production is that of an instantaneous step change in the driving force $\mathcal{F}(t)$ so that we have

$$\mathcal{F}(t) = \begin{cases} \mathcal{F}_0 & t < t_0, \\ \mathcal{F}_1 & t \geq t_0, \end{cases} \tag{6.5}$$

and

$$\langle v \rangle(t) = \begin{cases} \mathcal{F}_0/m\gamma & t < t_0, \\ \left(\mathcal{F}_1 + e^{-\gamma(t-t_0)}(\mathcal{F}_0 - \mathcal{F}_1)\right)/m\gamma & t \geq t_0. \end{cases} \tag{6.6}$$

Performing the relevant integrals in Eqs. (5.40), (5.55), (5.66) and (5.71) we then obtain

$$\frac{1}{k_B} \frac{d\langle \Delta \mathcal{S}_{\mathrm{tot}} \rangle^{\mathrm{F}}_{p_0^{\mathrm{F,st}}}}{dt} = \begin{cases} \mathcal{F}_0^2/m\gamma k_B T & t < t_0, \\ \left(\mathcal{F}_0 + \mathcal{F}_1(e^{\gamma(t-t_0)} - 1)\right)^2 e^{-2\gamma(t-t_0)}/m\gamma k_B T & t \geq t_0, \end{cases} \tag{6.7}$$

$$\frac{1}{k_B} \frac{d\langle \Delta \mathcal{S}_1 \rangle^{\mathrm{F}}_{p_0^{\mathrm{F,st}}}}{dt} = \begin{cases} 0 & t < t_0, \\ e^{-2\gamma(t-t_0)}(\mathcal{F}_0 - \mathcal{F}_1)^2/m\gamma k_B T & t \geq t_0, \end{cases} \tag{6.8}$$

$$\frac{1}{k_B} \frac{d\langle \Delta \mathcal{S}_2 \rangle^{\mathrm{F}}_{p_0^{\mathrm{F,st}}}}{dt} = \begin{cases} \mathcal{F}_0^2/m\gamma k_B T & t < t_0, \\ \mathcal{F}_1^2/m\gamma k_B T & t \geq t_0, \end{cases} \tag{6.9}$$

and

$$\frac{1}{k_B} \frac{d\langle \Delta \mathcal{S}_3 \rangle^{\mathrm{F}}_{p_0^{\mathrm{F,st}}}}{dt} = \begin{cases} 0 & t < t_0, \\ -2e^{-\gamma(t-t_0)}\mathcal{F}_1(\mathcal{F}_1 - \mathcal{F}_0)/m\gamma k_B T & t \geq t_0. \end{cases} \tag{6.10}$$

Choosing the specific case of a reversal of the driving force such that it changes from $\mathcal{F}_0 = 1$ to $\mathcal{F}_1 = -1$ at time $t_0 = 1$ and employing units $k_B = m = \gamma = T = 1$, we can generate the results shown in Figs. 6.1 and 6.2.

We note first that the mean rates of change of all three contributions $\Delta\mathcal{S}_{\rm tot}$, $\Delta\mathcal{S}_1$ and $\Delta\mathcal{S}_2$ are positive,

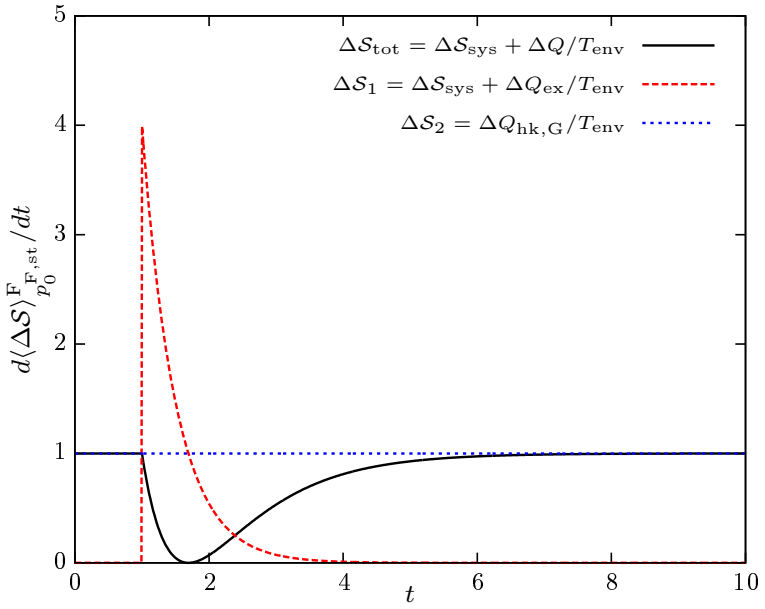

Figure 6.1: Positive mean rates of dimensionless entropy change against time for a non-conservatively forced Langevin particle, where we consider the transition between stationary states of a driven particle on a ring with $\mathcal{F}_0 = 1$, $\mathcal{F}_1 = -1$, $t_0 = 1$ and $k_B = m = \gamma = T = 1$.

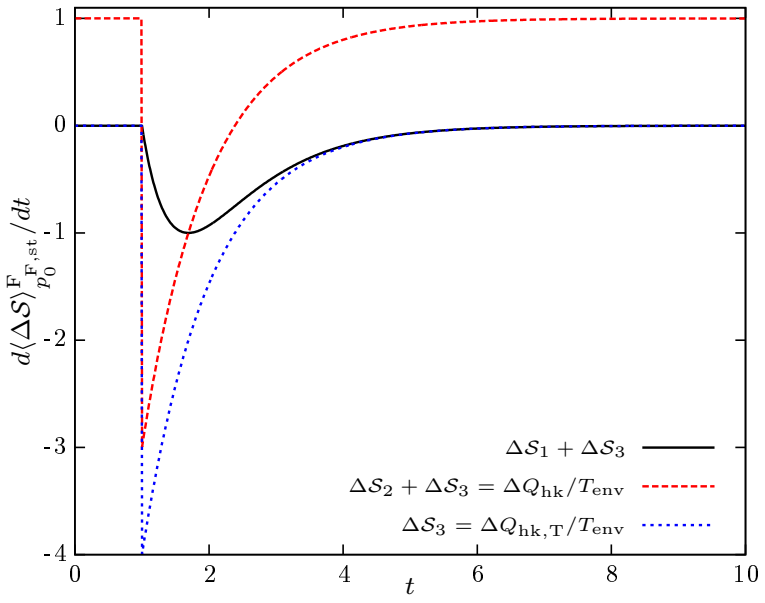

Figure 6.2: Unbounded mean rates of dimensionless entropy change for a non-conservatively forced Langevin particle, where we consider the transition between stationary states of a driven particle on a ring with $\mathcal{F}_0 = 1$, $\mathcal{F}_1 = -1$, $t_0 = 1$ and $k_B = m = \gamma = T = 1$.

reflecting their adherence to an IFT. All three mean rates of change are constant for $t < t_0 = 1$, are perturbed by the change in direction of the force, and relax back to constant values consistent with the transition between the stationary states. A key feature of this behaviour is that upon perturbation, the total entropy production rate *decreases* which would not emerge using an over-damped description of the dynamics. This feature can be explained by the existence of the $d\langle\Delta\mathcal{S}_3\rangle^{\mathrm{F}}_{p_0^{\mathrm{F}}}/dt$ contribution to the mean entropy production rate, which may take negative values depending on the relationship between the instantaneous distribution and the stationary distribution. In this specific case, the large negative value for $d\langle\Delta\mathcal{S}_3\rangle^{\mathrm{F}}_{p_0^{\mathrm{F}}}/dt$ indicates that upon reversal of the force the instantaneous distribution corresponds to particle motion, on average, in a direction counter to that expected to result from the new value of the force. The velocity distribution does relax, of course, to the stationary distribution that corresponds to the new value of the force and so the mean rate of change of $\Delta\mathcal{S}_3$ decays away. An important point to draw from Fig. 6.2 is that $\Delta\mathcal{S}_3$, $\Delta\mathcal{S}_1 + \Delta\mathcal{S}_3$ and $\Delta\mathcal{S}_2 + \Delta\mathcal{S}_3$ cannot be expected, in general, to be positive, reflecting that they cannot be expressed in the form of Eq. (4.2) and thus do not obey IFTs. This means previous approaches where the entropy production can always be divided into two positive quantities [36, 98, 99] and the house-keeping heat can be expected to obey an IFT [94], do not extend to the system considered here.

We consider this example to be a helpful illustration of how entropy production cannot always be divided into two contributions which derive from relaxation, and an absence of detailed balance owing to a non-equilibrium constraint, respectively. Explicitly, the non-equilibrium constraint here is the constant force which produces entropy in the stationary state by inducing a constant flux around the ring. The mean rate of entropy production in that stationary state is characterised by $d\langle\Delta\mathcal{S}_2\rangle^{\mathrm{F}}_{p_0^{\mathrm{F}}}/dt$ which remains constant throughout the process owing to the constant magnitude of the force which is applied. However, both $\Delta\mathcal{S}_2$ and $\Delta\mathcal{S}_3$ are non-zero only in the presence of a non-equilibrium constraint which breaks detailed balance. At the same time the *mean* rate of change of $\Delta\mathcal{S}_3$ is non-zero only when the distribution is relaxing to a new stationary solution in the same manner as $\Delta\mathcal{S}_1$. Whilst $\Delta\mathcal{S}_1$ describes the entropy production that arises from an evolution of the probability distribution of a general set of variables, $\Delta\mathcal{S}_3$ expresses what $\Delta\mathcal{S}_1$ explicitly leaves out: the additional impact of relaxation on entropy production that relates to the a priori physical specification of the variables as odd or even. Clearly, given that the non-equilibrium constraint is a force of constant magnitude, reflected by the constant $d\langle\Delta\mathcal{S}_2\rangle^{\mathrm{F}}_{p_0^{\mathrm{F}}}/dt$, it is reasonable to consider the sum of $\Delta\mathcal{S}_1$ and $\Delta\mathcal{S}_3$ as the contribution that arises due to relaxation to a new stationary state, particularly when the form of its mean rate of change in Fig. 6.2 is contrasted with that of $\Delta\mathcal{S}_{\mathrm{tot}}$, $\Delta\mathcal{S}_1$ and $\Delta\mathcal{S}_2$ in Fig. 6.1. We can make the analysis complete by considering the SDEs for all contributions. The explicit Itō forms of Eqs. (5.27), (5.53), (5.62) and (5.68) are given as

$$k_B^{-1} d\Delta\mathcal{S}_{\mathrm{tot}} = -\frac{m}{k_B T}\langle v\rangle dv - \frac{m}{k_B T}\left(v - \langle v\rangle\right)\frac{d\langle v\rangle}{dt}dt + \frac{\mathcal{F}(t)}{k_B T}dx \tag{6.11}$$

$$k_B^{-1} d\Delta\mathcal{S}_1 = \frac{1}{k_B T}\left(\frac{\mathcal{F}(t)}{\gamma} - m\langle v\rangle\right)dv - \frac{m}{k_B T}\left(v - \langle v\rangle\right)\frac{d\langle v\rangle}{dt}dt \tag{6.12}$$

$$k_B^{-1} d\Delta\mathcal{S}_2 = \frac{\mathcal{F}(t)}{\gamma k_B T}dv + \frac{\mathcal{F}(t)}{k_B T}dx \tag{6.13}$$

$$k_B^{-1} d\Delta\mathcal{S}_3 = -\frac{2\mathcal{F}(t)}{\gamma k_B T}dv \tag{6.14}$$

and illustrate the behaviour of all the contributions. $d\Delta\mathcal{S}_{\mathrm{tot}}$ is only zero when $\langle v\rangle = 0$, $\mathcal{F} = 0$ and

$d\langle v\rangle/dt = 0$ meaning the system is in the equilibrium state. $d\Delta\mathcal{S}_1$ is zero whenever $\langle v\rangle = \mathcal{F}/m\gamma$ and $d\langle v\rangle/dt = 0$ corresponding to any stationary state, equilibrium or otherwise, whilst $d\Delta\mathcal{S}_2$ and $d\Delta\mathcal{S}_3$ contribute independently of properties of the distribution (namely $\langle v\rangle$), but only when the non-equilibrium constraint is present such that $\mathcal{F}(t) \neq 0$. $d\Delta\mathcal{S}_3$ however, has a mean contribution of zero at stationarity since $\langle dv\rangle = 0$ for any stationary state. We can calculate distributions of all the contributions, as measured from the force reversal, numerically using the above SDEs and demonstrate the validity of IFTs, where appropriate, in Figs. 6.3 and 6.4. We observe that all distributions take Gaussian form, to be expected as the model is essentially a recasting of the over-damped dragged oscillator found in [123] where the further, but non-general, detailed fluctuation theorem symmetry $p^{\mathrm{F}}(\Delta\mathcal{S}_{\mathrm{tot}} = A)_{p_0^{\mathrm{F,st}}}/p^{\mathrm{F}}(\Delta\mathcal{S}_{\mathrm{tot}} = -A)_{p_0^{\mathrm{F,st}}} = \exp(A)$ has been noted to hold over finite times [123], but stressed elsewhere [114] to be coincidental. Further insight into this coincidence can be derived from the form of the SDEs which yield Gaussian distributions (for the given initial conditions) since they comprise only drift and additive noise terms (that is, no terms of the form $f(v) * dv$). Such properties however, do not distract from the nature of the contributions which can be readily observed: the distributions in $\Delta\mathcal{S}_1$ and $\Delta\mathcal{S}_3$ develop fastest at first reflecting the initially fast response of the distribution to the change in force. However, distributions for both $\Delta\mathcal{S}_2$ and $\Delta\mathcal{S}_{\mathrm{tot}}$ develop steadily, owing to their contributions being characterised by steady heat dissipation. As such, as time progresses, the distribution of $\Delta\mathcal{S}_1$ ceases to develop as the system reaches the new stationary state and the distributions of $\Delta\mathcal{S}_2$ and $\Delta\mathcal{S}_{\mathrm{tot}}$ continue to shift to the right until they eventually dominate. Similarly for $\Delta\mathcal{S}_3$, we observe here that the distribution stops evolving despite receiving non-zero contributions. For completeness we investigate

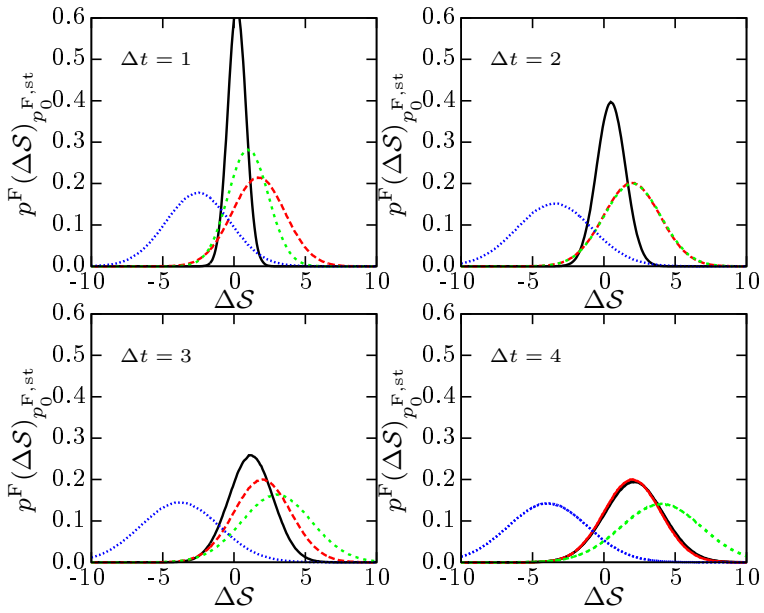

Figure 6.3: Distributions of entropy productions $\Delta\mathcal{S}_{\mathrm{tot}}$ (solid black line), $\Delta\mathcal{S}_1$ (wide dashed red line), $\Delta\mathcal{S}_2$ (narrow dashed green line) and $\Delta\mathcal{S}_3$ (dotted blue line) measured at times $\Delta t = t - t_0 = 1$, $\Delta t = 2$, $\Delta t = 3$ and $\Delta t = 4$ after the reversal of the force for $\mathcal{F}_0 = 1$, $\mathcal{F}_1 = -1$, $t_0 = 1$ and $k_B = m = \gamma = T = 1$. Note that for $\Delta t = 4$ the lines for $\Delta\mathcal{S}_{\mathrm{tot}}$ and $\Delta\mathcal{S}_1$ overlap. We performed 7.5×10^6 Monte Carlo runs with time step $dt = 1 \times 10^{-3}$ to generate the results.

the same model with a less trivial time dependence in the non-conservative force, along with its approach to the over-damped limit where such systems have been considered previously [98, 153]. We employ the force protocol

$$\mathcal{F}(t) = 1.5 - 0.5\tanh(-5(t-1)) \tag{6.15}$$

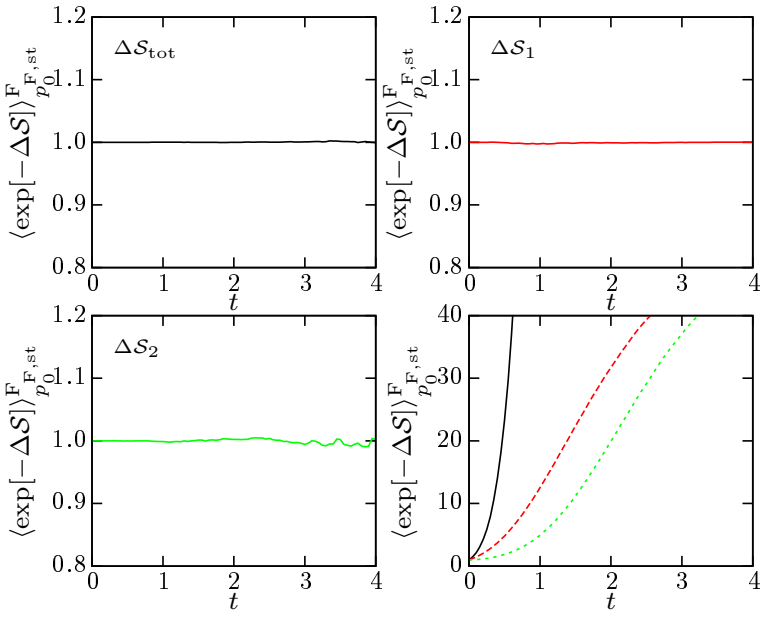

Figure 6.4: Illustration of adherence to IFTs by consideration of the average $\langle\exp[-\Delta\mathcal{S}]\rangle^{\rm F}_{p_0^{\rm F}}$ against time, $\Delta t = t - t_0$ after the force reversal, for $\Delta\mathcal{S}_{\rm tot}$, $\Delta\mathcal{S}_1$ and $\Delta\mathcal{S}_2$ (indicated) and the failure to adhere to an IFT of $\Delta\mathcal{S}_3$ (solid black line, fourth subplot), $\Delta\mathcal{S}_1 + \Delta\mathcal{S}_3$ (wide dashed red line, fourth subplot) and $\Delta\mathcal{S}_2 + \Delta\mathcal{S}_3$ (narrow dashed green line, fourth subplot) for $\mathcal{F}_0 = 1$, $\mathcal{F}_1 = -1$, $t_0 = 1$ and $k_B = m = \gamma = T = 1$.

and perform the calculations numerically for two values of damping coefficient, $\gamma = 1$ and $\gamma = 5$. We point out again that the meaning of $d\langle\Delta\mathcal{S}_2\rangle^{\rm F}_{p_0^{\rm F}}/dt$ for this system is easily elucidated since the non-equilibrium constraint, $\mathcal{F}(t)$, being phase space independent, leads to $J_v^{\rm ir,st} \propto p^{\rm F,st}$ so that

$$\frac{1}{k_B}\frac{d\langle\Delta\mathcal{S}_2\rangle^{\rm F}_{p_0^{\rm F}}}{dt} = \frac{1}{k_B}\frac{d\langle\Delta\mathcal{S}_{\rm tot}\rangle^{\rm F,st}_{p_0^{\rm F,st}}}{dt} = \frac{\mathcal{F}(t)^2}{m\gamma k_B T}. \tag{6.16}$$

The mean contributions for such a protocol for two values of the damping coefficient, again starting from the stationary state, are shown in Fig. 6.5. Note that in this case the contribution $d\langle\Delta\mathcal{S}_3\rangle^{\rm F}_{p_0^{\rm F}}/dt$ is positive. This reflects the fact that as the non-conservative force decreases, the instantaneous distribution corresponds to a greater average particle flux in x than would be expected from the stationary distribution that corresponds to the instantaneous value of the force. As such, calculating the entropy contribution arising due to $\mathcal{F}(t)$ using its instantaneous value ($d\langle\Delta\mathcal{S}_2\rangle^{\rm F}_{p_0^{\rm F}}/dt$) underestimates the actual instantaneous particle flux in x and thus entropy production. This discrepancy is then corrected by $d\langle\Delta\mathcal{S}_3\rangle^{\rm F}_{p_0^{\rm F}}/dt$. As γ increases, the asymmetry of the stationary state (in velocity) decreases and the contribution from $\Delta\mathcal{S}_3$ diminishes. Consequently, the two stationary distributions become increasingly similar, meaning the contribution $\Delta\mathcal{S}_1$ also diminishes rendering the total entropy production almost entirely comprised of the contribution from $\Delta\mathcal{S}_2$. When the full over-damped limit is taken $\Delta\mathcal{S}_2$ is the only contribution and the results map onto those found in [98].

6.2 Entropy Production in the Presence of Dry Friction

We have seen how the three contributions to entropy production contribute for a non-conservatively forced Langevin particle. This model however, has, to a certain extent, somewhat unremarkable features

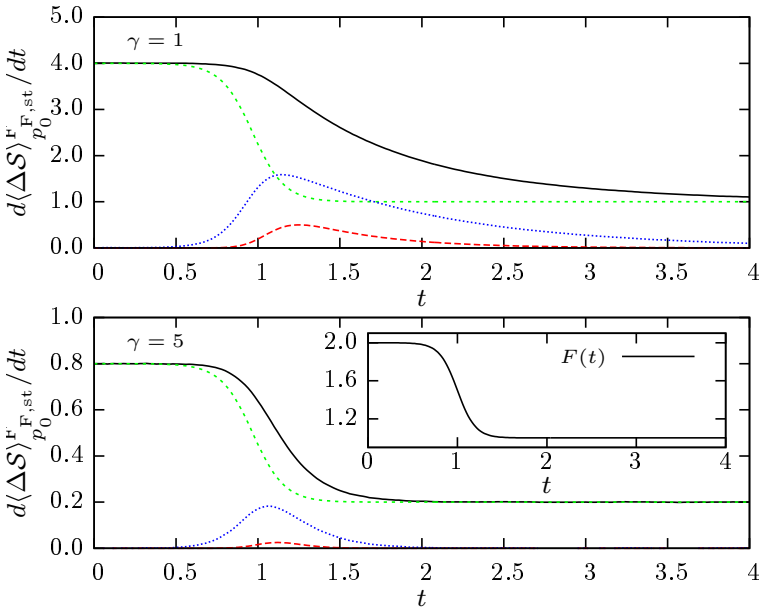

Figure 6.5: Mean rates of change of $\Delta\mathcal{S}_1$ (wide dashed red line), $\Delta\mathcal{S}_2$ (narrow dashed green line), $\Delta\mathcal{S}_3$ (dotted blue line), and their sum $\Delta\mathcal{S}_{\mathrm{tot}}$ (solid black line) for a non-conservatively forced Langevin particle starting in the stationary state with a time dependence in the applied force given by Eq. (6.15), units $k_B = m = T = 1$ and damping coefficients $\gamma = 1$ (top) and $\gamma = 5$ (bottom).

since, for the solvable situations we considered, the solution to the Fokker-Planck equation is always Gaussian. As such it is of interest to investigate the properties of the entropy production for more complicated, perhaps non-linear, models which don't have this property. In such cases the question of how to proceed then becomes one of finding an interesting enough system with a solvable Fokker-Planck equation. An example which is solvable, through an explicit representation of its propagator, is that of dry friction [154]. This is where the frictional term is of a constant magnitude and depends solely on the sign of the velocity. If such a term were analogously used as a position dependent force within an over-damped regime the potential it would arise from would be a wedge shape. We then consider the equations of motion for dynamics that contain this new dry friction term in addition to the original viscous friction term. This then provides a model of a Langevin particle like we have seen previously, such that there is a noisy dissipative environment, with an additional friction source to model a medium independent stick-slip character in its dynamics. These are then given by

$$\begin{aligned} dx &= v dt \\ dv &= -\gamma_{\mathrm{d}}\sigma(v)dt - \gamma v dt + \frac{\mathcal{F}}{m}dt + \sqrt{2D}dW \end{aligned} \tag{6.17}$$

where γ_{d} is the strength of the dry friction, $\sigma(v)$ is the sign function, γ is the normal, viscous, damping coefficient and D is the diffusion constant providing a measure of the strength of the noise, taken to be white and Gaussian, represented by the Wiener process. We note the absence of an explicit temperature: this is not, necessarily, a thermal system, but rather a model of stick slip dynamics subject to fluctuations. One may of course argue that the noise constitutes a temperature provided by the environment which acts through the viscous friction and noise strength, which would cause us to write $D = k_B T\gamma/m$, but the stationary solution would not be the Maxwell distribution. We note that owing to the discontinuity one usually provides an additional convention for the sign function at the origin

such as $\sigma(0) = 0$ as in [155].

We write the stationary solution in terms of a non-equilibrium potential so that we have

$$p^{\mathrm{F,st}}(v) = \exp\left[-\phi\left(v\right)\right], \tag{6.18}$$

which for the above dynamics requires

$$\phi\left(v\right) = \frac{1}{2}\left(\sqrt{\frac{\gamma}{D}}|v| + \gamma_d\sqrt{\frac{1}{\gamma D}}\right)^2 - \frac{\mathcal{F}}{mD}v + \ln Z, \tag{6.19}$$

where Z is a normalisation constant and plays the role of a non-equilibrium partition sum. When coming to construct an increment in the medium entropy production we should be somewhat careful since A_v^{ir} is discontinuous at the origin and so must recognise that the resultant expression

$$\begin{aligned} k_B^{-1} d\Delta\mathcal{S}_{\mathrm{med}} &= -\frac{\left(\gamma_{\mathrm{d}}\sigma\left(v\right) + \gamma v\right)}{D} \circ dv + \frac{\mathcal{F}\left(\gamma_{\mathrm{d}}\sigma\left(v\right) + \gamma v\right)}{mD}dt \\ &= -\frac{\gamma_{\mathrm{d}}}{D}\sigma\left(v\right) \circ dv - \frac{\gamma}{2D}d\left(v^2\right) + \frac{\mathcal{F}\left(\gamma_{\mathrm{d}}\sigma\left(v\right) + \gamma v\right)}{mD}dt, \end{aligned} \tag{6.20}$$

strictly, like the original SDE for dry friction, is defined for the piecewise domains $v > 0$ and $v < 0$ separately. A practical consequence of this is that when simulating such an SDE numerically, we cannot confidently use the above expression to represent an increment in the medium entropy when approximating it with a non-infinitesimal increment that crosses the origin. As such it is instructive to point out how to interpret the above SDE more practically and to highlight how it is implemented in simulation. To do so we consider the discontinuous parts of the increment in the medium entropy as the sum of two sub-infinitesimal increments which both approach and start from the origin asymptotically. As such, we take, for example, a transition from v which is negative to v' which is positive, and consider the transitions $v \to v_-$ and $v_+ \to v'$ where v_- asymptotically approaches the origin from below and v_+ from above. In this case the first term of Eq. (6.20) is equivalent to

$$\begin{aligned} \lim_{v_-\to 0^-} \frac{1}{2D}(-\gamma_{\mathrm{d}}\sigma(v) - \gamma_{\mathrm{d}}\sigma(v_-))(v_- - v) + \lim_{v_+\to 0^+} \frac{1}{2D}(-\gamma_{\mathrm{d}}\sigma(v_+) - \gamma_{\mathrm{d}}\sigma(v'))(v' - v_+) & \\ = -\lim_{v_-\to 0^-} \frac{\gamma_{\mathrm{d}}}{D}\sigma(v)(v_- - v) - \lim_{v_+\to 0^+} \frac{\gamma_{\mathrm{d}}}{D}\sigma(v')(v' - v_+) & \\ = -\frac{\gamma_{\mathrm{d}}}{D}\left(\sigma(v')v - \sigma(v)v\right) & \\ = -\frac{\gamma_{\mathrm{d}}}{D}d|v| & \end{aligned} \tag{6.21}$$

which reflects the energy change from the wedge like potential associated with the dry friction term. The third term, by defining the fraction of the transition spent in the initial domain about the origin before crossing the origin (for example the fraction for which $v < 0$) as α_v, is given by

$$\lim_{v_-\to 0^-} \frac{\mathcal{F}\gamma_{\mathrm{d}}}{mD}\sigma(v)\alpha_v dt + \lim_{v_+\to 0^+} \frac{\mathcal{F}\gamma_{\mathrm{d}}}{mD}\sigma(v_+)(1-\alpha_v)dt = \frac{\mathcal{F}\gamma_{\mathrm{d}}}{mD}\left(\alpha_v\sigma(v) + (1-\alpha_v)\sigma(v')\right)dt. \tag{6.22}$$

Consequently, we numerically implement the expression

$$k_B^{-1} d\Delta\mathcal{S}_{\mathrm{med}} = -\frac{\gamma_{\mathrm{d}}}{D}d|v| - \frac{\gamma}{2D}d\left(v^2\right) + \frac{\mathcal{F}\gamma_{\mathrm{d}}}{mD}\left(\alpha_v\sigma(v) + (1-\alpha_v)\sigma(v')\right)dt + \frac{\mathcal{F}\gamma}{mD}vdt \tag{6.23}$$

which, whilst not strictly an SDE since it explicitly involves start and end points of the transition and

the quantities $d|v|$ and α_v, reduces to the result in Eq. (6.20) for $\sigma(v) = \sigma(v')$ and can thus be effectively used as a general expression for an increment in the medium entropy change. Using this result and the expected form of the adjoint dynamics (in particular the equivalence between the term in $d|v|$ and the equivalent expression in the non-equilibrium potential) we may construct the SDEs for the three contributions to entropy production as before. These are then given as

$$k_B^{-1} d\Delta\mathcal{S}_1 = -d\ln\left(p^{\mathrm{F}}(x,v,t)\right) - \frac{\gamma}{2D} d\left(v^2\right) + \frac{\mathcal{F}}{mD} dv - \frac{\gamma_{\mathrm{d}}}{D} d|v| \tag{6.24}$$

$$k_B^{-1} d\Delta\mathcal{S}_2 = \frac{\gamma_{\mathrm{d}}\mathcal{F}}{mD}\left(\alpha_v \sigma(v) + (1-\alpha_v)\sigma(v')\right) dt + \frac{\gamma\mathcal{F}}{mD} dx + \frac{\mathcal{F}}{mD} dv \tag{6.25}$$

$$k_B^{-1} d\Delta\mathcal{S}_3 = -\frac{2\mathcal{F}}{mD} dv. \tag{6.26}$$

These can then be used numerically to investigate the properties of such quantities. However, we do not, in general, have an expression for $p^{\mathrm{F}}(x,v,t)$. In order to obtain a usable result in certain circumstances we utilise the form of the propagator in velocity space given by Touchette et al. [155]. The propagator is given as an expansion of parabolic cylinder functions [156] in terms of the eigenvalues, λ_n, of the characteristic equation. It has the form

$$p^{\mathrm{F}}(v',\tau|v,0) = p^{\mathrm{F,st}}(v') + \sum_{n=1}^{\infty} \exp\left(-\gamma\lambda_n\tau\right) \frac{U_n\left(\sqrt{\frac{\gamma}{D}}v'\right) V_n\left(\sqrt{\frac{\gamma}{D}}v\right)}{Z_n}. \tag{6.27}$$

Here λ_n is found from the characteristic equation

$$\begin{aligned}\lambda_n \Bigg(& D_{\lambda_n}\left(\sqrt{\frac{1}{\gamma D}}\left(\gamma_{\mathrm{d}} + \frac{\mathcal{F}}{m}\right)\right) D_{\lambda_n - 1}\left(\sqrt{\frac{1}{\gamma D}}\left(\gamma_{\mathrm{d}} - \frac{\mathcal{F}}{m}\right)\right) \\ & + D_{\lambda_n}\left(\sqrt{\frac{1}{\gamma D}}\left(\gamma_{\mathrm{d}} - \frac{\mathcal{F}}{m}\right)\right) D_{\lambda_n - 1}\left(\sqrt{\frac{1}{\gamma D}}\left(\gamma_{\mathrm{d}} + \frac{\mathcal{F}}{m}\right)\right)\Bigg) = 0\end{aligned} \tag{6.28}$$

where D_{λ_n} is the parabolic cylinder function. Here we have

$$\begin{aligned} U_n\left(\sqrt{\frac{\gamma}{D}}v\right) &= e^{-\frac{\phi(v)}{2}} D_{\lambda_n}\left(\sqrt{\frac{1}{\gamma D}}\left(\gamma v + \gamma_{\mathrm{d}} - \frac{\mathcal{F}}{m}\right)\right) \qquad v \geq 0, \\ U_n\left(\sqrt{\frac{\gamma}{D}}v\right) &= e^{-\frac{\phi(v)}{2}} \frac{D_{\lambda_n}\left(\sqrt{\frac{1}{\gamma D}}\left(\gamma_{\mathrm{d}} - \frac{\mathcal{F}}{m}\right)\right)}{D_{\lambda_n}\left(\sqrt{\frac{1}{\gamma D}}\left(\gamma_{\mathrm{d}} + \frac{\mathcal{F}}{m}\right)\right)} D_{\lambda_n}\left(\sqrt{\frac{1}{\gamma D}}\left(-\gamma v + \gamma_{\mathrm{d}} + \frac{\mathcal{F}}{m}\right)\right) \qquad v \leq 0, \end{aligned} \tag{6.29}$$

$$V_n\left(\sqrt{\frac{\gamma}{D}}v\right) = U_n\left(\sqrt{\frac{\gamma}{D}}v\right) e^{\phi(v)} \tag{6.30}$$

and normalisation constant

$$\begin{aligned} Z_n = \lambda_n & D_{\lambda_n}\left(\sqrt{\frac{1}{\gamma D}}\left(\gamma_{\mathrm{d}} - \frac{\mathcal{F}}{m}\right)\right) D_{\lambda_n - 1}\left(\sqrt{\frac{1}{\gamma D}}\left(\gamma_{\mathrm{d}} - \frac{\mathcal{F}}{m}\right)\right) \\ & \times \frac{\partial}{\partial\lambda_n} \ln\left|\frac{D_{\lambda_n - 1}\left(\sqrt{\frac{1}{\gamma D}}\left(\gamma_{\mathrm{d}} - \frac{\mathcal{F}}{m}\right)\right) D_{\lambda_n}\left(\sqrt{\frac{1}{\gamma D}}\left(\gamma_{\mathrm{d}} + \frac{\mathcal{F}}{m}\right)\right)}{D_{\lambda_n - 1}\left(\sqrt{\frac{1}{\gamma D}}\left(\gamma_{\mathrm{d}} + \frac{\mathcal{F}}{m}\right)\right) D_{\lambda_n}\left(\sqrt{\frac{1}{\gamma D}}\left(\gamma_{\mathrm{d}} - \frac{\mathcal{F}}{m}\right)\right)}\right|. \end{aligned} \tag{6.31}$$

In practice, a set of λ_n to a given cut off are found numerically from the characteristic equation along with the corresponding normalisation coefficients. The parabolic cylinder functions are then calculated using library functions for the gamma function and confluent hypergeometric functions. We may then use the propagator to calculate the transient probability distribution corresponding to the transition from some

given initial distribution to the relevant stationary state. Choosing a given stationary distribution which is different to that used in the propagator (which the system will subsequently relax to) then allows us to examine the transition between stationary states brought about by an instantaneous change in any of the system parameters which could be, for example, the strength of the non-conservative force, the damping coefficient etc.

Defining all parameters a to have initial, $a_0 = [\gamma_0, \gamma_{\mathrm{d},0}, \mathcal{F}_0, m_0, D_0]$, and final, $a_1 = [\gamma_1, \gamma_{\mathrm{d},1}, \mathcal{F}_1, m_1, D_1]$, values before and after the transition we may write

$$p^{\mathrm{F}}(v', \tau) = p^{\mathrm{F,st}}(v', a_1) + \sum_{n=1}^{\infty} \frac{\exp\left[-\gamma_1 \lambda_n(a_1)\tau\right] U_n\left(\sqrt{\frac{\gamma_1}{D_1}}v', a_1\right)}{Z_n(a_1)} C_n(a_1) \tag{6.32}$$

where

$$C_n(a_1) = \int_{-\infty}^{\infty} p^{\mathrm{F}}(v, 0) V_n\left(\sqrt{\frac{\gamma_1}{D_1}}v, a_1\right) dv \tag{6.33}$$

which for $p^{\mathrm{F}}(v, 0) = p^{\mathrm{F,st}}(v, a_0)$ is

$$\begin{aligned} p^{\mathrm{F,st}}(v, a_0) = \sqrt{\frac{\gamma_0}{D_0}} \Bigg[\sqrt{\frac{\pi}{2}} \Bigg(& \exp\left[\frac{\mathcal{F}_0}{2\gamma_0 D_0 m_0}\left(\frac{\mathcal{F}_0}{m_0} - 2\gamma_{\mathrm{d},0}\right)\right] \left(1 + \operatorname{erf}\left(\frac{\frac{\mathcal{F}_0}{m_0} - \gamma_{\mathrm{d},0}}{\sqrt{2\gamma_0 D_0}}\right)\right) \\ & + \exp\left[\frac{\mathcal{F}_0}{2\gamma_0 D_0 m_0}\left(\frac{\mathcal{F}_0}{m_0} + 2\gamma_{\mathrm{d},0}\right)\right] \operatorname{erfc}\left(\frac{\frac{\mathcal{F}_0}{m_0} + \gamma_{\mathrm{d},0}}{\sqrt{2\gamma_0 D_0}}\right)\Bigg)\Bigg]^{-1} \\ & \times \exp\left[-\frac{1}{2}\left(\sqrt{\frac{\gamma_0}{D_0}}|v| + \gamma_{\mathrm{d},0}\sqrt{\frac{1}{\gamma_0 D_0}}\right)^2 + \frac{\mathcal{F}_0}{m_0 D_0} v\right] \end{aligned} \tag{6.34}$$

which is simply the stationary distribution for a_0 with the normalisation coefficient written explicitly. One may represent the parabolic cylinder functions using gamma and confluent hypergeometric functions for which there are standard library functions for their numerical evaluation. As such one can numerically find a set of λ_n, Z_n, C_n and describe the time dependent distribution for v. Practically, one can initialise the distribution simply as all stationary states are piecewise Gaussian. An example is given in Fig. 6.6. Clearly, the representation will be poor for small times and progressively improve.

6.2.1 Example Implementations

Instantaneous Temperature Change

In order to consider all three entropy contributions we must consider situations where we have a means of representing the transient distribution. Using the propagator expansion detailed above we are able to do this when there is an instantaneous change in system parameters. The first of these situations we consider is for a step change in the noise strength, or in more physical terms, temperature to which the particle is exposed. Considering an increase in the noise strength this is characterised by a broadening of the distribution, the outline of which is illustrated in Fig. 6.6 using the expansion detailed above with 80 terms for which the approximated distribution becomes usable after $t \simeq 0.1$ as checked by convergence to that found by numerical simulation. Noting here that the non-equilibrium potential is given by

$$\phi\left(v, \lambda^{\mathrm{F}}(t)\right) = \frac{1}{2}\left(\sqrt{\frac{\gamma}{D(\lambda^{\mathrm{F}}(t))}}|v| + \gamma_d\sqrt{\frac{1}{\gamma D(\lambda^{\mathrm{F}}(t))}}\right)^2 - \frac{\mathcal{F}}{mD(\lambda^{\mathrm{F}}(t))}v + \ln Z(\lambda^{\mathrm{F}}(t)), \tag{6.35}$$

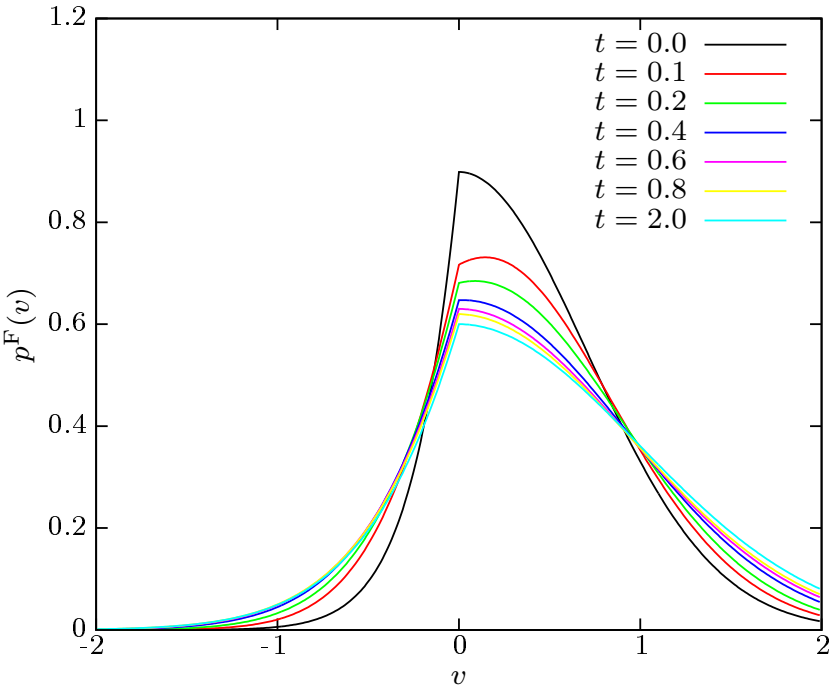

Figure 6.6: Evolution of the velocity distribution following an instantaneous increase in noise strength from $D_0 = 0.5$ to $D_1 = 1$ with $\gamma = 1 = \gamma_{\mathrm{d}} = m = \mathcal{F} = 1$ using 80 terms in the expansion in Eq. (6.32).

.

where we specify $D(\lambda^{\mathrm{F}}(t)) = D(t) = D_0$ for $t < 0$ and $D(t) = D_1$ for $t \geq 0$, and write $\beta_i = D_i^{-1}$, we may consider the stationary to stationary limit of the process and verify a relevant fluctuation theorem for the contribution $\Delta\mathcal{S}_1$. The entropy production contribution for such a process may be simplified as

$$\begin{aligned} k_B^{-1}\Delta\mathcal{S}_1^{\mathrm{st}\to\mathrm{st}} &= \ln\left[\frac{p^{\mathrm{F,st}}(v(0),D_0)}{p^{\mathrm{F,st}}(v(\tau),D_1)}\frac{p^{\mathrm{F,st}}(v(\tau),D_1)}{p^{\mathrm{F,st}}(v(0),D_1)}\right] \\ &= \ln\left[\frac{p^{\mathrm{F,st}}(v(0),D_0)}{p^{\mathrm{F,st}}(v(0),D_1)}\right] \\ &= \left(\gamma_{\mathrm{d}}|v(0)| + \frac{\gamma(v(0))^2}{2} + \frac{\gamma_{\mathrm{d}}^2}{2\gamma} - \frac{\mathcal{F}}{m}v(0)\right)(\beta_1 - \beta_0) + \ln\frac{Z_1}{Z_0} \end{aligned} \tag{6.36}$$

where Z_i is the partition sum. Since this only depends on $v = v(0)$ we may simply write

$$p(k_B^{-1}\Delta\mathcal{S}_1^{\mathrm{st}\to\mathrm{st}})d\left(k_B^{-1}\Delta\mathcal{S}_1^{\mathrm{st}\to\mathrm{st}}\right) = p^{\mathrm{F,st}}(v(0))dv(0). \tag{6.37}$$

With

$$v(0) = \frac{\mathcal{F} \pm m\gamma_d}{m\gamma} \pm \sqrt{\frac{2\left(\left(k_B^{-1}\Delta\mathcal{S}_1 - \ln\frac{Z_1}{Z_0}\right) + \frac{(\beta_1-\beta_0)\mathcal{F}}{m^2\gamma}\left(\mathcal{F} \pm 2m\gamma_{\mathrm{d}}\right)\right)}{\gamma(\beta_1-\beta_0)}} \tag{6.38}$$

we thus find, choosing the physical root,

$$\begin{aligned} p^{\mathrm{F}}(k_B^{-1}\Delta\mathcal{S}_1^{\mathrm{st}\to\mathrm{st}})_{p_0^{\mathrm{F,st}}} = Z_0^{-1}&\left(2(\beta_1-\beta_0)\gamma\left((k_B^{-1}\Delta\mathcal{S}_1^{\mathrm{st}\to\mathrm{st}} - \ln\frac{Z_1}{Z_0}) + \frac{(\beta_1-\beta_0)\mathcal{F}}{2m^2\gamma}(\mathcal{F} - 2m\gamma_d)\right)\right)^{-\frac{1}{2}} \\ &\times \exp\left[-\frac{\beta_0\left(k_B^{-1}\Delta\mathcal{S}_1^{\mathrm{st}\to\mathrm{st}} - \ln\frac{Z_1}{Z_0}\right)}{(\beta_1-\beta_0)}\right]. \end{aligned} \tag{6.39}$$

The symmetry used in the construction of $\Delta\mathcal{S}_1$ arises from the comparison of the normal dynamics and protocol with that of the adjoint dynamics with reversed protocol. As such we can envisage the appropriate reversed adjoint process which, since we have considered $\Delta\mathcal{S}_1^{\mathrm{st}\to\mathrm{st}}$ in the forward process, we also specify starts in the stationary state. Specifically, if we consider the forward protocol to consist of the step change $\Delta\beta = (\beta_1 - \beta_0)$ at $t = 0$ starting from the stationary state characterised by β_0 then the reverse adjoint process consists of the step change $\Delta\beta^{\mathrm{R}} = -\Delta\beta$ at $t = \tau$ starting from the stationary state characterised by β_1 under the adjoint dynamics. However, under the adjoint dynamics, the quantity $\Delta\mathcal{S}_1^{\mathrm{R,ad,st}\to\mathrm{st}}$ is formed only of stationary distributions which are common to both the adjoint and usual forward dynamics. Consequently we can readily find

$$k_B^{-1}\Delta\mathcal{S}_1^{\mathrm{R,ad,st}\to\mathrm{st}} = -\left(\gamma_{\mathrm{d}}|v^{\mathrm{R}}(\tau)| + \frac{\gamma(v^{\mathrm{R}}(\tau))^2}{2} + \frac{\gamma_{\mathrm{d}}^2}{2\gamma} - \frac{\mathcal{F}}{m}v^{\mathrm{R}}(\tau)\right)(\beta_1 - \beta_0) - \ln\frac{Z_1}{Z_0}. \tag{6.40}$$

Given the reversed protocol we also know that in the reversed adjoint process the system is in the stationary state up to $t = \tau$ and so, as above for $\Delta\mathcal{S}_1^{\mathrm{R,ad,st}\to\mathrm{st}}$ and $v(0)$, we can relate the distribution of $\Delta\mathcal{S}_1^{\mathrm{R,ad,st}\to\mathrm{st}}$ and $v^{\mathrm{R}}(\tau)$, choosing the physical root, thus finding

$$\begin{aligned} p^{\mathrm{R,ad}}(k_B^{-1}\Delta\mathcal{S}_1^{\mathrm{R,ad,st}\to\mathrm{st}})_{p_0^{\mathrm{R,st}}} &= \exp\left[-\frac{\beta_1\left(k_B^{-1}\Delta\mathcal{S}_1^{\mathrm{R,ad,st}\to\mathrm{st}} - \ln\frac{Z_0}{Z_1}\right)}{(\beta_0 - \beta_1)}\right] \\ &\times Z_1^{-1}\left(2(\beta_0 - \beta_1)\gamma\left((k_B^{-1}\Delta\mathcal{S}_1^{\mathrm{R,ad,st}\to\mathrm{st}} - \ln\frac{Z_0}{Z_1}) + \frac{(\beta_0 - \beta_1)\mathcal{F}}{2m^2\gamma}(\mathcal{F} - 2m\gamma_d)\right)\right)^{-\frac{1}{2}}. \end{aligned} \tag{6.41}$$

However, because of the parity of this contribution ($\hat{\mathcal{A}}[\vec{\boldsymbol{x}}^{\mathrm{R}}] = \Delta\mathcal{S}_1^{\mathrm{R,ad,st}\to\mathrm{st}}(v^{\mathrm{R}}(\tau)) = -\Delta\mathcal{S}_1^{\mathrm{st}\to\mathrm{st}}(v(0))$ $= -\mathcal{A}[\vec{\boldsymbol{x}}]$) arising since we have specified the stationary to stationary case we expect to observe a fluctuation theorem. As such we may write

$$\begin{aligned} &p^{\mathrm{F}}(k_B^{-1}\Delta\mathcal{S}_1^{\mathrm{st}\to\mathrm{st}})_{p_0^{\mathrm{F,st}}}\exp[-k_B^{-1}\Delta\mathcal{S}_1^{\mathrm{st}\to\mathrm{st}}] = \\ &\quad Z_1^{-1}\left(2(\beta_0 - \beta_1)\gamma\left((-k_B^{-1}\Delta\mathcal{S}_1^{\mathrm{st}\to\mathrm{st}} - \ln\frac{Z_0}{Z_1}) + \frac{(\beta_0 - \beta_1)\mathcal{F}}{2m^2\gamma}(\mathcal{F} - 2m\gamma_d)\right)\right)^{-\frac{1}{2}}\exp\left[\ln\frac{Z_1}{Z_0}\right] \\ &\qquad\times\exp\left[-\frac{\beta_0\left(k_B^{-1}\Delta\mathcal{S}_1^{\mathrm{st}\to\mathrm{st}} - \ln\frac{Z_1}{Z_0}\right)}{(\beta_1 - \beta_0)}\right]\exp\left[-\frac{\beta_1 - \beta_0}{\beta_1 - \beta_0}k_B^{-1}\Delta\mathcal{S}_1^{\mathrm{st}\to\mathrm{st}}\right] \\ &\quad= Z_1^{-1}\left(2(\beta_0 - \beta_1)\gamma\left((-k_B^{-1}\Delta\mathcal{S}_1^{\mathrm{st}\to\mathrm{st}} - \ln\frac{Z_0}{Z_1}) + \frac{(\beta_0 - \beta_1)\mathcal{F}}{2m^2\gamma}(\mathcal{F} - 2m\gamma_d)\right)\right)^{-\frac{1}{2}} \\ &\qquad\times\exp\left[-\frac{\beta_1}{\beta_0 - \beta_1}\left(-k_B^{-1}\Delta\mathcal{S}_1^{\mathrm{st}\to\mathrm{st}} - \ln\frac{Z_0}{Z_1}\right)\right] \\ &\quad= p^{\mathrm{R,ad}}(-k_B^{-1}\Delta\mathcal{S}_1^{\mathrm{R,ad,st}\to\mathrm{st}})_{p_0^{\mathrm{R,st}}} \end{aligned} \tag{6.42}$$

demonstrating a Crooks-like fluctuation theorem for $\Delta\mathcal{S}_1$. We may observe the above forms as slightly modified chi-squared distributions reducing exactly when $\mathcal{F} = 0$. In such cases we may see that it has degree of freedom parameter equal to one. If we were to define our system as consisting of several non-interacting versions of the same system then we would find that the degree of freedom parameter represents the number of particles we consider producing a result equivalent to a convolution of single particle distributions. Interestingly, when setting $\mathcal{F} = 0$ we observe that the functional form for $\Delta\mathcal{S}_1$ is identical to that for the same process in a system where the dry friction term is absent, as in Eq. (6.1) (although the partition sums will differ). However, if we only include the dry friction term (equivalent

to setting $\gamma = 0$ in Eq. (6.17)) we observe an exponential (Erlang) distribution of the form

$$p^{\mathrm{F}}(k_B^{-1}\Delta\mathcal{S}_1^{\mathrm{st}\to\mathrm{st}})_{p_0^{\mathrm{F,eq}}} = \frac{2\beta_0}{(\beta_1-\beta_0)}\exp\left[-\frac{\beta_0\left(k_B^{-1}\Delta\mathcal{S}_1^{\mathrm{st}\to\mathrm{st}} - \ln\frac{Z_1}{Z_0}\right)}{(\beta_1-\beta_0)}\right] \tag{6.43}$$

suggesting the distribution of $\Delta\mathcal{S}_1$ is dominated by higher order terms in the (non-)equilibrium potential.

The full evolution of the transient distribution of $\Delta\mathcal{S}_1$, utilising the propagator expansion, for a particle subject to viscous damping, dry friction and non-conservative forcing is shown in Fig. 6.7 where we observe a peaked distribution steadily approaching the modified chi-squared distribution given above as the process becomes equivalent to the stationary to stationary process. Since we have included a

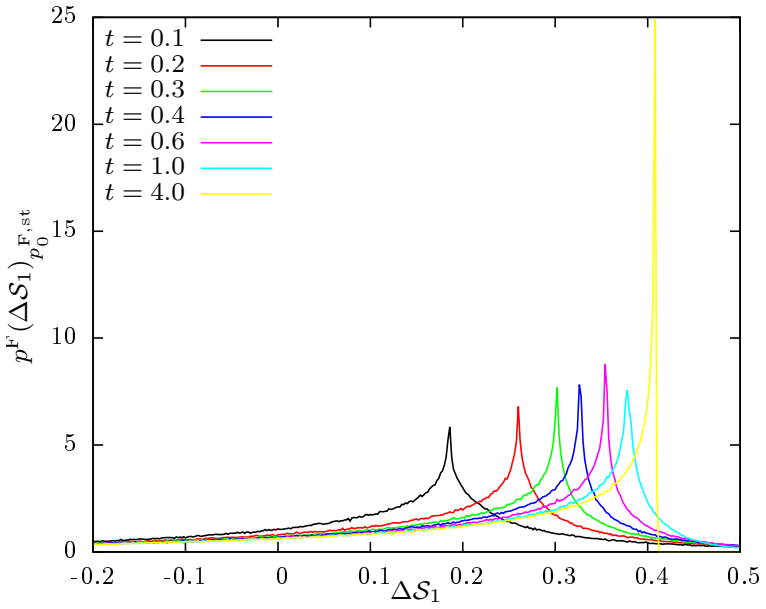

Figure 6.7: Distributions of entropy production contribution $\Delta\mathcal{S}_1$ for times indicated after an instantaneous increase in noise strength from $D_0 = 0.5$ to $D_1 = 1.0$ with $\mathcal{F} = m = \gamma = \gamma_{\mathrm{d}} = 1$. We observe a peaked distribution that steadily approaches the modified chi-squared distribution given in Eq. (6.39).

non-conservative force we also have contributions $\Delta\mathcal{S}_2$ and $\Delta\mathcal{S}_3$. Immediately we see there are some key similarities between the dry friction model and that of the usual Langevin equation. Most notable, is that the nature of the non-equilibrium constraint is the same yielding similar behaviour in both $\Delta\mathcal{S}_2$ and $\Delta\mathcal{S}_3$. For example it is straightforward to show

$$\frac{1}{k_B}\frac{d\langle\Delta\mathcal{S}_2\rangle^{\mathrm{F}}_{p_0^{\mathrm{F}}}}{dt} = \frac{\mathcal{F}(t)^2}{m^2 D} \tag{6.44}$$

given any driving. We illustrate the mean behaviour, as found by a Monte Carlo average of the appropriate SDEs in Eqs. (6.24) to (6.26), of all contributions arising from an instantaneous increase noise strength in Fig. 6.8 where, for comparison, we also show the result for $d\langle\Delta\mathcal{S}_{\mathrm{tot}}\rangle^{\mathrm{F}}_{p^{\mathrm{F,st}},0}/dt$ obtained for the case where $\gamma_d = 0$. The latter can be achieved, analytically through Eq. (5.40), quite simply by noting the solution is always Gaussian following a reversal at time t_0. Using the same methods as in Sect. 6.1 then allows us to parametrise the solution through the Gaussian's variance so that for $t > t_0$, $\sigma^2(t) = e^{-2\gamma(t-t_0)}(D_0 - D_1 + D_1 e^{2\gamma(t-t_0)})$. A consequence of this Gaussian form is that

$d\langle\Delta\mathcal{S}_3\rangle^{\mathrm{F}}_{p_0^{\mathrm{F,st}}}/dt = 0$, something we do not observe when dry friction is included leading to a more complicated structure in the mean total entropy production rate characterised by a non-monotonic decrease in the transition between the stationary states.

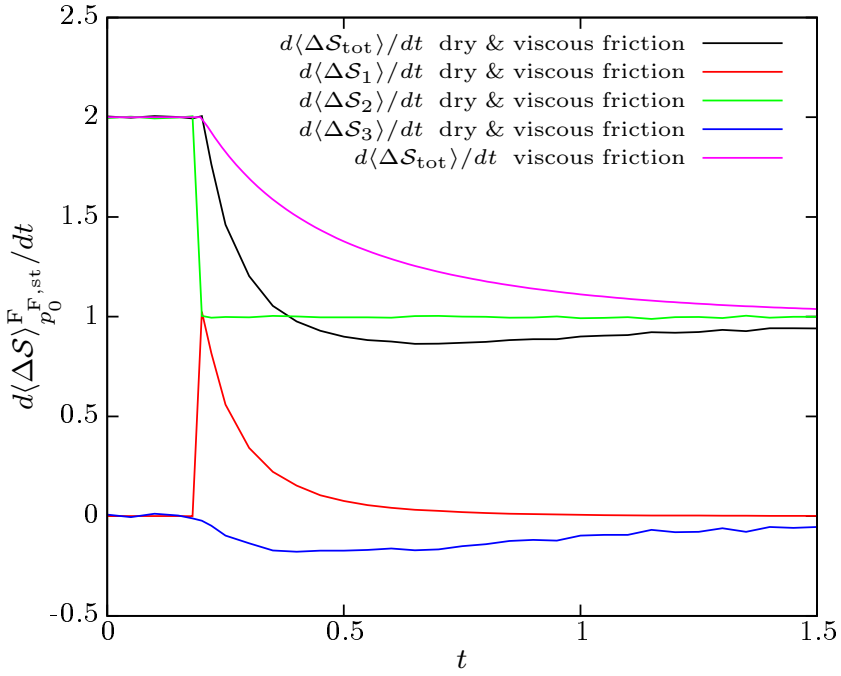

Figure 6.8: Mean entropy production rate contributions for a Langevin particle with both dry and viscous friction following an instantaneous increase in noise strength from $D_0 = 0.5$ to $D_1 = 1$ at time $t = 0.2$ with $\gamma = m = \gamma_d = \mathcal{F} = 1$ calculated using a Monte Carlo average of the SDEs in Eqs. (6.24) to (6.26). Also shown is the mean entropy production rate for the same Langevin particle without dry friction ($\gamma_d = 0$) calculated by an integral of the form in Eq. (5.40).

Force Reversal

For contrast with the previously considered force reversal for the usual Langevin dynamics with viscous friction we can consider the same process in the presence of dry friction. The evolution of the distribution in such a process is illustrated in Fig. 6.9 and as before we found the distribution to be accurate after around $t \simeq 0.1$ when using 80 expansion terms in Eq. (6.32). The question we may then ask is how does the presence of dry friction affect the irreversibility in a force reversal process. We found that the behaviour of $\Delta\mathcal{S}_1$, $\Delta\mathcal{S}_2$ and $\Delta\mathcal{S}_3$ was qualitatively the same both in distribution and mean as for the process in the absence of dry friction, however they were not quantitatively the same giving the total entropy production and thus total irreversibility a different character. The results for two different values of dry friction constant are given in Fig. 6.10. We observe that the introduction of dry friction, whilst having no marked effect on the irreversibility (or entropy production) in the stationary state, shortens the timescale of the transition, as might be expected from the increased damping effect, but also increases the minimum observed irreversibility (specifically when compared to the viscous, $\gamma_d = 0$, case seen in Fig. 6.1) throughout a transition between stationary states. Further increases in the strength of the dry friction amplify both these effects. Such results are to be contrasted with the force reversal results for viscous friction only where we observed a reduction of the total entropy production to zero at $t \simeq 0.6$ after the force reversal as the Gaussian velocity distribution passed through the origin.

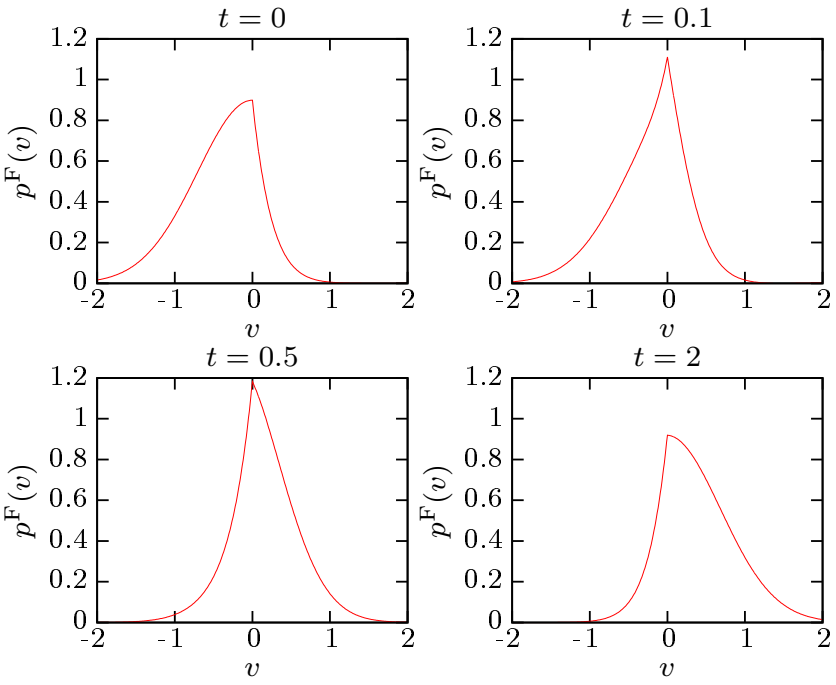

Figure 6.9: Evolution of the velocity distribution following a reversal of force from $\mathcal{F}_0 = -1$ to $\mathcal{F}_1 = 1$ with $\gamma_0 = \gamma_1 = 1$, $\gamma_{\mathrm{d},0} = \gamma_{\mathrm{d},1} = 1$, $D_0 = D_1 = 0.5$ and $m_0 = m_1 = 1$ using 80 terms in the expansion.

.

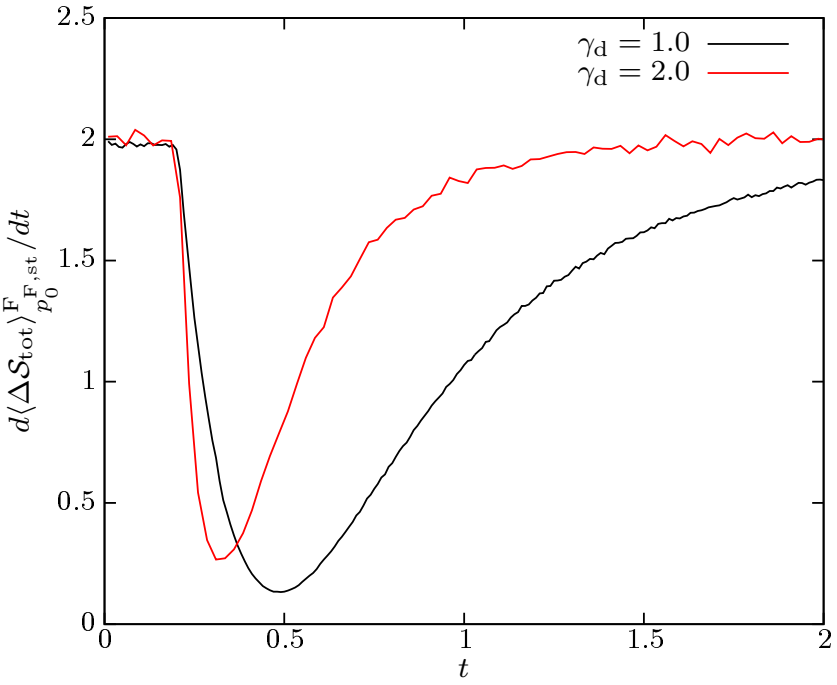

Figure 6.10: Evolution of the mean contribution to the total entropy production following a reversal of force from $\mathcal{F}_0 = -1$ to $\mathcal{F}_1 = 1$ from the stationary state at $t = 0.2$ with $\gamma_0 = \gamma_1 = 1$, $D_0 = D_1 = 0.5$ and $m_0 = m_1 = 1$ using 80 terms in the expansion for two different values of dry friction parameter $\gamma_{\mathrm{d},0} = \gamma_{\mathrm{d},1} = 1$ and $\gamma_{\mathrm{d},0} = \gamma_{\mathrm{d},1} = 2$.

6.3 Relativistic Drift Diffusion, Non-linear Brownian Motion and State Dependent $\langle \Delta \mathcal{S}_2 \rangle^{\mathrm{F}}_{p_0^{\mathrm{F}}}$ Contributions

So far for continuous systems we have seen the general properties of the three contributions to entropy production and have observed that, in the mean, $\Delta\mathcal{S}_1$ and $\Delta\mathcal{S}_3$ contribute transiently, decaying to zero as expected, but with different rates, which along with a constant contribution to $\langle \Delta \mathcal{S}_2 \rangle^{\mathrm{F}}_{p_0^{\mathrm{F}}}$ from a state independent non-equilibrium constraint, defines the form of the mean total entropy production. We now however, consider situations where the $\langle \Delta \mathcal{S}_2 \rangle^{\mathrm{F}}_{p_0^{\mathrm{F}}}$ contribution has a transient character in the absence of driving. Typically this can be achieved by either having a state dependent non-equilibrium constraint such that there is a state dependent irreversible flux or by the presence of multiplicative noise.

An example of where this can be observed is for relativistic Brownian motion. When studying exceptionally fast dynamics the usual Brownian motion has the distinct shortcoming that it permits velocities that exceed the speed of light, as it can receive fluctuations of any size according to the Wiener process. In an attempt to resolve such an issue there is the so-called theory of relativistic Brownian motion which applies the laws of special relativity to continuous stochastic dynamics [157, 158]. The main features of such an approach are that the fluctuations are driven in the momentum space and that the magnitude of these fluctuations is dependent on the momentum in such a way that the magnitude of the velocity is bound by the speed of light. One finds that the strength of the fluctuations follow the Lorentz transformation such that one may construct a consistent SDE for the particle momentum, p_v, in the laboratory frame is given by

$$dp_v = -\gamma p_v dt + \mathcal{F} dt + \sqrt{2k_B T m \gamma \left(1 + \frac{p_v^2}{m^2 c^2}\right)^{\frac{1}{2}}} * dW. \tag{6.45}$$

Since

$$v = \frac{c p_v}{\sqrt{m^2 c^2 + p_v^2}} \tag{6.46}$$

we understand a fluctuation to infinity in p_v corresponds to a fluctuation to the bound $v = c$. Fluctuation theorems have been studied in the context of relativistic Brownian motion before [159], but not in situations where the stationary distribution is non-equilibrium which makes our main division of entropy production relevant. It is well recognised [157, 160] that since we have multiplicative noise we must choose a discretisation procedure for which each leads to different physical behaviour and resultant distributions. It is however, generally accepted that to achieve the Maxwell-Juttner equilibrium distribution [157, 158] one must choose the Hänggi-Klimontovich or post-point evaluation, which we also recognise as the choice which removes all noise induced contributions to the medium entropy change. As such, we formally write

$$dp_v = -\gamma p_v dt + \mathcal{F} dt + \sqrt{2k_B T m \gamma \left(1 + \frac{p_v^2}{m^2 c^2}\right)^{\frac{1}{2}}} \bullet dW. \tag{6.47}$$

By examining the previously unconsidered case of a relativistic particle in a non-equilibrium stationary state by introducing a non-conservative force as before, we first identify the stationary state

$$p^{\mathrm{F,st}}(p_v) \propto \exp\left[-\frac{mc^2}{k_B T}\left(\sqrt{1 + \frac{p_v^2}{m^2 c^2}} - \frac{\mathcal{F}}{mc\gamma}\mathrm{arcsinh}\left[\frac{p_v}{mc}\right]\right)\right] \tag{6.48}$$

which satisfies

$$\frac{\partial}{\partial p_v}\left(\gamma p_v - \mathcal{F}\right) p^{\mathrm{F}}(p_v) + \frac{\partial}{\partial p_v} k_B T m \gamma \left(1 + \frac{p_v^2}{m^2 c^2}\right)^{\frac{1}{2}} \frac{\partial p^{\mathrm{F}}(p_v)}{\partial p_v} = 0, \tag{6.49}$$

being the Fokker-Planck equation according to the Hänggi-Klimontovich interpretation, which allows us to find the following SDEs for entropy production contributions

$$k_B^{-1} d\Delta\mathcal{S}_{\mathrm{tot}} = -d\left(\ln p^{\mathrm{F}}(p_v, x, t)\right) - \frac{p_v}{m k_B T}\left(1 + \frac{p_v^2}{m^2 c^2}\right)^{-\frac{1}{2}} \circ dp_v + \frac{\mathcal{F} p_v}{m k_B T}\left(1 + \frac{p_v^2}{m^2 c^2}\right)^{-\frac{1}{2}} dt \tag{6.50}$$

$$k_B^{-1} d\Delta\mathcal{S}_1 = -d\left(\ln p^{\mathrm{F}}(p_v, x, t)\right) - \frac{p_v}{m k_B T}\left(1 + \frac{p_v^2}{m^2 c^2}\right)^{-\frac{1}{2}} \circ dp_v + \frac{\mathcal{F}}{m k_B T \gamma}\left(1 + \frac{p_v^2}{m^2 c^2}\right)^{-\frac{1}{2}} \circ dp_v \tag{6.51}$$

$$k_B^{-1} d\Delta\mathcal{S}_2 = \frac{\mathcal{F}}{m k_B T \gamma}\left(1 + \frac{p_v^2}{m^2 c^2}\right)^{-\frac{1}{2}} \circ dp_v + \frac{\mathcal{F} p_v}{m k_B T}\left(1 + \frac{p_v^2}{m^2 c^2}\right)^{-\frac{1}{2}} dt \tag{6.52}$$

$$k_B^{-1} d\Delta\mathcal{S}_3 = -\frac{2\mathcal{F}}{m k_B T \gamma}\left(1 + \frac{p_v^2}{m^2 c^2}\right)^{-\frac{1}{2}} \circ dp_v. \tag{6.53}$$

Despite knowledge of the stationary distribution in such a system, finding an analytical form for the transient, relaxing, probability density function is more challenging with no obvious solution. We can however, numerically investigate the contributions which do not rely on this information, namely the contributions to the house-keeping heat.

We may immediately find the expression for $d\langle\Delta\mathcal{S}_2\rangle^{\mathrm{F}}_{p_0^{\mathrm{F}}}/dt$ by our usual methods, noticing that as before we have a state independent ratio of irreversible current to probability density function,

$$\frac{J^{\mathrm{ir}}_{p_v}(p_v, x, t)}{p^{\mathrm{F}}(p_v, x, t)} = -\mathcal{F}, \tag{6.54}$$

but a state dependent mean entropy contribution due to the multiplicative noise

$$\begin{aligned}\frac{1}{k_B}\frac{d\langle\Delta\mathcal{S}_2\rangle^{\mathrm{F}}_{p_0^{\mathrm{F}}}}{dt} &= \int dp_v\, p^{\mathrm{F}}(p_v, x, t) \frac{\mathcal{F}^2}{k_B T m \gamma}\left(1 + \frac{p_v^2}{m^2 c^2}\right)^{-\frac{1}{2}} \\ &= \left\langle \frac{\mathcal{F}^2}{k_B T m \gamma}\left(1 + \frac{p_v^2}{m^2 c^2}\right)^{-\frac{1}{2}} \right\rangle.\end{aligned} \tag{6.55}$$

This is the first example we have seen where a phase space independent non-equilibrium constraint (the non-conservative force) has led to a distribution dependent contribution to the generalised house-keeping heat. This is a reflection of how entropy production is intimately related to the irreversibility manifest in the dynamics. That is to say, whilst the constant force generates a constant ratio of irreversible current to probability density function (in the stationary state) which we know is the cause of entropy production, the irreversibility and thus the actual magnitude of the contribution, will be inversely dependent on the noise strength. Consequently we observe that a given irreversible current results in larger contributions to $\Delta\mathcal{S}_2$ the closer to the origin in momentum space it is. This is of the form of the Lorentz contraction correction term and motivates a measure of the mean, conditional, phase space dependent, contribution to the entropy production rate. This is simply the positive quantity that is being 'averaged' in the expressions for the mean contributions and allows us to identify the thermodynamically important and unimportant parts of phase space for the contribution $\Delta\mathcal{S}_2$. For our example of a relativistic Brownian particle subject to a constant non-conservative force this is given in Fig. 6.11.

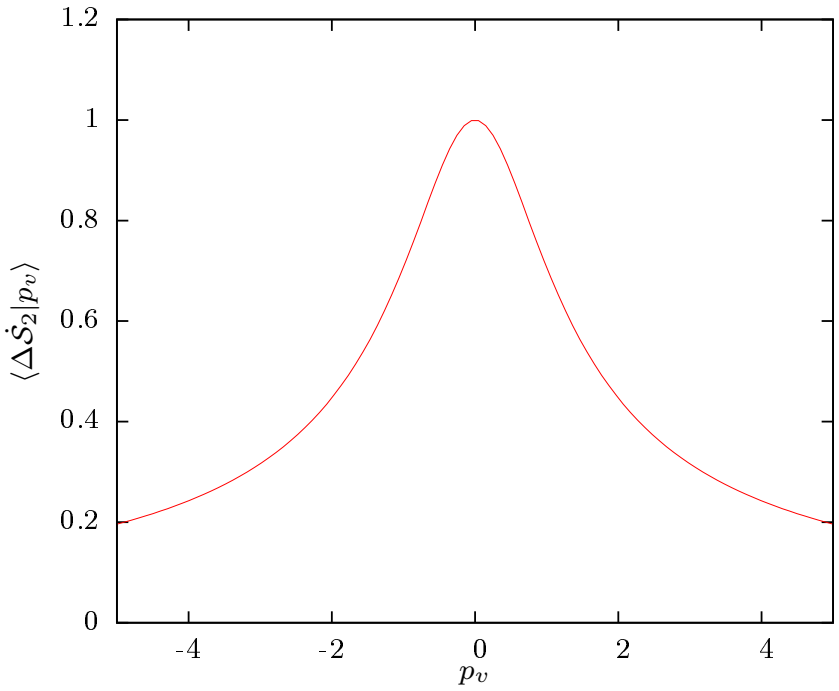

Figure 6.11: Phase space conditional mean contribution to the generalised house-keeping heat for a relativistic Langevin particle using units $k_B T = m = c = \gamma = \mathcal{F} = 1$.

Noticing that the function is symmetric about the origin, we may gain understanding of the mean total contribution by imagining the expected form of the transient probability density function placed on such a distribution: for a force reversal from the stationary state, the momentum distribution will shift from one side of the mean momentum dependent production rate to the other. As such it will move through the higher contributing region at the origin meaning a force reversal for relativistic Brownian motion is characterised by a transient increase in the expected generalised house-keeping heat before returning to its original value by symmetry of the initial and final distributions. This transient behaviour in the generalised house-keeping heat (and its analogue where there are only even variables) means that its place within an operational non-equilibrium thermodynamics cannot be simply expressed, as is often the case in the literature, as the constant heat flow to maintain the steady state, but is rather better expressed as the positive contribution to irreversibility arising from a non-equilibrium constraint. This transient nature is illustrated for this example from a numerical approach in Fig. 6.12.

A similar system which exhibits this behaviour is that of non-linear Brownian motion where

$$dv = -\gamma(|v|)v dt + \frac{\mathcal{F}}{m} dt + \sqrt{2k_B T \gamma(|v|)} m \bullet dW. \tag{6.56}$$

Once again assuming a Hänggi-Klimontovich interpretation, any even function $\gamma(|v|)$ will give rise to the entropy contributions

$$k_B^{-1} d\Delta \mathcal{S}_{\text{tot}} = -d\left(\ln p^{\text{F}}(x, v, t)\right) - d\left(\frac{mv^2}{2k_B T}\right) + \frac{\mathcal{F}}{k_B T} dx \tag{6.57}$$

$$k_B^{-1} d\Delta \mathcal{S}_1 = -d\left(\ln p^{\text{F}}(x, v, t)\right) - d\left(\frac{mv^2}{2k_B T}\right) + \frac{\mathcal{F}}{k_B T \gamma(v)} \circ dv \tag{6.58}$$

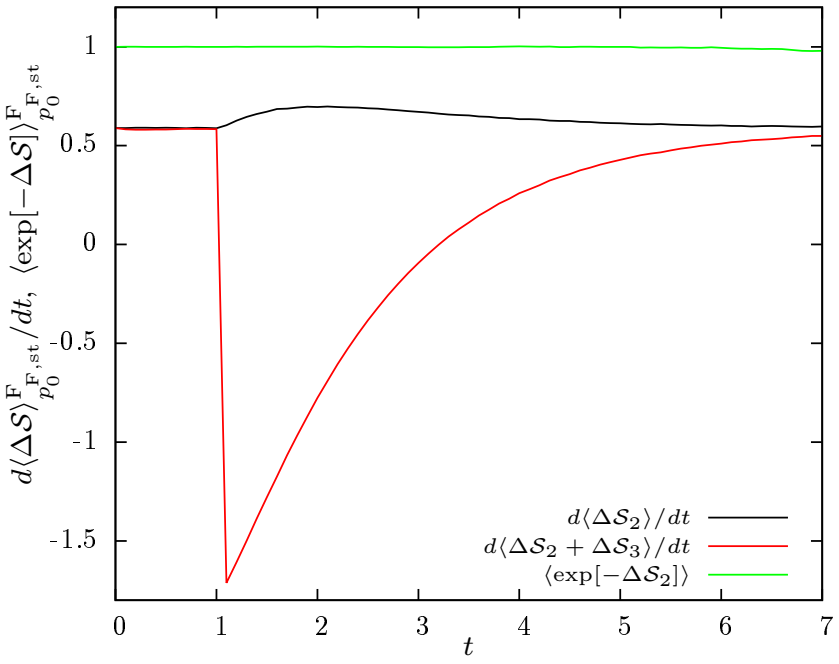

Figure 6.12: Mean contributions to the generalised and total house-keeping heats and the adherence of the former to an IFT for a force reversal from the stationary state for a relativistic Langevin particle for $k_BT = \gamma = m = c = \mathcal{F} = 1$.

$$k_B^{-1} d\Delta\mathcal{S}_2 = \frac{\mathcal{F}}{k_BT\gamma(v)} \circ dv + \frac{\mathcal{F}}{k_BT} dx \tag{6.59}$$

$$k_B^{-1} d\Delta\mathcal{S}_3 = -\frac{2\mathcal{F}}{k_BT\gamma(v)} \circ dv. \tag{6.60}$$

and thus a mean contribution

$$\frac{1}{k_B} \frac{d\langle\Delta\mathcal{S}_2\rangle^{\mathrm{F}}_{p_0^{\mathrm{F}}}}{dt} = \left\langle \frac{\mathcal{F}^2}{k_BTm\gamma(v)} \right\rangle. \tag{6.61}$$

Finally, similar behaviour can be seen in systems with state dependent non-equilibrium constraints such as, for example,

$$dv = -\gamma v dt + \frac{\mathcal{F}(v)}{m} dt + \sqrt{2k_BT\gamma} m dW, \tag{6.62}$$

which would lead to

$$\frac{1}{k_B} \frac{d\langle\Delta\mathcal{S}_2\rangle^{\mathrm{F}}_{p_0^{\mathrm{F}}}}{dt} = \left\langle \frac{\mathcal{F}^{\mathrm{rev}}(v)^2}{k_BTm\gamma} \right\rangle, \tag{6.63}$$

where $\mathcal{F}^{\mathrm{rev}}(v)$ is the reversible part of the velocity dependent applied force. One might think, given all the examples considered so far, that the parity dependent form of $d\langle\Delta\mathcal{S}_2\rangle^{\mathrm{F}}_{p_0^{\mathrm{F}}}/dt$ given in Eq. (5.66) is superfluous since every expression for the conditional mean production rate has been explicitly symmetric. This has arisen since we have defined natural boundaries on v with a non-conservative force (periodic boundaries in odd variables would surely be unphysical) giving $J_v(\boldsymbol{x}) = 0$ automatically rendering $J_v^{\mathrm{ir,st}}(\boldsymbol{x})/p^{\mathrm{st}}(\boldsymbol{x})$ symmetric by virtue of the necessary symmetry in $J_v^{\mathrm{rev,st}}(\boldsymbol{x})/p^{\mathrm{st}}(\boldsymbol{x}) = A^{\mathrm{rev}}(\boldsymbol{x})$. However, we point out that situations can arise where the current in, for example, v is coupled to other system variables allowing a non-zero total current and thus an asymmetric $J_v^{\mathrm{ir,st}}(\boldsymbol{x})/p^{\mathrm{st}}(\boldsymbol{x})$ by exploiting the higher dimensionality to produce rotation. Such an example is of heat transport in one dimension which we consider next which we note also serves as a justification for the inclusion of the odd velocity

variable in the dynamics as one cannot provide a satisfactory description without it.

6.4 Thermal Transport

Here we present an example of a situation which necessitates the use of odd variables in order to describe entropy production adequately by considering heat transport due to diffusion in one spatial dimension in the presence of a spatially dependent temperature field. In particular we contrast this approach to that which considers a bound particle simultaneously in contact with two reservoirs with different temperatures [98, 115, 161, 162]. Mathematically this system can be modelled without odd (velocity) variables by employing the over-damped limit and constructing a multiplicative SDE and Fokker-Planck equation of the form

$$dx = \frac{\mathcal{F}(x)}{m\gamma}dt + \sqrt{\frac{2k_BT(x)}{m\gamma}}dW \tag{6.64}$$

and

$$\frac{\partial p^{\mathrm{F}}(x,t)}{\partial t} = -\frac{\partial}{\partial x}\left(\frac{\mathcal{F}(x)p^{\mathrm{F}}(x,t)}{m\gamma}\right) + \frac{\partial^2}{\partial x^2}\left(\frac{k_BT(x)p^{\mathrm{F}}(x,t)}{m\gamma}\right) \tag{6.65}$$

where again, m is the particle mass, γ the damping coefficient and $\mathcal{F}(x)$ the force operating on the particle which for simplicity we state arises from a defined potential. We note the Itō form of both (for a discussion of the resolution of the Itō-Stratonovich dilemma in this case see, for example, [163–165]). This Fokker-Planck equation has a stationary distribution

$$p^{\mathrm{F,st}}(x) = \frac{\mathcal{N}m}{k_BT(x)}\exp\left[\int_0^x dx'\ \frac{\mathcal{F}(x')}{k_BT(x')}\right] \tag{6.66}$$

where $\mathcal{N}$ is a normalisation constant. We can quite readily identify the terms $A_x^{\mathrm{ir}} = \mathcal{F}(x)/m\gamma$, $A_x^{\mathrm{rev}} = 0$ and $D_x(x) = k_BT(x)/m\gamma$. However, when we come to construct the entropy production in the stationary state from Eq. (5.27) as

$$\begin{aligned} k_B^{-1}d\Delta\mathcal{S}_{\mathrm{tot}} =& \frac{A_x^{\mathrm{ir}}(x)}{D_x(x)} \circ dx - \frac{1}{D_x(x)}\frac{\partial D_x(x)}{\partial x} \circ dx - \frac{1}{p^{\mathrm{F,st}}(x)}\frac{\partial p^{\mathrm{F,st}}(x)}{\partial x} \circ dx \\ =& \left[\frac{\mathcal{F}(x)}{k_BT(x)} - \frac{1}{T(x)}\frac{\partial T(x)}{\partial x} - \frac{1}{p^{\mathrm{F,st}}(x)}\left(-\frac{1}{T(x)}\frac{\partial T(x)}{\partial x}p^{\mathrm{F,st}}(x) + \frac{\mathcal{F}(x)}{k_BT(x)}p^{\mathrm{F,st}}(x)\right)\right] \circ dx \\ =& 0 \end{aligned} \tag{6.67}$$

we find that there is zero entropy production for all trajectories. This may be understood either physically by recognising that in the over-damped limit one demands that the velocity distribution relaxes instantaneously thereby preventing any heat transfer due to temperature inhomogeneities, or geometrically by recognising the impossibility of having stationary flow, and thus entropy production, for a system in one dimension with natural boundaries.

To provide a satisfactory representation and to understand the entropy production in such a system we need to consider the more realistic under-damped dynamics in full phase space where we retain both position and velocity coordinates, x and v, which are even and odd under time reversal, respectively.

The SDEs and Fokker-Planck equation are now given as

$$\begin{aligned} dx &= vdt \\ dv &= -\gamma v dt + \frac{\mathcal{F}(x)}{m}dt + \sqrt{\frac{2k_BT(x)\gamma}{m}}dW \end{aligned} \tag{6.68}$$

and

$$\frac{\partial p^{\mathrm{F}}(x,v,t)}{\partial t} = -v\frac{\partial p^{\mathrm{F}}(x,v,t)}{\partial x} - \frac{\partial}{\partial v}\left(\left(\frac{\mathcal{F}(x)}{m} - \gamma v\right)p^{\mathrm{F}}(x,v,t)\right) + \frac{k_BT(x)\gamma}{m}\frac{\partial^2 p^{\mathrm{F}}(x,v,t)}{\partial v^2}. \tag{6.69}$$

We may, as before, then identify the terms $A^{\mathrm{ir}}_x = 0$, $A^{\mathrm{rev}}_x = v$, $A^{\mathrm{ir}}_v = -\gamma v$, $A^{\mathrm{rev}}_v = \mathcal{F}(x)/m$, $D_x = 0$ and $D_v = k_BT(x)\gamma/m$. By Eq. (5.27) the entropy production is

$$\begin{aligned} k_B^{-1}d\Delta\mathcal{S}_{\mathrm{tot}} &= -d(\ln p^{\mathrm{F}}(x,v,t)) - \frac{mv}{k_BT(x)} \circ dv + \frac{\mathcal{F}v}{k_BT(x)}dt \\ &= -d(\ln p^{\mathrm{F}}(x,v,t)) - \frac{1}{k_BT(x)}d\left(\frac{mv^2}{2}\right) + \frac{\mathcal{F}}{k_BT(x)}dx \end{aligned} \tag{6.70}$$

using $v \circ dv = (1/2)(v' + v)(v' - v)$ and $vdt = dx$, and noting that x is now deterministic, meaning the integration rules are irrelevant. The second and third terms correctly reproduce the form of the change in medium entropy as heat transfer to the environment, equal to negative heat transfer to the particle (in agreement with the result found in stochastic energetics [45]), divided by the instantaneous temperature, and do so only by virtue of the consideration of odd and even variables.

We can use this SDE to produce distributions of entropy production and verify relevant fluctuation theorems. To do so, however, requires knowledge of the solution to the Fokker-Planck equation, for which there is no simple analytical form. To proceed we restrict ourselves to the stationary state and utilise the expansion found in [165] and [166] which expresses the stationary solution as a series expansion about the over-damped distribution:

$$p^{\mathrm{F,st,over}}(x,v) = \frac{\mathcal{N}m}{k_BT(x)}\exp\left[\int_0^x dx'\,\frac{\mathcal{F}(x')}{k_BT(x')}\right]\sqrt{\frac{m}{2\pi k_BT(x)}}\exp\left[-\frac{mv^2}{2k_BT(x)}\right], \tag{6.71}$$

where $\mathcal{N}$ is determined by normalisation, such that

$$\begin{aligned} p^{\mathrm{F,st}}(x,v) &= p^{\mathrm{F,st,over}}(x,v) + \sum_{i=1}^{\infty}(1/\gamma)^i p_i(x,v) \\ &= \sum_{i=0}^{\infty}(1/\gamma)^i p_i(x,v). \end{aligned} \tag{6.72}$$

$p_i(x,v)$ has a general form

$$p_i(x,v) = \sum_{k=a_i}^{k=b_i}\frac{c_{i,k}(x)H_k(v\sqrt{m/k_BT(x)})}{\sqrt{2\pi k_BT(x)/m}}\exp\left[-\frac{mv^2}{2k_BT(x)}\right], \tag{6.73}$$

where constants a_i, b_i and functions $c_{i,k}(x)$ are found by an iterative procedure, and $H_k(y)$ are Hermite polynomials defined as

$$H_k(y) = (-1)^k e^{\frac{y^2}{2}}\frac{d^k}{dy^k}e^{\frac{-y^2}{2}}. \tag{6.74}$$

Defining the thermal velocity,

$$v_{\mathrm{th}}(x) = \sqrt{\frac{k_B T(x)}{m}}, \tag{6.75}$$

the terms not automatically zero and other requisite expressions are described below. For zeroth order we have

$$c_{0,0}(x) = p_0(x) = \frac{\mathcal{N} m}{k_B T(x)} \exp\left[\int_0^x dx' \, \frac{\mathcal{F}(x')}{k_B T(x')}\right] \tag{6.76}$$

where $p_0(x)$ is the marginalised over-damped solution such that $p^{\mathrm{F,st,over}}(x,v) = p_0(x,v)$. First order terms are:

$$\begin{aligned}
c'_{0,0}(x) &= \frac{\mathcal{F}(x) - k_B T'(x)}{k_B T(x)} c_{0,0}(x) \\
c_{1,1}(x) &= \frac{\mathcal{F}(x)}{m v_{\mathrm{th}}(x)} c_{0,0}(x) - c'_{0,0}(x) v_{\mathrm{th}}(x) - 2 c_{0,0}(x) v'_{\mathrm{th}}(x) = 0 \\
c_{1,3}(x) &= -\frac{p_0(x)}{3} v'_{\mathrm{th}}(x).
\end{aligned} \tag{6.77}$$

Second order terms are given as:

$$\begin{aligned}
c_{2,2}(x) &= -\frac{3}{2}\left(c'_{1,3}(x) v_{\mathrm{th}}(x) + 3 c_{1,3}(x) v'_{\mathrm{th}}(x)\right) \\
&= p_0(x)\left(\frac{(v'_{\mathrm{th}}(x))^2}{2} + \frac{\mathcal{F} v'_{\mathrm{th}}(x)}{2 m v_{\mathrm{th}}(x)} + \frac{v_{\mathrm{th}}(x) v''_{\mathrm{th}}(x)}{2}\right) \\
c_{2,4}(x) &= -\frac{c'_{1,3}(x) v_{\mathrm{th}}(x)}{4} - 2 c_{1,3}(x) v'_{\mathrm{th}}(x) + \frac{\mathcal{F}(x)}{4 m v_{\mathrm{th}}(x)} c_{1,3}(x) \\
&= p_0(x)\left(\frac{(v'_{\mathrm{th}}(x))^2}{2} + \frac{v_{\mathrm{th}}(x) v''_{\mathrm{th}}(x)}{12}\right) \\
c_{2,6}(x) &= -\frac{1}{6} c_{1,3}(x) v'_{\mathrm{th}}(x) \\
&= p_0(x) \frac{(v'_{\mathrm{th}}(x))^2}{18} \\
c_{2,0}(x) &= p_0(x)\left[-\int_0^x dx' \, \frac{1}{p_0(x')}\left[2 c'_{2,2}(x') + 4 c_{2,2}(x') \frac{v'_{\mathrm{th}}(x')}{v_{\mathrm{th}}(x')}\right]\right. \\
&\quad \left. + \int_{-\infty}^{\infty} dx' \, p_0(x') \left[\int_0^{x'} dx'' \, \frac{1}{p_0(x'')}\left[2 c'_{2,2}(x'') + 4 c_{2,2}(x'') \frac{v'_{\mathrm{th}}(x'')}{v_{\mathrm{th}}(x'')}\right]\right]\right] \\
&= 2 p_0(x)\left[-\left(\frac{(v'_{\mathrm{th}}(x))^2}{2} + \frac{\mathcal{F}(x) v'_{\mathrm{th}}(x)}{2 m v_{\mathrm{th}}(x)} + \frac{v_{\mathrm{th}}(x) v''_{\mathrm{th}}(x)}{2}\right)\right. \\
&\quad - \int_0^x dx' \frac{\mathcal{F}(x')}{m v_{\mathrm{th}}^2(x')}\left(\frac{(v'_{\mathrm{th}}(x'))^2}{2} + \frac{\mathcal{F}(x') v'_{\mathrm{th}}(x')}{2 m v_{\mathrm{th}}(x')} + \frac{v_{\mathrm{th}}(x') v''_{\mathrm{th}}(x')}{2}\right) \\
&\quad \left. + \int_{-\infty}^{+\infty} dx' p_0(x') (v'_{\mathrm{th}}(x'))^2 + \int_{-\infty}^{+\infty} dx' p_0(x') \int_0^{x'} dx'' \frac{\mathcal{F}(x'')}{m v_{\mathrm{th}}^2(x'')}\left(\frac{(v'_{\mathrm{th}}(x''))^2}{2} + \frac{\mathcal{F}(x'') v'_{\mathrm{th}}(x'')}{2 m v_{\mathrm{th}}(x'')}\right)\right].
\end{aligned} \tag{6.78}$$

Third order terms are:

$$\begin{aligned}
c'_{2,2}(x) &= c_{2,2}(x)\left(\frac{\mathcal{F}(x)}{m v_{\mathrm{th}}^2(x)} - \frac{2 v'_{\mathrm{th}}(x)}{v_{\mathrm{th}}(x)}\right) \\
&\quad + p_0(x)\left(\frac{\mathcal{F}(x) v''_{\mathrm{th}}(x)}{2 m v_{\mathrm{th}}(x)} + \frac{\mathcal{F}'(x) v'_{\mathrm{th}}(x)}{2 m v_{\mathrm{th}}(x)} - \frac{\mathcal{F}(x)(v'_{\mathrm{th}}(x))^2}{2 m v_{\mathrm{th}}^2(x)} + 3\frac{v'_{\mathrm{th}}(x) v''_{\mathrm{th}}(x)}{2} + \frac{v_{\mathrm{th}}(x) v'''_{\mathrm{th}}(x)}{2}\right)
\end{aligned}$$

$$
\begin{aligned}
c'_{2,4}(x) &= c_{2,4}(x)\left(\frac{\mathcal{F}(x)}{mv_{\rm th}^2(x)} - \frac{2v'_{\rm th}(x)}{v_{\rm th}(x)}\right) + p_0(x)\left(\frac{13}{12}v'_{\rm th}(x)v''_{\rm th}(x) + \frac{1}{12}v_{\rm th}(x)v'''_{\rm th}(x)\right)\\
c'_{2,6}(x) &= c_{2,4}(x)\left(\frac{\mathcal{F}(x)}{mv_{\rm th}^2(x)} - \frac{2v'_{\rm th}(x)}{v_{\rm th}(x)}\right) + p_0(x)\frac{v'_{\rm th}(x)v''_{\rm th}(x)}{9}\\
c_{3,3}(x) &= -\frac{4}{3}\left(c'_{2,4}(x)v_{\rm th}(x) + 4c_{2,4}(x)v'_{\rm th}(x)\right)\\
&\quad - \frac{1}{3}\left(c'_{2,2}(x)v_{\rm th}(x) + 6c_{2,2}(x)v'_{\rm th}(x) - \frac{\mathcal{F}(x)}{mv_{\rm th}(x)}c_{2,2}(x)\right) - \frac{1}{3}c_{2,0}(x)v'_{\rm th}(x)\\
c_{3,5}(x) &= -\frac{6}{5}\left(c'_{2,6}(x)v_{\rm th}(x) + 6c_{2,6}(x)v'_{\rm th}(x)\right)\\
&\quad - \frac{1}{5}\left(c'_{2,4}(x)v_{\rm th}(x) + 10c_{2,4}(x)v'_{\rm th}(x) - \frac{\mathcal{F}(x)}{mv_{\rm th}(x)}c_{2,4}(x)\right) - \frac{1}{5}c_{2,2}(x)v'_{\rm th}(x)\\
c_{3,7}(x) &= -\frac{1}{7}\left(c'_{2,6}(x)v_{\rm th}(x) + 14c_{2,6}(x)v'_{\rm th}(x) - \frac{\mathcal{F}(x)}{mv_{\rm th}(x)}c_{2,6}(x)\right) - \frac{1}{7}c_{2,4}(x)v'_{\rm th}(x)\\
c_{3,9}(x) &= -\frac{1}{9}c_{2,6}(x)v'_{\rm th}(x),
\end{aligned}
\tag{6.79}
$$

and fourth order:

$$
\begin{aligned}
c'_{3,5}(x) &= -\frac{6}{5}\left(7c'_{2,6}(x)v'_{\rm th}(x) + c''_{2,6}(x)v_{\rm th}(x) + 6c_{2,6}(x)v''_{\rm th}(x)\right)\\
&\quad - \frac{1}{5}\left(c''_{2,6}(x)v_{\rm th}(x) + 11c'_{2,6}(x)v'_{\rm th}(x) + 10c_{2,6}(x)v''_{\rm th}(x)\right.\\
&\quad \left. -\frac{\mathcal{F}(x)}{mv_{\rm th}(x)}c'_{2,4}(x) - \left(\frac{\mathcal{F}'(x)}{mv_{\rm th}(x)} - \frac{\mathcal{F}(x)v'_{\rm th}(x)}{mv_{\rm th}^2(x)}\right)c_{2,4}(x)\right) - \frac{1}{5}\left(c_{2,2}(x)v''_{\rm th}(x) + c'_{2,2}(x)v'_{\rm th}(x)\right)\\
c'_{3,7}(x) &= -\frac{1}{7}\left(c''_{2,6}(x)v_{\rm th}(x) + 15c'_{2,6}(x)v'_{\rm th}(x) + 14c_{2,6}(x)v''_{\rm th}(x)\right.\\
&\quad \left. -\frac{\mathcal{F}(x)}{mv_{\rm th}(x)}c'_{2,6}(x) - \left(\frac{\mathcal{F}'(x)}{mv_{\rm th}(x)} - \frac{\mathcal{F}(x)v'_{\rm th}(x)}{mv_{\rm th}^2(x)}\right)c_{2,6}(x)\right) - \frac{1}{7}\left(c_{2,4}(x)v''_{\rm th}(x) + c'_{2,4}(x)v'_{\rm th}(x)\right)\\
c'_{3,9}(x) &= -\frac{1}{9}\left(c'_{2,6}(x)v'_{\rm th}(x) + c_{2,6}(x)v''_{\rm th}(x)\right)\\
c_{4,2}(x) &= -\frac{3}{2}\left(c'_{3,3}(x)v_{\rm th}(x) + 3c_{3,3}(x)v'_{\rm th}(x)\right)\\
c_{4,4}(x) &= -\frac{5}{4}\left(c'_{3,5}(x)v_{\rm th}(x) + 5c_{3,5}(x)v'_{\rm th}(x)\right) - \frac{1}{4}\left(c'_{3,5}(x)v_{\rm th}(x) + 8c_{3,3}(x)v'_{\rm th}(x) - \frac{\mathcal{F}(x)}{mv_{\rm th}(x)}c_{3,3}(x)\right)\\
c_{4,6}(x) &= -\frac{7}{6}\left(c'_{3,7}(x)v_{\rm th}(x) + 7c_{3,7}(x)v'_{\rm th}(x)\right)\\
&\quad - \frac{1}{6}\left(c'_{3,5}(x)v_{\rm th}(x) + 12c_{3,5}(x)v'_{\rm th}(x) - \frac{\mathcal{F}(x)}{mv_{\rm th}(x)}c_{3,5}(x)\right) - \frac{1}{6}c_{3,3}(x)v'_{\rm th}(x)\\
c_{4,8}(x) &= -\frac{9}{8}\left(c'_{3,9}(x)v_{\rm th}(x) + 9c_{3,9}(x)v'_{\rm th}(x)\right)\\
&\quad - \frac{1}{8}\left(c'_{3,7}(x)v_{\rm th}(x) + 16c_{3,7}(x)v'_{\rm th}(x) - \frac{\mathcal{F}(x)}{mv_{\rm th}(x)}c_{3,7}(x)\right) - \frac{1}{8}c_{3,5}(x)v'_{\rm th}(x)\\
c_{4,10}(x) &= -\frac{1}{10}\left(c'_{3,9}(x)v_{\rm th}(x) + 20c_{3,9}(x)v'_{\rm th}(x) - \frac{\mathcal{F}(x)}{mv_{\rm th}(x)}c_{3,9}(x)\right) - \frac{1}{10}c_{3,7}(x)v'_{\rm th}(x)\\
c_{4,12}(x) &= -\frac{1}{12}c_{3,9}(x)v'_{\rm th}(x)\\
c_{4,0}(x) &= p_0(x)\left[-\int_0^x dx'\,\frac{1}{p_0(x')}\left[2c'_{4,2}(x') + 4c_{4,2}(x')\frac{v'_{\rm th}(x)}{v_{\rm th}(x)}\right]\right.\\
&\quad \left. + \int_{-\infty}^{\infty} dx'\,p_0(x')\left[\int_0^{x'} dx''\,\frac{1}{p_0(x'')}\left[2c'_{4,2}(x'') + 4c_{4,2}(x'')\frac{v'_{\rm th}(x'')}{v_{\rm th}(x'')}\right]\right]\right].
\end{aligned}
\tag{6.80}
$$

Whilst the expansion has the formal deficiency that the expansion parameter is not unit-less it suffices

for a theoretical illustration where we can consider it in a limit where it is appropriate. The form of the correction terms is illustrated for an example system in Fig. 6.13.

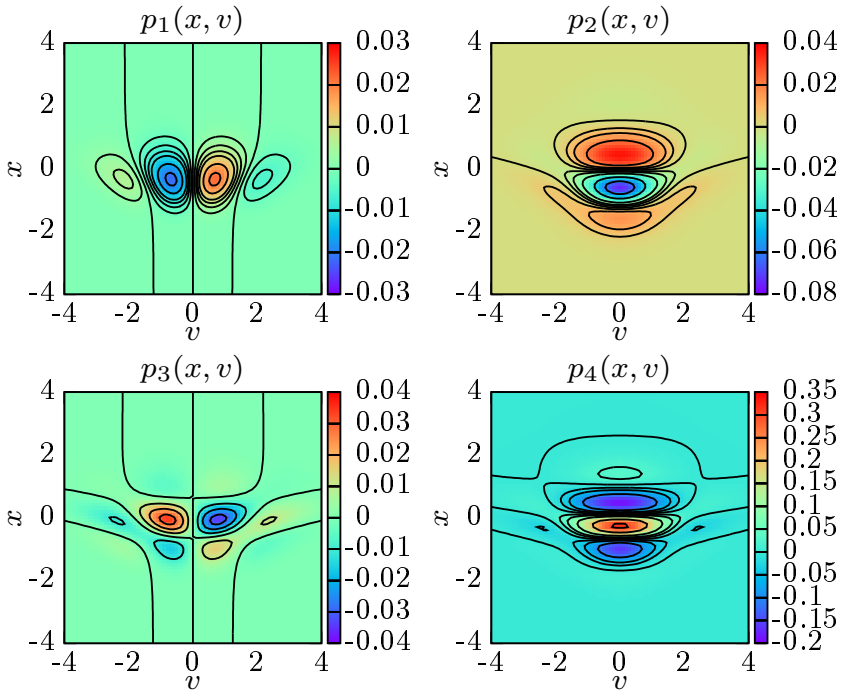

Figure 6.13: Contour maps of the expansion terms $p_1(x,v)$, $p_2(x,v)$, $p_3(x,v)$ and $p_4(x,v)$ for a harmonic trap $\mathcal{F}(x) = -x$, temperature profile $T(x) = 1 + \frac{1}{2}\tanh x$ and $m = k_B = 1$.

As an illustration of the structure of the mean conditional contribution to $\Delta\mathcal{S}_2$ and how it reduces to the total entropy production we calculate it from the appropriate integral up to second order in γ^{-1} where we first observe asymmetric behaviour and yet is analytically tractable. This is possible since the integral form of the $c_{2,0}(x)$ term does not contribute to $J_v^{\mathrm{ir,st}}(\boldsymbol{x})/p^{\mathrm{F,st}}(\boldsymbol{x})$ since the zeroth order Hermite polynomial introduces no further v dependence on top of the Gaussian form. If we write

$$\begin{aligned}\frac{1}{k_B}\frac{d\langle\Delta\mathcal{S}_2\rangle^{\mathrm{F}}_{p_0^{\mathrm{F}}}}{dt} &= k_B^{-1}\langle\langle\Delta\dot{\mathcal{S}}_2|x,v\rangle\rangle \\ &= \left\langle \gamma^{-1}\langle\Delta\dot{\mathcal{S}}_2^1|x,v\rangle + \gamma^{-2}\langle\Delta\dot{\mathcal{S}}_2^2|x,v\rangle + O(\gamma^{-3})\right\rangle, \end{aligned} \tag{6.81}$$

noting the absence of a zeroth order term owing to a lack of stationary state entropy production in the over-damped limit, we find

$$\langle\Delta\dot{\mathcal{S}}_2^1|x,v\rangle = \frac{((mv^2 - k_BT(x))T'(x))^2}{4mk_BT^3(x)} \tag{6.82}$$

and

$$\begin{aligned}&\langle\Delta\dot{\mathcal{S}}_2^2|x,v\rangle = \\ &\frac{vT'(x)(mv^2 - k_BT(x))\left(6\mathcal{F}(x)T(x)T'(x) - (mv^2 + 3k_BT(x))(T'(x))^2 + 2mv^2T(x)T''(x)\right)}{12mk_B(T(x))^4}.\end{aligned} \tag{6.83}$$

Of note, is that when equipartition in 1D holds, $\langle mv^2\rangle = k_BT(x)$, such as in equilibrium (or the over-damped limit), the contributions vanish. Importantly, because of the structure of the Hermite

polynomials, $\langle \Delta\dot{\mathcal{S}}_2^n|x,v\rangle$ is even in v for odd n and vice versa. This means that $\langle \Delta\dot{\mathcal{S}}_2|x,v\rangle$ is generally asymmetric and is something we didn't observe in the presence of a non-conservative force. However, in the *stationary* state we expect $d\langle \Delta\mathcal{S}_{\rm tot}\rangle^{\rm F,st}_{p_0^{\rm F,st}}/dt = d\langle \Delta\mathcal{S}_2\rangle^{\rm F,st}_{p_0^{\rm F,st}}/dt$ despite the parity difference in their expressions. This means we require the integral of the odd part of $\langle \Delta\dot{\mathcal{S}}_2|x,v\rangle$, with respect to v, to vanish in the stationary state. Since the expansion must hold for all γ, we then also require the integrals of all odd contributions for every γ^{-n} to vanish individually. Starting from the fact that all even nth order contributions in γ^{-1} to $\langle \Delta\dot{\mathcal{S}}_2|x,v\rangle$ are odd with respect to v we realise it is the integrals over these quantities that we require to vanish in the stationary state. Since we can represent the integral in the stationary state by integrating over such quantities multiplied by the expansion form of $p^{\rm F,st}(x,v)$, we then immediately identify that any contribution, even in γ^{-1}, must vanish in the stationary state. This is because they are necessarily formed of integrals, over v, of the form $p_i(x,v)\times\langle \Delta\dot{\mathcal{S}}_2^n|x,v\rangle$ where one of $p_i(x,v)$ or $\langle \Delta\dot{\mathcal{S}}_2^n|x,v\rangle$ is odd in v with the other even. An example of this is for second order in γ^{-1} given by the integral

$$\int dx \int dv\; p_0(x,v)\langle \Delta\dot{\mathcal{S}}_2^2|x,v\rangle + p_1(x,v)\langle \Delta\dot{\mathcal{S}}_2^1|x,v\rangle. \tag{6.84}$$

Both terms are odd in v and so disappear when integrated. The consequence of this is that in the stationary state there are no $\gamma^{-2},\gamma^{-4}\ldots$ contributions to the total entropy production. However, in the third order contribution to $d\langle \Delta\mathcal{S}_2\rangle^{\rm F,st}_{p_0^{\rm F,st}}/dt$ there is a contribution based on the term $\langle \Delta\dot{\mathcal{S}}_2^2|x,v\rangle$, which is odd with respect to v, of the form

$$\int dx \int dv\; p_1(x,v)\langle \Delta\dot{\mathcal{S}}_2^2|x,v\rangle. \tag{6.85}$$

Since $p_1(x,v)$ is also odd in v, this expression is even and so doesn't vanish directly from the form of the Hermite polynomials. We emphasise, however, that this integral must vanish. Such an expected property is based on the relationship between the stationary current and probability distribution, defined everywhere, and so we must instead consider the integral over all of the phase space rather than just the velocity. To give confidence in the result, after giving explicit forms for $\langle \Delta\dot{\mathcal{S}}_2^1|x,v\rangle$ and $\langle \Delta\dot{\mathcal{S}}_2^2|x,v\rangle$ for a given system, we demonstrate that this integral does indeed vanish, numerically for that specific system. Such a property then provides an illustration of the result

$$\begin{aligned}\frac{1}{k_B}\frac{d\langle \Delta\mathcal{S}_2\rangle^{\rm F,st}_{p_0^{\rm F,st}}}{dt} &= \int dx\int dv\;\frac{mp^{\rm F,st}(x,v)}{k_B\gamma T(x)}\left(\frac{J^{\rm ir,st}(x,-v)}{p^{\rm F,st}(x,-v)}\right)^2\\ &= \int dx\int dv\;\frac{mp^{\rm F,st}(x,v)}{k_B\gamma T(x)}\left(\frac{J^{\rm ir,st}(x,v)}{p^{\rm F,st}(x,v)}\right)^2\\ &= \frac{1}{k_B}\frac{d\langle \Delta\mathcal{S}_{\rm tot}\rangle^{\rm F,st}_{p_0^{\rm F,st}}}{dt},\end{aligned} \tag{6.86}$$

describing how the mean entropy production contribution associated with the generalised house-keeping heat is equivalent to the total entropy production in the stationary state despite the parity difference in their expressions. For the specific choices

$$T(x) = 1 + \frac{\tanh(x)}{2} \tag{6.87}$$

and

$$\mathcal{F}(x) = -x \tag{6.88}$$

we find

$$\langle \Delta \dot{\mathcal{S}}_2^1 | x, v \rangle = \frac{\mathrm{sech}^4(x)(2k_B - 2mv^2 + k_B \tanh(x))^2}{8mk_B(2 + \tanh(x))^3} \tag{6.89}$$

and

$$\langle \Delta \dot{\mathcal{S}}_2^2 | x, v \rangle = v \mathrm{sech}^6(x)(2k_B - 2mv^2 + k_B \tanh(x)) \times \frac{(6k_B + 2mv^2 + 3k_B \tanh(x) + 4(2\cosh(x) + \sinh(x))(3x\cosh(x) + 2mv^2 \sinh(x)))}{24mk_B(2 + \tanh(x))^4}. \tag{6.90}$$

These terms are illustrated in Fig. 6.14. Though somewhat unwieldy, by identifying the entropically important parts of phase space this illustrates a system where again the contribution $d\langle \Delta \mathcal{S}_2 \rangle_{p_0^{\mathrm{F}}}^{\mathrm{F}}/dt$ can behave transiently in the mean with relaxation of the distribution, but also has an explicit asymmetric dependence on odd variables necessitating the structure of the general expression for the mean contribution in Eq. (5.66). The integral found in Eq. (6.85), required to vanish in order to be consistent with the

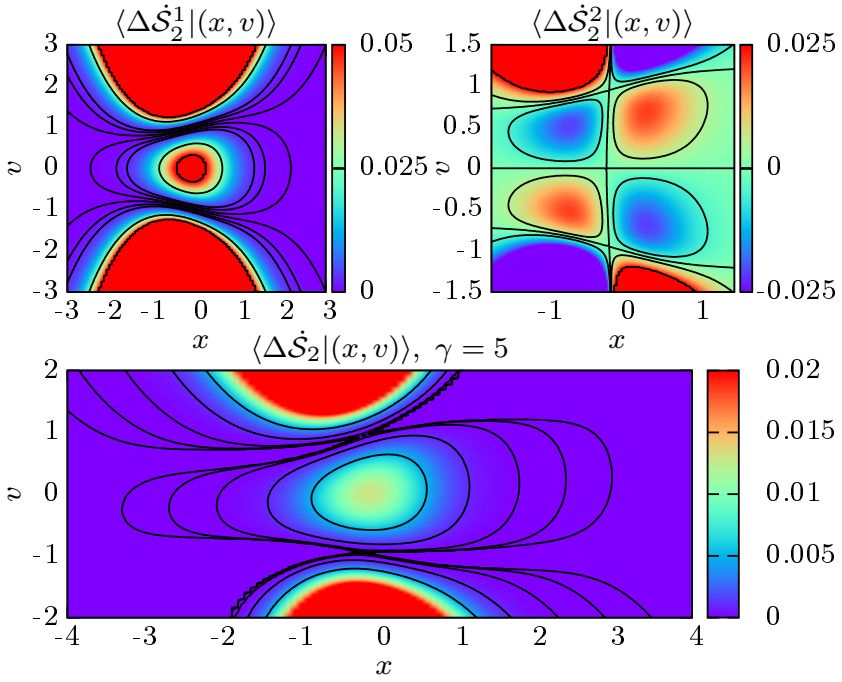

Figure 6.14: First, second order and combined contributions to the phase space conditional mean contribution to the entropy contribution $\Delta \mathcal{S}_2$ for a Langevin particle in harmonic trap $\mathcal{F} = -x$, temperature field $T(x) = 1 + \tanh(x)/2$, $m = k_B = 1$ and with $\gamma = 5$ for the combined contribution. In each plot the range is artificially bounded by $[0 : 0.05]$, $[-0.025 : 0.025]$ an $[0 : 0.02]$ respectively near the edges in order to elucidate the structure close to the origin.

three contributions to entropy production, is then illustrated in Fig. 6.15 by demonstrating the phase space dependence of the integrand and demonstrating the dependence of its value on the upper bound in position. This illustrates how the integral vanishes as we increase the bounds to cover all space. We note that the integral over all phase space, performed numerically, was found to be negligibly small at $\sim 6.0 \times 10^{-17}$.

As a further example of this asymmetry in the quantity $\langle \Delta \dot{\mathcal{S}}_2 | x, v \rangle$ we present the mean contributions to $\Delta \mathcal{S}_2$ and $\Delta \mathcal{S}_3$ for the system described here, calculated by means of a Monte Carlo average of the appropriate SDEs, relaxing from two different initial distributions which we allow to be the time reverse

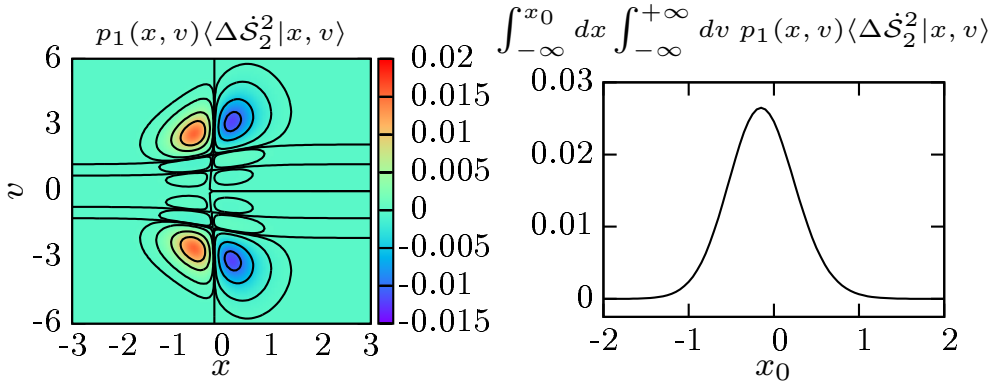

Figure 6.15: Phase space dependence of the third order contribution $p_1(x,v)\langle\Delta\dot{\mathcal{S}}_2^2|x,v\rangle$ and variation of the integral $\int_{-\infty}^{x_0} dx \int_{-\infty}^{+\infty} dv\ p_1(x,v)\langle\Delta\dot{\mathcal{S}}_2^2|x,v\rangle$ demonstrating that the integral vanishes in the stationary state enabling the result $d\langle\Delta\mathcal{S}_{\rm tot}\rangle^{\rm F,st}_{p_0^{\rm F,st}}/dt = d\langle\Delta\mathcal{S}_2\rangle^{\rm F,st}_{p_0^{\rm F,st}}/dt$.

of each other. We choose a simple Gaussian function designed to centre on the key asymmetry in the second order contribution to $\langle\Delta\dot{\mathcal{S}}_2|x,v\rangle$ in the positive x domain as illustrated in Fig. 6.14. As such we use parameters $\langle x(0)\rangle = 0.5$, $\langle v(0)\rangle \pm 0.5$ and $\sigma = 0.15$. The results for the relaxation period are shown in Fig. 6.16. We observe distinct behaviour in both $\Delta\mathcal{S}_2$ and $\Delta\mathcal{S}_3$ as the system relaxes depending on the

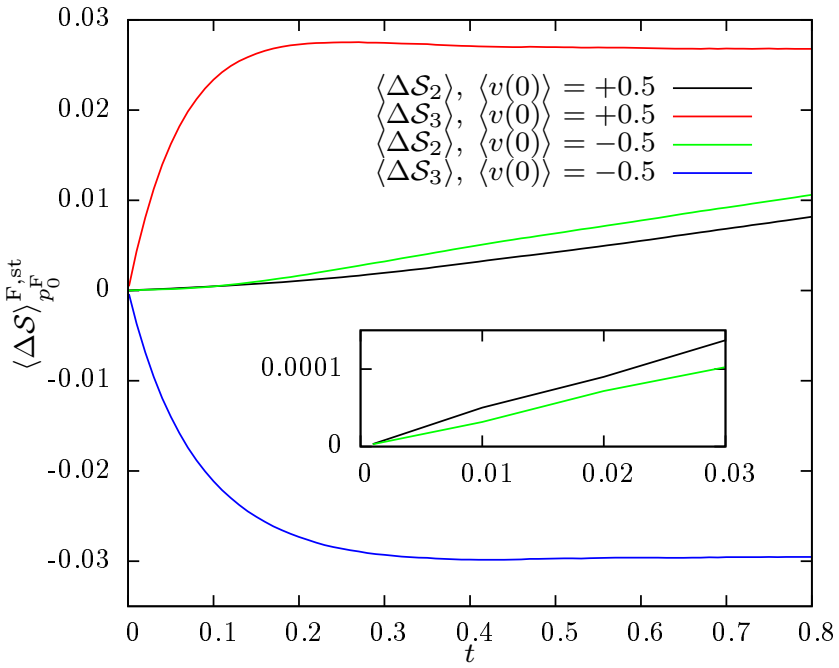

Figure 6.16: Mean contributions $\Delta\mathcal{S}_2$ and $\Delta\mathcal{S}_3$ for a Langevin particle in harmonic trap $\mathcal{F} = -x$, temperature field $T(x) = 1+\tanh(x)/2$, $m = 1$ and with $\gamma = 5$ subject to an initial starting distribution $p^{\rm F,st}(x,v,t=0) \propto \exp[-((x-\langle x(0)\rangle)^2+(v-\langle v(0)\rangle)^2)/2\sigma^2]$ for $\langle x(0)\rangle = 0.5$, $\langle v(0)\rangle = \pm 0.5$ and $\sigma = 0.15$.

initial distribution and note that for small times where the distribution is still approximately Gaussian according to the initial distribution the contributions reflect those expected both from Eq. (5.71) and Fig. 6.14. Specifically, a peaked potential will generally have a negative $\dot{p}(x,v,t)$ and, when centred on $v = 0.5$, $\ln(p^{\rm F,st}(x,v)/p^{\rm F,st}(x,-v))$ will be positive leading to a positive $d\langle\Delta\mathcal{S}_3\rangle^{\rm F}_{p_0^{\rm F}}/dt$ and negative for

$v = -0.5$. Similarly, when the distribution is centred at $v = 0.5$ the asymmetric second order contribution adds to $\langle \Delta\dot{\mathcal{S}}_2|x,v\rangle$ giving a larger $d\langle\Delta\mathcal{S}_2\rangle^{\mathrm{F}}_{p_0^{\mathrm{F}}}/dt$ than a distribution centred at $v = -0.5$ where the second order term reduces the contribution. This then provides numerical evidence for the expected transient asymmetric contributions to $d\langle\Delta\mathcal{S}_2\rangle^{\mathrm{F}}_{p_0^{\mathrm{F}}}/dt$.

Given the expression in Eq. (6.82) we may calculate a first order contribution to the mean, stationary, total entropy production, which is equivalent to the integral of $\gamma^{-1} \times p_0(x,v) \times \langle \Delta\dot{\mathcal{S}}_2^1|x,v\rangle$, and find

$$\frac{1}{k_B}\frac{d\langle\Delta\mathcal{S}_{\mathrm{tot}}\rangle^{\mathrm{F,st}}_{p_0^{\mathrm{F,st}}}}{dt} \simeq \int_{-\infty}^{+\infty} dx\ \frac{k_B p_0(x)}{2m\gamma T(x)}\left(\frac{\partial T(x)}{\partial x}\right)^2 \tag{6.91}$$

where

$$p_0(x) = \int_{-\infty}^{+\infty} dv\ p_0(x,v). \tag{6.92}$$

We emphasise that there is no second order contribution in the stationary state; this arises directly from the structure of $\langle \Delta\dot{\mathcal{S}}_2|x,v\rangle$ and is consistent with the result $d\langle\Delta\mathcal{S}_2\rangle^{\mathrm{F,st}}_{p_0^{\mathrm{F,st}}}/dt = d\langle\Delta\mathcal{S}_{\mathrm{tot}}\rangle^{\mathrm{F,st}}_{p_0^{\mathrm{F,st}}}/dt$. Third order contributions can be calculated, but are somewhat unilluminating with a complicated dependence on $\mathcal{F}(x)$, $c_{2,0}(x)$ and high order derivatives of the temperature field. Importantly, however, we find that the above expression is equivalent to the first order approximation found from the expected phenomenological expression for dimensionless internal entropy generation [167]

$$\frac{\dot{S}_{\mathrm{int}}}{k_B} = \int_{-\infty}^{+\infty} dx\ J_Q(x)\frac{\partial}{\partial x}\left(\frac{1}{k_B T(x)}\right) \tag{6.93}$$

where $J_Q(x)$ is the stationary heat current defined as

$$J_Q(x) = \int_{-\infty}^{+\infty} dv\ \frac{1}{2}mv^3 p^{\mathrm{F,st}}(x,v). \tag{6.94}$$

We now demonstrate numerically, an equivalence between these two approaches by calculating the mean total entropy production in the stationary state from a Monte Carlo averaging of the appropriate SDEs and an integral over the stationary heat current as predicted by the stationary solution found by means of the expansion up to fourth order in γ^{-1} for a range of γ, again for $\mathcal{F}(x) = -x$ and temperature profile

$$T(x) = 1 + \frac{1}{2}\tanh(x). \tag{6.95}$$

The results are shown in Fig. 6.17. The results show that the dimensionless entropy production obtained by performing the integral in Eq. (6.93) using a numerically calculated $p^{\mathrm{F,st}}(x,v)$ agrees well with that obtained by averaging the SDE in Eq. (6.70) by Monte Carlo simulation of the underlying particle dynamics. This provides both confidence in the definition of the path dependent entropy production and the accuracy in the expansion for the values of γ^{-1} being utilised. Specifically, this is because in the stationary state $\langle\Delta\mathcal{S}_{\mathrm{sys}}\rangle^{\mathrm{F,st}}_{p_0^{\mathrm{F,st}}} = 0$ regardless of the accuracy of $p^{\mathrm{F,st}}(x,v)$ meaning we have agreement between a microscopic approach found by averaging individual heat transfers, independently of the distribution, and a integral explicitly dependent on it. We also see both results converge to the first order result for large γ in the linear regime. At this point, we note that total entropy production decreases as coupling to the environment increases which may seem counter-intuitive, but we emphasise that with increased coupling, despite greater heat transfer to and from the environment, there is highly diminished spatial heat *transport* (the latter being the cause of entropy production) as the system is brought closer

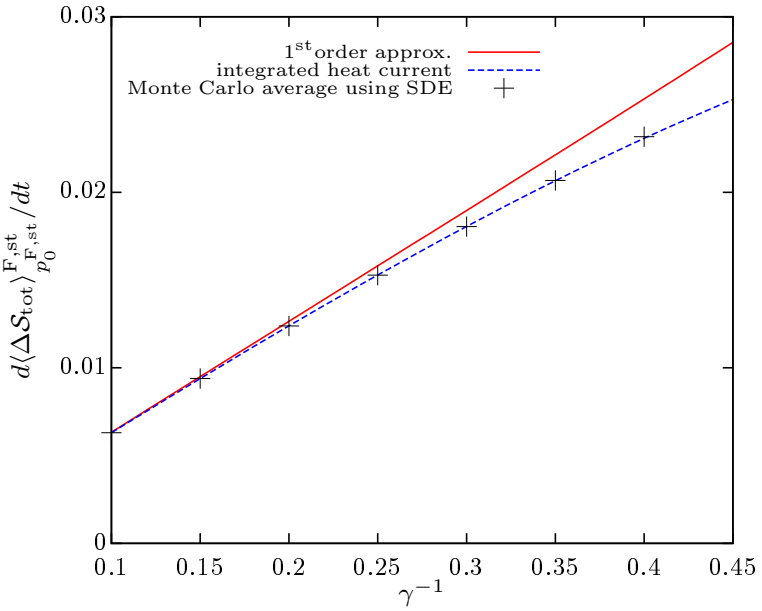

Figure 6.17: Mean dimensionless entropy production in the stationary state for a range of damping coefficients as predicted by a first order approximation in Eq. (6.91) (solid red line), an integral over the heat current, Eq. (6.93) (dashed blue line) and a Monte Carlo average based on the SDE in Eq. (6.70) (crosses). Simulations were performed by initialisation of particles into the stationary distribution using a simple reject/accept algorithm along with a burn in time of $t = 10$. We performed 3×10^7 Monte Carlo runs utilising a forward Euler discretisation method with time step $dt = 1.0 \times 10^{-3}$ to solve the SDE in Eq. (6.70).

to a local equilibrium.

Given an observed equivalence between the phenomenological entropy production and the mean path dependent entropy production we can use the SDE for entropy production (Eq. (6.70)) to move beyond a classical description of such mean productions to one described by Jarzynski, Seifert, Sekimoto and others [10, 33, 34] where we can identify entropy generating and destroying trajectories. We can explicitly calculate the distribution of total entropy production which is shown for $\gamma = 10$ in Fig. 6.18 for various process intervals, along with a demonstration that it adheres to an IFT throughout. Additionally, since we consider the stationary state we can demonstrate a detailed fluctuation theorem of the form $p^{\rm F,st}(\Delta\mathcal{S}_{\rm tot} = A)_{p_0^{\rm F,st}}/p^{\rm F,st}(\Delta\mathcal{S}_{\rm tot} = -A)_{p_0^{\rm F,st}} = \exp(A)$ [12] as shown in Fig. 6.19.

Finally we point out that, being in the stationary state, $d\langle\Delta\mathcal{S}_3\rangle^{\rm F,st}_{p_0^{\rm F,st}}/dt = 0$, but since it is a non-equilibrium stationary state that is asymmetric in the odd velocity variable we have $\Delta\mathcal{S}_3 \neq 0$ in detail, as is clear in Eq. (5.68). We can demonstrate the increasing range of values of $\Delta\mathcal{S}_3$ as γ is reduced and the system is taken further away from local equilibrium, with its symmetric velocity distribution, by generating the distribution of $\Delta\mathcal{S}_3$ using Eq. (5.68) for a given time interval, as shown in Fig. 6.20. Such a result highlights the fact that although a non-zero $d\langle\Delta\mathcal{S}_3\rangle^{\rm F}_{p_0^{\rm F}}/dt$ is only possible during relaxation as shown by Eq. (5.71), the specific evolution of $\Delta\mathcal{S}_3$ for each trajectory is brought about by non-equilibrium constraints that cause the stationary solution to depart from equilibrium.

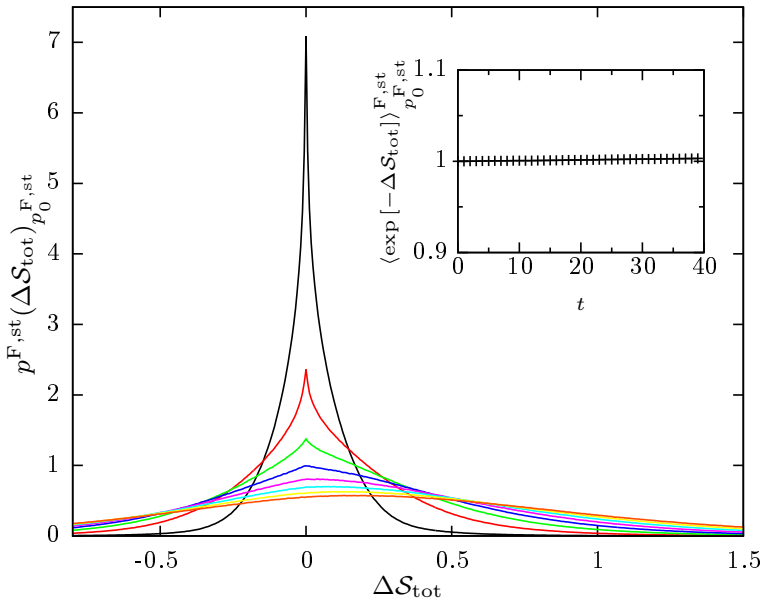

Figure 6.18: Distributions of dimensionless total entropy production $\Delta\mathcal{S}_{\mathrm{tot}}$ for a Langevin particle in the stationary state for temperature field $T(x) = 1 + (1/2)\tanh(x)$ and harmonic potential $\mathcal{F}(x) = -x$, for $\gamma = 10$ together with a demonstration of adherence to an IFT. Distributions shown are for process intervals from $t = 2$ (narrowest) to $t = 44$ (widest) in steps of 6 units.

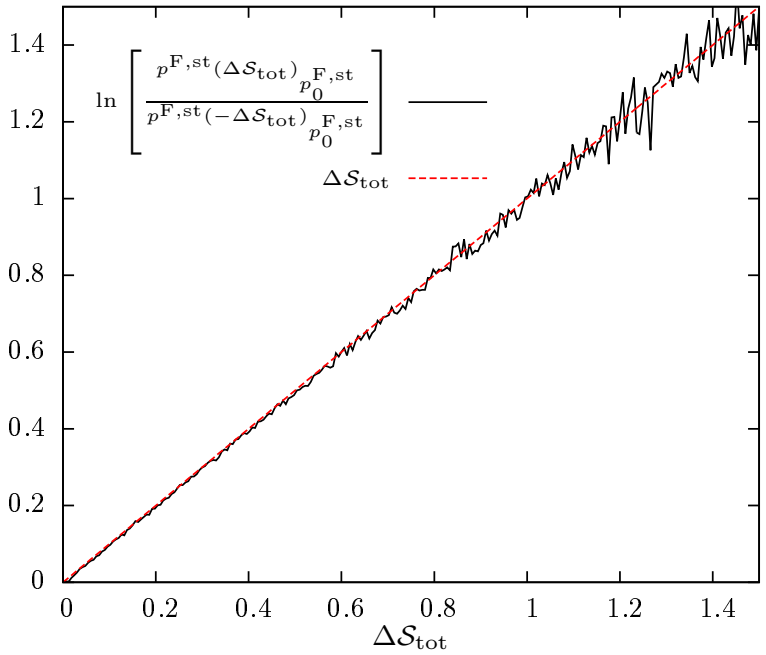

Figure 6.19: Verification of a detailed fluctuation theorem for a Langevin particle in the stationary state for temperature field $T(x) = 1 + (1/2)\tanh(x)$ and harmonic potential $\mathcal{F}(x) = -x$, using data from simulation for $\gamma = 10$ at time $t = 8$.

6.5 Discrete Representations of Full Space and the Nature of Coarse Graining

Finally, we turn our attention to a class of systems which deal with a set of discrete states and as such are described by master equations in order to illustrate the effects of such a formalism in such systems.

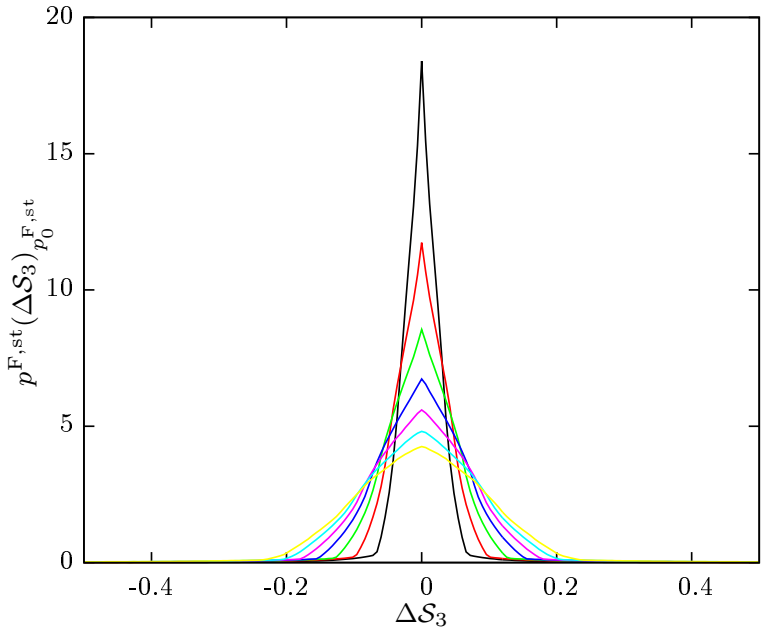

Figure 6.20: Distributions of $\Delta\mathcal{S}_3$ for a Langevin particle in the stationary state for temperature field $T(x) = 1 + (1/2)\tanh(x)$ and harmonic potential $\mathcal{F}(x) = -x$, evaluated at $t = 8$ for a range of γ from $\gamma^{-1} = 0.1$ (narrowest) to $\gamma^{-1} = 0.4$ (widest).

By introducing an odd variable into the dynamics we may obtain the three contributions to entropy laid out in previous chapters. At first it may certainly seem unnatural to deal with odd variables in this way, but there are many models in which odd variables are taken as discrete quantities, for example the Ising model and, albeit with a somewhat different purpose, lattice gas models. However, we prefer to leave the discussion as general as possible, simply considering that the probabilistic nature of the system is inherently related to the uncertainty with which one perceives the dynamics. If one is completely certain of the dynamics, entropy production, as defined here, vanishes. If one is solely uncertain of the environment, but completely certain in the dynamics (such that they can be resolved, or that we have no uncertainty in the underlying mechanics) of, for example, a colloidal particle, then a Langevin equation is appropriate and continuous representations of entropy production suffice. However, more generally, our uncertainty is much more profound. In an experiment governing the behaviour of a small particle, one might be unsure as to the nature of some coarse grained underlying dynamics and may only be able to resolve the position and velocity to within certain tolerances. If one can only measure, for example, whether the particle is moving one way or another, then, given no other information, we may only infer probabilistic behaviour of transitions between the two observed states. As such we present some simple toy systems which are discrete in nature, but with the introduction of odd variables which play the role of a rudimentary velocity as determined, in principle, by our degree of belief in how the system behaves given our uncertainty. We note however, that velocity variables being described by master equations is not a novel approach [168, 169]. We present some analytical results and demonstrate that the thermodynamics, qualitatively and quantitatively, differs depending on how one interprets the nature of the dynamics and the use of the variables in achieving them.

6.5.1 Introducing a Velocity Variable to a Random Walker

We consider a particle with dynamics modelled as a random walk on a lattice. In order to create a stationary state we employ periodic boundary conditions and may make this non-equilibrium by asserting that the probability of a move in each direction differ. We then however, introduce more information into our understanding of the model, by observing that if the particle makes a transition to the right this implies it has a positive velocity and that if it makes a transition to the left it has a negative velocity. As such by considering a one dimensional random walker we are able to 'infer' a probability distribution over velocity states. If we allow the random walker to move one lattice site at a time then there are only two inferred velocity variables, + and −. A schematic of this system is given in Fig. 6.21. Such a

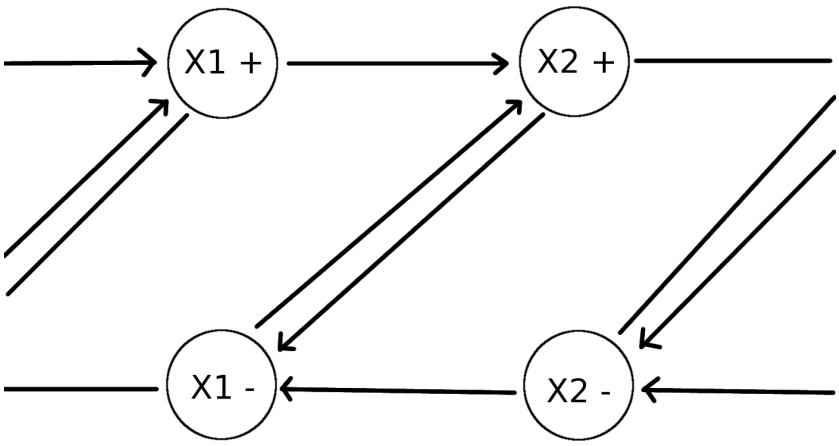

Figure 6.21: A random walker moves from left to right with transition rates $T(L \to R) = c + a$ and $T(R \to L) = c - a$. We then infer an implicit transition between velocity variables accounting for the diagonal transitions.

system is governed by the set of linear ordinary differential equations

$$\begin{aligned}\frac{dP^{\mathrm{F}}(X_i,+,t)}{dt} &= (c+a)P^{\mathrm{F}}(X_{i-1},-,t) + (c+a)P^{\mathrm{F}}(X_{i-1},+,t) - 2cP^{\mathrm{F}}(X_i,+,t)\\ \frac{dP^{\mathrm{F}}(X_i,-,t)}{dt} &= (c-a)P^{\mathrm{F}}(X_{i+1},-,t) + (c-a)P^{\mathrm{F}}(X_{i+1},+,t) - 2cP^{\mathrm{F}}(X_i,-,t).\end{aligned} \tag{6.96}$$

However, if we assert that the initial distribution is uniform in X then it will be for all time and so can write

$$\begin{aligned}\frac{dP^{\mathrm{F}}(X_i,+,t)}{dt} &= (c+a)P^{\mathrm{F}}(X_{i-1},-,t) - (c-a)P^{\mathrm{F}}(X_{i-1},+,t)\\ \frac{dP^{\mathrm{F}}(X_i,-,t)}{dt} &= -(c+a)P^{\mathrm{F}}(X_{i+1},-,t) + (c+a)P^{\mathrm{F}}(X_{i+1},+,t).\end{aligned} \tag{6.97}$$

This has stationary solution, for L spatial sites,

$$P^{\mathrm{F,st}}(X_i,\pm) = \frac{(c \pm a)}{2cL}. \tag{6.98}$$

Considering, in the same spirit as a force reversal in continuous systems, a switch between $a = a_0$ to $a = a_1$ the equations can be solved to give

$$P^{\mathrm{F}}(X_i, +, t) = \frac{c + a_1}{2cL} + \frac{a_0 - a_1}{2cL} e^{-2ct}$$
$$P^{\mathrm{F}}(X_i, -, t) = \frac{c - a_1}{2cL} + \frac{a_1 - a_0}{2cL} e^{-2ct}. \tag{6.99}$$

We may then consider the entropic contributions for any particular particle behaviour, which amounts either to an instantaneous transition or residence in a given state. We have already seen that $\Delta\mathcal{S}_3$ consists only of jump contributions, that $\Delta\mathcal{S}_2$ consists of both jump and residence contributions, but that if the mean escape rate between time reversed states are equal that the latter is zero. Previously, $\Delta\mathcal{S}_1$ has been written (noting that we shall generally consider these discrete systems athermal and so shall omit the factor k_B)

$$\Delta\mathcal{S}_1 = \ln \frac{P^{\mathrm{F}}(\boldsymbol{x}_0, 0)}{P^{\mathrm{F}}(\boldsymbol{x}_N, \tau)} + \sum_{i=1}^{N} \ln \frac{P^{\mathrm{F,st}}(\boldsymbol{x}_i, \lambda^{\mathrm{F}}(t_i))}{P^{\mathrm{F,st}}(\boldsymbol{x}_{i-1}, \lambda^{\mathrm{F}}(t_i))}, \tag{6.100}$$

but to identify contributions we should further divide this into residence and jump terms such that

$$\Delta\mathcal{S}_1 = \sum_{i=1}^{N+1} \ln \frac{P^{\mathrm{F}}(\boldsymbol{x}_{i-1}, t_{i-1})}{P^{\mathrm{F}}(\boldsymbol{x}_{i-1}, t_i)} + \sum_{i=1}^{N} \ln \frac{P^{\mathrm{F}}(\boldsymbol{x}_{i-1}, t_i) P^{\mathrm{F,st}}(\boldsymbol{x}_i, \lambda^{\mathrm{F}}(t_i))}{P^{\mathrm{F}}(\boldsymbol{x}_i, t_i) P^{\mathrm{F,st}}(\boldsymbol{x}_{i-1}, \lambda^{\mathrm{F}}(t_i))}. \tag{6.101}$$

If we then consider the situation $a_1 = a$ and $a_0 = -a$ we may find the relevant contributions in table 6.1.

Behaviour $\boldsymbol{x}_i \to \boldsymbol{x}_j$	$T(\boldsymbol{x}_j \vert \boldsymbol{x}_i)$	$\Delta\mathcal{S}_1$	$\Delta\mathcal{S}_2$	$\Delta\mathcal{S}_3$
$X_i+ \to X_{i-1}-$	$c - a$	$\ln \frac{(c-a)(c+a(1-2e^{-2ct}))}{(c+a)(c-a(1-2e^{-2ct}))}$	$\ln \frac{c-a}{c+a}$	$-2\ln \frac{c-a}{c+a}$
$X_i- \to X_{i+1}+$	$c + a$	$\ln \frac{(c+a)(c-a(1-2e^{-2ct}))}{(c-a)(c+a(1-2e^{-2ct}))}$	$\ln \frac{c+a}{c-a}$	$-2\ln \frac{c+a}{c-a}$
$X_i+ \to X_{i+1}+$	$c + a$	0	$\ln \frac{c+a}{c-a}$	0
$X_i- \to X_{i-1}-$	$c - a$	0	$\ln \frac{c-a}{c+a}$	0
$X_i+ \quad t \to t + \Delta t$	$-2c$	$\ln \frac{(c+a(1-2e^{-2ct}))}{(c+a(1-2e^{-2c(t+\Delta t)}))}$	0	0
$X_i- \quad t \to t + \Delta t$	$-2c$	$\ln \frac{(c-a(1-2e^{-2ct}))}{(c-a(1-2e^{-2c(t+\Delta t)}))}$	0	0

Table 6.1: Transition rates and path dependent entropy contributions for all particle behaviour for the random walker with an introduced velocity.

We then propose to calculate the instantaneous mean production rates for all three quantities under such behaviour for which we construct a path average of the form

$$\langle A[\vec{\boldsymbol{x}}] \rangle^{\mathrm{F}}_{P^{\mathrm{F}}_0} = \sum_{N=0}^{\infty} \int_0^{\tau} dt_1 \ldots \int_{t_{N-1}}^{\tau} dt_N \sum_{\vec{\boldsymbol{x}}} p^{\mathrm{F}}[\vec{\boldsymbol{x}}]_{P^{\mathrm{F}}_0} A[\vec{\boldsymbol{x}}] \tag{6.102}$$

where in turn, the inner sum represents a summation over all path sequences with N transitions, the integrals average over all possible transition times that satisfy $\tau \geq t_N > t_{N-1} > \ldots > t_1$ and the outer sum represents summation over the number of transitions in a given path. We, however, only consider up to first order in dt for which we need only consider $N = 0$ or $N = 1$ transitions. Further, for such a short time dt, we may exchange the integrals over probability densities by probabilities as in Eq. (2.109),

to obtain

$$\langle A[\vec{\boldsymbol{x}}]\rangle^{\mathrm{F}}_{P^{\mathrm{F}}_0} = \sum_{N=0}^{1}\sum_{\vec{\boldsymbol{x}}} P^{\mathrm{F}}[\vec{\boldsymbol{x}}]_{P^{\mathrm{F}}_0} A[\vec{\boldsymbol{x}}], \tag{6.103}$$

but only consider path weights up to $O(dt)$. Examining such probabilities for paths with transitions in the small interval dt (Eq. (2.109)) we may then simply consider the probability of a transition $T(\boldsymbol{x}_i|\boldsymbol{x}_{i-1})dt$ and of residence $\exp[T(\boldsymbol{x}_i|\boldsymbol{x}_i)dt] \simeq 1 + T(\boldsymbol{x}_i|\boldsymbol{x}_i)dt$. As such the instantaneous averages may be written

$$\begin{aligned}
&\langle d\Delta\mathcal{S}_1\rangle^{\mathrm{F}}_{P^{\mathrm{F,st}}_0} = \\
&\sum_{i=1}^{L} P^{\mathrm{F}}(X_i,+,t)\left[(c-a)dt\ln\frac{(c-a)(c+a(1-2e^{-2ct}))}{(c+a)(c-a(1-2e^{-2ct}))} + (1-2cdt)\ln\frac{(c+a(1-2e^{-2ct}))}{(c+a(1-2e^{-2c(t+dt)}))}\right]\\
&+\sum_{i=1}^{L} P^{\mathrm{F}}(X_i,-,t)\left[(c+a)dt\ln\frac{(c+a)(c-a(1-2e^{-2ct}))}{(c-a)(c+a(1-2e^{-2ct}))} + (1-2cdt)\ln\frac{(c-a(1-2e^{-2ct}))}{(c-a(1-2e^{-2c(t+dt)}))}\right]\\
&=\frac{1}{2c}(c+a(1-2e^{-2ct}))\left[(c-a)dt\ln\frac{(c-a)(c+a(1-2e^{-2ct}))}{(c+a)(c-a(1-2e^{-2ct}))} - (1-2cdt)\frac{4ace^{-2ct}dt}{c+a(1-2e^{-2ct})}\right]\\
&+\frac{1}{2c}(c-a(1-2e^{-2ct}))\left[(c+a)dt\ln\frac{(c+a)(c-a(1-2e^{-2ct}))}{(c-a)(c+a(1-2e^{-2ct}))} + (1-2cdt)\frac{4ace^{-2ct}dt}{c-a(1-2e^{-2ct})}\right]\\
&=\frac{1}{2c}(c+a(1-2e^{-2ct}))\left[(c-a)dt\ln\frac{(c-a)(c+a(1-2e^{-2ct}))}{(c+a)(c-a(1-2e^{-2ct}))} - \frac{4ace^{-2ct}dt}{c+a(1-2e^{-2ct})}\right]\\
&+\frac{1}{2c}(c-a(1-2e^{-2ct}))\left[(c+a)dt\ln\frac{(c+a)(c-a(1-2e^{-2ct}))}{(c-a)(c+a(1-2e^{-2ct}))} + \frac{4ace^{-2ct}dt}{c-a(1-2e^{-2ct})}\right] + O(dt^2)\\
&=\frac{1}{2c}(c+a(1-2e^{-2ct}))\left[(c-a)dt\ln\frac{(c-a)(c+a(1-2e^{-2ct}))}{(c+a)(c-a(1-2e^{-2ct}))}\right]\\
&+\frac{1}{2c}(c-a(1-2e^{-2ct}))\left[(c+a)dt\ln\frac{(c+a)(c-a(1-2e^{-2ct}))}{(c-a)(c+a(1-2e^{-2ct}))}\right] + O(dt^2)
\end{aligned} \tag{6.104}$$

such that

$$\frac{d\langle\Delta\mathcal{S}_1\rangle^{\mathrm{F}}_{P^{\mathrm{F,st}}_0}}{dt} = 2ae^{-2ct}\ln\left[\frac{1+2ae^{-2ct}/(c-a)}{1-2ae^{-2ct}/(c+a)}\right]. \tag{6.105}$$

Similarly,

$$\begin{aligned}
\langle d\Delta\mathcal{S}_2\rangle^{\mathrm{F}}_{P^{\mathrm{F,st}}_0} &= \sum_{i=1}^{L} P^{\mathrm{F}}(X_i,+,t)\left[(c-a)dt\ln\frac{(c-a)}{(c+a)} + (c+a)dt\ln\frac{(c+a)}{(c-a)}\right]\\
&+\sum_{i=1}^{L} P^{\mathrm{F}}(X_i,-,t)\left[(c+a)dt\ln\frac{(c+a)}{(c-a)} + (c-a)dt\ln\frac{(c-a)}{(c+a)}\right]\\
&=\frac{1}{2c}(c+a(1-2e^{-2ct}))\left[(c-a)dt\ln\frac{(c-a)}{(c+a)} + (c+a)dt\ln\frac{(c+a)}{(c-a)}\right]\\
&+\frac{1}{2c}(c-a(1-2e^{-2ct}))\left[(c+a)dt\ln\frac{(c+a)}{(c-a)} + (c-a)dt\ln\frac{(c-a)}{(c+a)}\right]
\end{aligned} \tag{6.106}$$

which gives

$$\frac{d\langle\Delta\mathcal{S}_2\rangle^{\mathrm{F}}_{P^{\mathrm{F,st}}_0}}{dt} = 2a\ln\left[\frac{c+a}{c-a}\right]. \tag{6.107}$$

Finally

$$\begin{aligned}\langle d\Delta\mathcal{S}_3\rangle^{\mathrm{F}}_{P_0^{\mathrm{F,st}}} &= \sum_{i=1}^{L} P^{\mathrm{F}}(X_i,+,t)\left[-2(c-a)dt\ln\frac{(c-a)}{(c+a)}\right] \\ &+ \sum_{i=1}^{L} P^{\mathrm{F}}(X_i,-,t)\left[-2(c+a)dt\ln\frac{(c+a)}{(c-a)}\right] \\ &= \frac{1}{2c}(c+a(1-2e^{-2ct}))\left[-2(c-a)dt\ln\frac{(c-a)}{(c+a)}\right] \\ &+ \frac{1}{2c}(c-a(1-2e^{-2ct}))\left[-2(c+a)dt\ln\frac{(c+a)}{(c-a)}\right] \end{aligned} \tag{6.108}$$

which yields

$$\frac{d\langle\Delta\mathcal{S}_3\rangle^{\mathrm{F}}_{P_0^{\mathrm{F,st}}}}{dt} = -4ae^{-2ct}\ln\left[\frac{c+a}{c-a}\right]. \tag{6.109}$$

Again, we observe the expected behaviour in all quantities. $\langle\Delta\mathcal{S}_1\rangle^{\mathrm{F}}_{P_0^{\mathrm{F}}}$ and $\langle\Delta\mathcal{S}_3\rangle^{\mathrm{F}}_{P_0^{\mathrm{F}}}$ contribute transiently, whereas $\langle\Delta\mathcal{S}_2\rangle^{\mathrm{F}}_{P_0^{\mathrm{F}}}$ persists in the steady state along with explicit positivity in all but $\langle\Delta\mathcal{S}_3\rangle^{\mathrm{F}}_{P_0^{\mathrm{F}}}$. By plotting the contributions for such a process we find the results in Fig. 6.22.

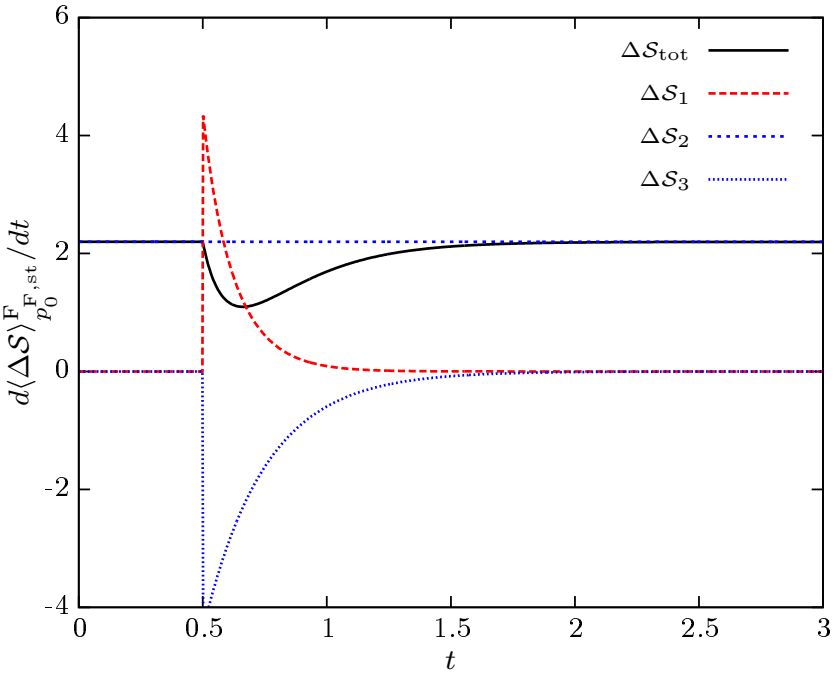

Figure 6.22: Entropy contributions for a reversal of the non-equilibrium constraint $a = 1$ with $c = 2$ from the stationary state for the random walker on a lattice with an introduced velocity variable.

6.5.2 Two State Ballistic Particle: A Modified Telegraph Process

Similar to the above model, we again consider a particle able to move in one dimension, but imagine now that intrinsically to the model it can take two defined velocities $+v$ and $-v$ and that this time fluctuations are driven in velocity space such that the spatial variable becomes irrelevant. For example

it could follow ballistically such that

$$x(t) = x(0) + \int_0^\tau v(t)dt. \tag{6.110}$$

This is in essence the motion (with slight adaptations) that defines the Kac or telegraph process [170] and the distribution in space is known to obey the telegrapher's equation. In this description there are only two states relating to forward and backwards motion. If the transition rates are given by $P(-|+) = (c-a)$ and $P(+|-) = (c+a)$ we once again have the master equation

$$\begin{aligned}\frac{dP^{\mathrm{F}}(+,t)}{dt} &= (c+a)P^{\mathrm{F}}(-,t) - (c-a)P^{\mathrm{F}}(+,t)\\ \frac{dP^{\mathrm{F}}(-,t)}{dt} &= (c-a)P^{\mathrm{F}}(+,t) - (c+a)P^{\mathrm{F}}(-,t)\end{aligned} \tag{6.111}$$

with solution

$$P^{\mathrm{F}}(\pm,t) = \frac{1}{2c}\left(c \pm a\left(1-2e^{-2ct}\right)\right). \tag{6.112}$$

Since, the solution is of the same form as for the adapted random walker it provides a good illustration of the thermodynamic difference resulting from the underlying microscopic uncertainty as their macroscopic behaviour is so similar. We find the three contributions to entropy production related to residence times and transitions and illustrate them in table 6.2. We then as before find the first order mean contribution

Behaviour $\boldsymbol{x}_i \to \boldsymbol{x}_j$	$T(\boldsymbol{x}_j\|\boldsymbol{x}_i)$	$\Delta\mathcal{S}_1$	$\Delta\mathcal{S}_2$	$\Delta\mathcal{S}_3$
$+ \to -$	$(c-a)$	$\ln\frac{(c-a)(c+a(1-2e^{-2ct}))}{(c+a)(c-a(1-2e^{-2ct}))}$	$\ln\frac{c-a}{c+a}$	$-2\ln\frac{c-a}{c+a}$
$- \to +$	$(c+a)$	$\ln\frac{(c+a)(c-a(1-2e^{-2ct}))}{(c-a)(c+a(1-2e^{-2ct}))}$	$\ln\frac{c+a}{c-a}$	$-2\ln\frac{c+a}{c-a}$
$+\ \ t \to t+\Delta t$	$(a-c)$	$\ln\frac{(c+a(1-2e^{-2ct}))}{(c+a(1-2e^{-2c(t+\Delta t)}))}$	$2a\Delta t$	0
$-\ \ t \to t+\Delta t$	$-(a+c)$	$\ln\frac{(c-a(1-2e^{-2ct}))}{(c-a(1-2e^{-2c(t+\Delta t)}))}$	$-2a\Delta t$	0

Table 6.2: Transition rates and path dependent entropy contributions for all particle behaviour for the telegraph process.

in dt. By consideration of all particle behaviour we see that $\langle\Delta\mathcal{S}_1\rangle^{\mathrm{F}}_{P_0^{\mathrm{F,st}}}$ and $\langle\Delta\mathcal{S}_3\rangle^{\mathrm{F}}_{P_0^{\mathrm{F,st}}}$ are of the same form, but that this is not the case for $\langle\Delta\mathcal{S}_2\rangle^{\mathrm{F}}_{P_0^{\mathrm{F}}}$ which is given by

$$\begin{aligned}\frac{d\langle\Delta\mathcal{S}_2\rangle^{\mathrm{F}}_{P_0^{\mathrm{F,st}}}}{dt} &= P^{\mathrm{F}}(+,t)\left[(c-a)\Delta\mathcal{S}_2^{+\to-} + \Delta\mathcal{S}_2^{+\to+}\right]\\ &\quad + P^{\mathrm{F}}(-,t)\left[(c+a)\Delta\mathcal{S}_2^{-\to+} + \Delta\mathcal{S}_2^{-\to-}\right] + O(dt)\\ &= P^{\mathrm{F}}(+,t)\left[(c-a)\ln\left[\frac{c-a}{c+a}\right] + 2a\right] + P(-,t)\left[(c+a)\ln\left[\frac{c+a}{c-a}\right] - 2a\right] + O(dt)\\ &= \frac{2a^2}{c}\left(1-2e^{-2ct}\right) + 2ae^{-2ct}\ln\left[\frac{c+a}{c-a}\right].\end{aligned} \tag{6.113}$$

As such, we now observe a time dependent contribution to the generalised house-keeping heat which reduces to a different value in the stationary state. Furthermore, the behaviour of the total entropy production is altered and is of the form shown in Fig. 6.23.

We point out that a subtle distinction in the dynamics, whilst leaving the solution to the master equations unchanged, results in a marked difference in the characterisation of the irreversibility of the process as defined by the entropy production.

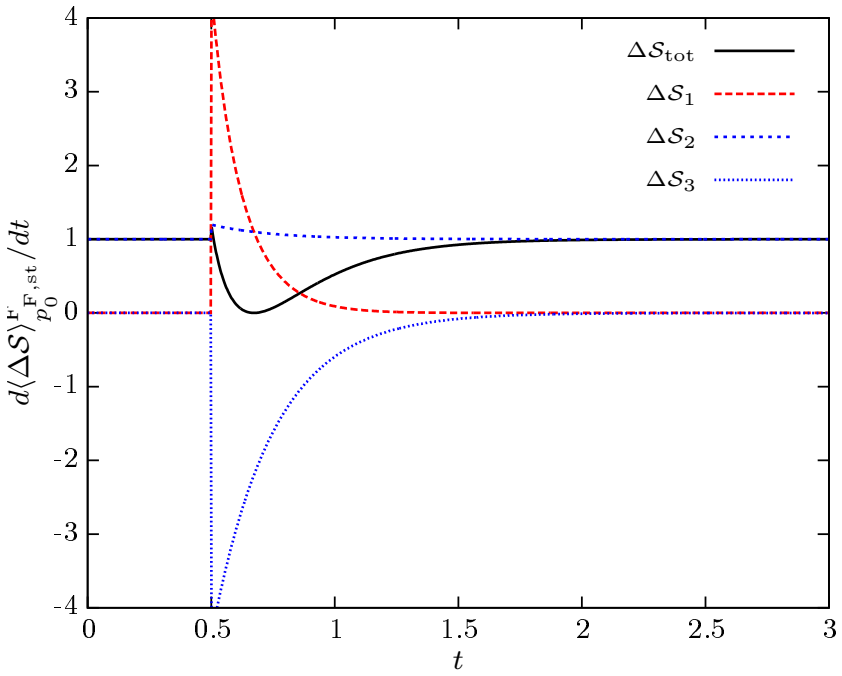

Figure 6.23: Entropy contributions for a reversal of the non-equilibrium constraint $a = 1$ with $c = 2$ from the stationary state.

We finally mention that one can simply verify the integral fluctuation theorems expected of $\Delta\mathcal{S}_1$, $\Delta\mathcal{S}_2$ and $\Delta\mathcal{S}_{\rm tot}$ by means of the same path averages (noting that if the model leads to vanishing contributions in entropy production, these should be included) such that, for example, up to $O(dt)$

$$\begin{aligned}\langle\exp[-\Delta\mathcal{S}_2]\rangle^{\rm F}_{P_0^{\rm F}} =&P^{\rm F}(+,t)\left[(c-a)dt\exp[-\Delta\mathcal{S}_2^{+\to-}]+(1-(c-a)dt)\exp[-\Delta\mathcal{S}_2^{+\to+}]\right]\\ &+P^{\rm F}(-,t)\left[(c+a)dt\exp[-\Delta\mathcal{S}_2^{-\to+}]+(1-(c+a)dt)\exp[-\Delta\mathcal{S}_2^{-\to-}]\right]+O(dt^2)\\ =&P^{\rm F}(+,t)\left[(c-a)dt\frac{c+a}{c-a}+(1-(c-a)dt)\exp[-2adt]\right]\\ &+P^{\rm F}(-,t)\left[(c+a)dt\frac{c-a}{c+a}+(1-(a+c)dt)\exp[2adt]\right]+O(dt^2)\\ =&P^{\rm F}(+,t)\left[1+O(dt^2)\right]+P(-,t)\left[1+O(dt^2)\right]\\ =&1.\end{aligned}\tag{6.114}$$

By virtue of the above, not only is the result $\langle\exp[-\Delta\mathcal{S}_2]\rangle^{\rm F}_{P_0^{\rm F}} = 1$ upheld for the range dt, but it also demonstrates $d\langle\exp[-\Delta\mathcal{S}_2]\rangle^{\rm F}_{P_0^{\rm F}}/dt = 0$, which along with the Markovian nature of the dynamics ensures the result holds for all time.

6.5.3 Simple Model of Thermal Conduction

Next we propose what we believe to be one of the simplest possible models for thermal conduction that allows for an appreciation all relevant thermodynamic quantities. It comprises only 4 states which we take to be a slow and fast velocity in both a left and right direction in a tight enough potential such that we deem there to be a singular position state. We consider that when our model possesses a negative velocity it comes in contact with a thermal wall on the left with some mean rate, upon which

the particle is reflected meaning the particle transitions into a positive velocity state and vice versa. By then specifying that the left thermal wall, on average favours conversion from a slow velocity to a fast velocity we may consider it 'hot' and by specifying that the right thermal wall, on average favours conversion from a fast velocity to a slow velocity we may consider it 'cool'. The exact nature of the transition rates which amount to this behaviour can take many forms, but we propose one that makes the individual contributions to entropy production somewhat simpler in form and write it in terms of constants A and thermal gradient parameter ΔT. The transition rates we propose take the form given in table 6.3. These in turn lead to the stationary probability distribution

Behaviour $\boldsymbol{x}_i \to \boldsymbol{x}_j$	$T(\boldsymbol{x}_j \vert \boldsymbol{x}_i)$
$V_\mathrm{f} \to -V_\mathrm{f}$	$A - \Delta T$
$V_\mathrm{f} \to -V_\mathrm{s}$	$A + \Delta T$
$V_\mathrm{s} \to -V_\mathrm{f}$	A
$V_\mathrm{s} \to -V_\mathrm{s}$	A
$-V_\mathrm{f} \to V_\mathrm{f}$	A
$-V_\mathrm{f} \to V_\mathrm{s}$	A
$-V_\mathrm{s} \to V_\mathrm{f}$	$A + \Delta T$
$-V_\mathrm{s} \to V_\mathrm{s}$	$A - \Delta T$

Table 6.3: All possible transitions and rates for the 4 state model of thermal conduction.

$$P^{\mathrm{F,st}}(V_\mathrm{f}) = P^{\mathrm{F,st}}(-V_\mathrm{s}) = \frac{A}{2(2A - \Delta T)}$$
$$P^{\mathrm{F,st}}(-V_\mathrm{f}) = P^{\mathrm{F,st}}(V_\mathrm{s}) = \frac{A - \Delta T}{2(2A - \Delta T)}. \tag{6.115}$$

We may consider the dynamics in continuous time, noticing the specific rate choice means contributions to $\Delta\mathcal{S}_2$ arising from residence times are 0 for all states. We may, as before write down expressions for the entropy contributions, which for terms dependent on $P^\mathrm{F}(\pm V_{\mathrm{f,s}})$ we leave in implicit form, in table 6.4. Starting from equilibrium, $P^\mathrm{F}(\pm V_{\mathrm{f,s}}) = 0.25$, we may solve the relevant master equation numerically for

Behaviour	$\Delta\mathcal{S}_1$	$\Delta\mathcal{S}_2$	$\Delta\mathcal{S}_3$
$V_\mathrm{f} \to -V_\mathrm{f}$	$\ln \frac{P^\mathrm{F}(V_\mathrm{f},t)}{P^\mathrm{F}(-V_\mathrm{f},t)} + \ln\left(1 - \frac{\Delta T}{A}\right)$	$\ln\left(1 - \frac{\Delta T}{A}\right)$	$-2\ln\left(1 - \frac{\Delta T}{A}\right)$
$V_\mathrm{f} \to -V_\mathrm{s}$	$\ln \frac{P^\mathrm{F}(V_\mathrm{f},t)}{P^\mathrm{F}(-V_\mathrm{s},t)}$	$\ln\left(1 + \frac{\Delta T}{A}\right)$	0
$V_\mathrm{s} \to -V_\mathrm{f}$	$\ln \frac{P^\mathrm{F}(V_\mathrm{f},t)}{P^\mathrm{F}(-V_\mathrm{f},t)}$	$-\ln\left(1 + \frac{\Delta T}{A}\right)$	0
$V_\mathrm{s} \to -V_\mathrm{s}$	$\ln \frac{P^\mathrm{F}(V_\mathrm{s},t)}{P^\mathrm{F}(-V_\mathrm{s},t)} - \ln\left(1 - \frac{\Delta T}{A}\right)$	$-\ln\left(1 - \frac{\Delta T}{A}\right)$	$2\ln\left(1 - \frac{\Delta T}{A}\right)$
$-V_\mathrm{f} \to V_\mathrm{f}$	$\ln \frac{P^\mathrm{F}(-V_\mathrm{f},t)}{P^\mathrm{F}(V_\mathrm{f},t)} - \ln\left(1 - \frac{\Delta T}{A}\right)$	$-\ln\left(1 - \frac{\Delta T}{A}\right)$	$2\ln\left(1 - \frac{\Delta T}{A}\right)$
$-V_\mathrm{f} \to V_\mathrm{s}$	$\ln \frac{P^\mathrm{F}(-V_\mathrm{f},t)}{P^\mathrm{F}(V_\mathrm{s},t)}$	$-\ln\left(1 + \frac{\Delta T}{A}\right)$	0
$-V_\mathrm{s} \to V_\mathrm{f}$	$\ln \frac{P^\mathrm{F}(-V_\mathrm{s},t)}{P^\mathrm{F}(V_\mathrm{s},t)}$	$\ln\left(1 + \frac{\Delta T}{A}\right)$	0
$-V_\mathrm{s} \to V_\mathrm{s}$	$\ln \frac{P^\mathrm{F}(-V_\mathrm{s},t)}{P^\mathrm{F}(V_\mathrm{s},t)} + \ln\left(1 - \frac{\Delta T}{A}\right)$	$\ln\left(1 - \frac{\Delta T}{A}\right)$	$-2\ln\left(1 - \frac{\Delta T}{A}\right)$
$V_\mathrm{f}\ \ t \to t + \Delta t$	$\ln \frac{P^\mathrm{F}(V_\mathrm{f},t)}{P^\mathrm{F}(V_\mathrm{f},t+\Delta t)}$	0	0
$V_\mathrm{s}\ \ t \to t + \Delta t$	$\ln \frac{P^\mathrm{F}(V_\mathrm{s},t)}{P^\mathrm{F}(V_\mathrm{s},t+\Delta t)}$	0	0
$-V_\mathrm{f}\ \ t \to t + \Delta t$	$\ln \frac{P^\mathrm{F}(-V_\mathrm{f},t)}{P^\mathrm{F}(-V_\mathrm{f},t+\Delta t)}$	0	0
$-V_\mathrm{s}\ \ t \to t + \Delta t$	$\ln \frac{P^\mathrm{F}(-V_\mathrm{s},t)}{P^\mathrm{F}(-V_\mathrm{s},t+\Delta t)}$	0	0

Table 6.4: All possible behaviour and corresponding entropy productions for the 4 state model of thermal conduction.

a given (even) protocol

$$\Delta T(t) = \begin{cases} \frac{1}{4}(1-\cos(\pi t)) & t < 1, \\ 0.5 & 1 \leq t \leq 4, \\ \frac{1}{4}(1+\cos(\pi(t-4))) & t > 4, \end{cases} \tag{6.116}$$

which yields the result in Fig. 6.24. Further, by utilising the numerical solution we may calculate the

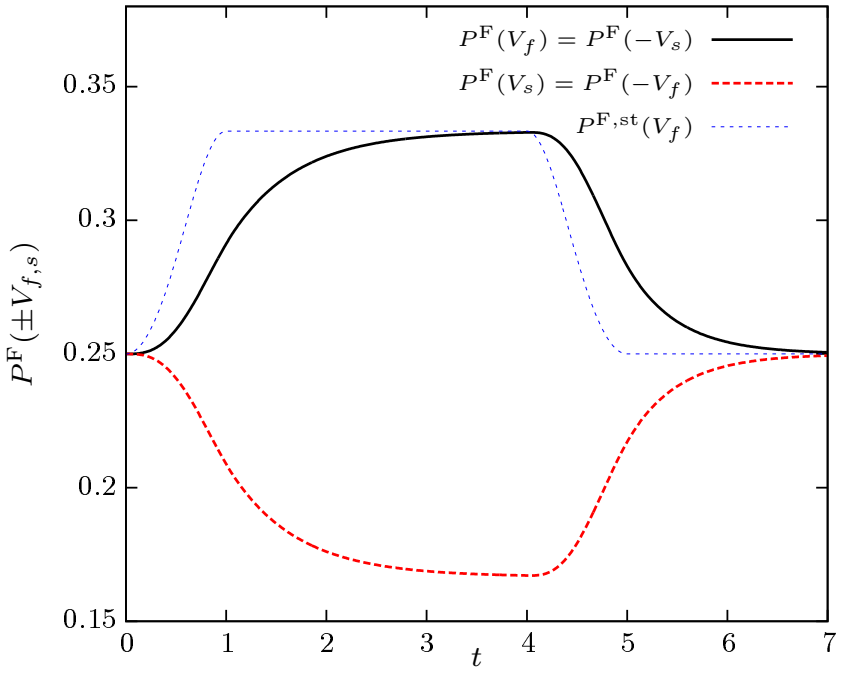

Figure 6.24: Solution for $P^{\mathrm{F}}(V_{\mathrm{f}})$, $P^{\mathrm{F}}(V_{\mathrm{s}})$, $P^{\mathrm{F}}(-V_{\mathrm{f}})$, $P^{\mathrm{F}}(-V_{\mathrm{s}})$ for protocol $\Delta T(t)$ starting from an equilibrium stationary state.

instantaneous mean entropy production rates given such a protocol which are given in Fig. 6.25. By comparing the two figures we may once again benefit from an intuitive understanding of the three different contributions to entropy production which is particularly clear given the simplicity of the model used and the simple relation between the stationary distribution and the non-equilibrium constraint. As one expects we find a rigorously positive contribution $d\langle\Delta\mathcal{S}_1\rangle^{\mathrm{F}}_{P_0^{\mathrm{F}}}/dt$ which is present when the distribution is different to the stationary distribution and so indicates relaxation through evolution of the probability distribution. $d\langle\Delta\mathcal{S}_2\rangle^{\mathrm{F}}_{P_0^{\mathrm{F}}}/dt$ follows the evolution of the stationary distribution, $P^{\mathrm{F,st}}(V)$, and contributes when it differs from the equilibrium value $P^{\mathrm{F,eq}}(V) = 0.25$ and indicates the entropy production expected in the stationary state. $d\langle\Delta\mathcal{S}_3\rangle^{\mathrm{F}}_{P_0^{\mathrm{F}}}/dt$ contributes when the distribution differs from the stationary distribution like $d\langle\Delta\mathcal{S}_1\rangle^{\mathrm{F}}_{P_0^{\mathrm{F}}}/dt$, but can take either sign depending on whether the instantaneous distribution is 'further' or 'closer' to the equilibrium state than the state the system is relaxing towards; or more accurately whether calculation of entropy production due to the non-equilibrium constraint in the stationary state over or underestimates the entropy production that arises transiently because of the same non-equilibrium constraint. Finally, the sum of all three is the total entropy production, which as the measure of irreversibility taken to be synonymous with degree of non-equilibrium behaviour, follows the progression of the distribution away from the equilibrium distribution which arises when the non-equilibrium constraint vanishes ($\Delta T = 0$).

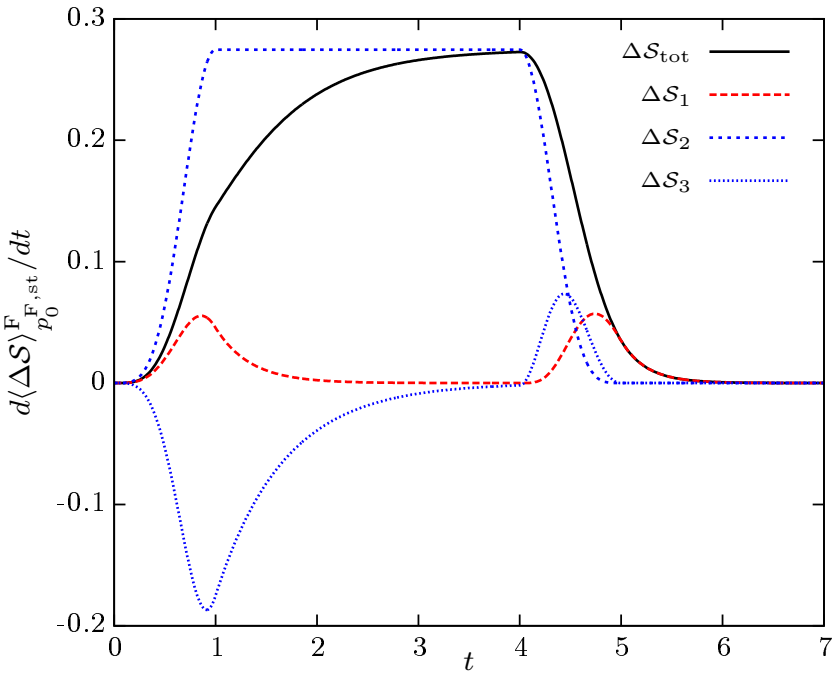

Figure 6.25: Entropy contributions for evolution from the stationary state following the protocol $\Delta T(t)$.

Chapter 7

Discussion and Conclusions

We have presented a refinement of some of the attempts to define non-equilibrium entropy production that have arisen in recent years with our main result being that if one considers time reversal fully by allowing more general transformations in its operation, both the form of the total entropy production and its division into contributions relevant to the thermodynamics of steady states needs modification. The first of these modifications is manifest in an extra term appearing in the total entropy production for systems on discrete state spaces based on mean escape rates with its inclusion providing an update to the original mean entropy production as described by Schnakenberg [58]. The revision to the division of the total entropy production so as to reflect the operational non-equilibrium thermodynamics arises due to the fact that one cannot define an alternative dynamics which possess the same stationary state whilst reversing only the entropy producing current. Or perhaps more clearly, it arises when one notices that there is a portion of the probability current which is not entropy producing. This has the effect of introducing an additional entropy production contribution, based solely on variables and protocols which transform to different values upon time reversal and which, in the mean, behaves transiently. This contribution, however, is not bounded like the remaining two contributions meaning that previous assertions that the mean total entropy production can always be divided into two rigorously non-negative components [36] are incomplete. The properties and implications of this novel entropy division have then been explored both in discrete and continuous state spaces. In particular, in continuous state spaces the general treatment of arbitrary correlated, multiplicative SDEs with explicit discussion of the delicate use of short time propagators for defining entropy productions with multiplicative noise is something we also believe to be novel. We have then illustrated these ideas with some simple examples both in discrete and continuous state space. First we have seen that by considering the velocity in the description of the system we may observe a finer structure in the total entropy production dependent on specific velocity dependent features such as alternative and additional friction sources and generalisations that allow for relativistic speeds even for simple processes such as transitions between steady states provided by non-conservative forces. We then saw that in order to describe entropy production due to thermal transport, odd variables are not simply an addition to make the model more sophisticated, but actually necessary in order to provide a meaningful description and saw that such a system, allowing non-zero currents by exploiting the higher dimensionality, invokes the parity dependent structure in the definition of the entropy contribution associated with the generalised house-keeping heat. Finally, we illustrated the concepts for discrete state space models highlighting that the thermodynamics is subtly dependent on the microscopic behaviour, not just the macroscopic properties of the distribution and utilised a very simple model of heat transport so as to gain a clear understanding of the meaning of each of the three entropy production contributions.

Such results, however, by refining the manner in which entropy production is defined, highlight some of the more subtle conceptual issues surrounding the definition of such a stochastic entropy production based on functionals that obey fluctuation theorems. We have seen that treating entropy production in this way means that depending on how one models the dynamics one can reach qualitatively different behaviour in the expected thermodynamics, for example, by considering full phase space we observe a finer structure in the entropy production than if the over-damped limit is utilised. These properties should reinforce a general attitude of caution; entropy production has a long history as a somewhat nebulous quantity straddling both thermodynamic and information theoretic concepts and its role as a path dependent functional is no different, particularly when one considers the structure of the system entropy, which by containing information about the entire distribution or ensemble, necessarily introduces an information-like entropy contribution. On the one hand entropy production as explicit irreversibility has the advantage of connecting such a macroscopic consequence to a thermodynamic quantity, but we should recognise that the irreversibility (which we somewhat arbitrarily insert a priori when we use stochastic dynamics) in the dynamics is reliant upon the stochastic behaviour which is inherently connected with the uncertainty we perceive in the dynamics. As such if we have a different degree of belief, or uncertainty, in the dynamics we observe different thermodynamics as evidenced in the difference between the results from the two discrete full phase models despite both having the same time dependent probability distributions and indeed the existence of a third entropy production contribution if one includes odd variables in the dynamics. In practice, this may mean that if one were to capture the effective dynamics of a system in a model by means of a stochastic interpretation, the irreversibility predicted may be misleading when compared to that observed practically in the real system depending on whether the approximation used in the model has coarse grained away some important feature and of course whether it is accurate at all. We note, of course, that we assume the underlying dynamics of some real system to be governed by reversible equations of motion: we expect the irreversibility we perceive to originate from the practical uncertainty that arises in trying to measure the dynamics. This, indeed, is what we are trying to capture in such models and so one could argue that there is a responsibility to match the specific stochastic behaviour with the uncertainty one observes experimentally. As such we stress that all the quantities derived and considered in this treatment represent the thermodynamics of a model and that one must always justify the model if one is to infer any real life consequence. The reliability of the model, however, is not the only source of ambiguity. By introducing odd time reversal transformations such issues arise in a slightly different manner through a certain freedom in what precisely one means by time reversal and how this impacts on the irreversibility. This may arise, for example, for an externally applied magnetic field or torque which does not feature in a Hamiltonian. Further, we have seen that time reversal can become even more ambiguous when one considers protocols that utilise feedback.

In general, what one may say about the path dependent entropy production in such systems is that it represents the irreversibility of the model one is utilising according to the definition of irreversibility that the modeller deems relevant with an onus on the modeller to justify such choices. We note that such a quantity will then always obey the relevant fluctuation theorems regardless of this definition; the existence of the fluctuation theorems and resultant second law inequalities are simply not enough to determine thermodynamic relevance. Indeed, the fact that the path dependent entropy productions obey the fluctuation theorems should also not be surprising when one examines their structure. In fact it has been pointed out that the fluctuation theorems themselves are strictly tautological statements [114] identifying that the existence of the symmetries are not, in and of themselves, remarkable. This

again stresses that it is the identification of relevant thermodynamic quantities as logarithmic ratios of path probabilities that is necessary to give them meaning which brings us back to our reasons for taking great care when defining time reversal.

Nevertheless, we conclude that defining the entropy production as we have done has further elucidated the nature of the non-equilibrium thermodynamics relevant to such stochastic systems. This development, arising from a further, yet demonstrably relevant, division of the dissipated heat transfer, leads to a greater complexity in the structure of the irreversibility characterised by entropy production. Given the broad relevance to any non-equilibrium system with odd variables we expect to observe further richness in the phenomenology of entropy production in many potential applications including, for example, any model that is based on Hamiltonian dynamics or that includes magnetic fields or moments.

Part II

Spatially Local Parallel Tempering

Foreword regarding the structure of work on spatially local parallel tempering

The following work describes and illustrates the design and implementation of a novel technique called spatially local parallel tempering. After an introduction, the necessary theory, formulation of the problem, how to measure any efforts to overcome it and its context within the literature are given in chapter 9. The specifics of the algorithm are then set out in chapter 10 through a series of toy models each progressively more complicated so as they introduce the key aspects of the algorithm. Some details of the algorithm, and particularly the quantities defined in order to test its effectiveness, are made specifically for the application to such models, but it is made clear when this is the case. The numerical results for the application to such models are given in chapter 11 and include evidence of convergence to the correct limiting distribution and of linear scaling, being the primary objective of such a technique. Finally a discussion is given in chapter 12.

Chapter 8

Introduction

In physics and chemistry much of the predictive power of computer simulation relies on the ability to calculate expectation values of physical quantities at a defined temperature at equilibrium. The high dimensional integrals required for an exact solution dictate that an approximation is required which is achieved by sampling the configuration space. The most common methods used to generate these configurations are Markov Chain Monte Carlo (MCMC) and Molecular Dynamics (MD). The power and simplicity of these methods mean that they have become the workhorses of computer simulation, however they struggle when the energy landscape consists of numerous minima separated by energy barriers which restrict sampling trajectories between them at the sampling temperature.

This issue of sampling systems with these well separated 'metastable' regions is so common that there has been a great deal of effort devoted to devising techniques which can overcome it. These are often referred to as 'accelerated sampling' and include a wide range of techniques including umbrella sampling [171], J-walking [172], multicanonical sampling [173], simulated tempering [174] and Wang Landau sampling [175]. One of the most successful however, is known as replica exchange Monte Carlo or parallel tempering [176–185].

As the temperature increases, transitions between these metastable regions become increasingly likely as the free energy surface flattens. It is this property which parallel tempering exploits by using independent high temperature simulations to generate configurations for the original simulation being run at a more prohibitive lower temperature. The idea is to simultaneously model the same system independently with each instance of the system being called a 'replica'. Each replica, now labelled i, is then simulated at its own temperature T_i using conventional MCMC or MD moves. Parallel Tempering then involves introducing a new kind of simulation move. Using a probabilistic attempt frequency it is proposed that the entire configurations of two of the replicas are swapped. This move is then accepted or rejected in such a way that the thermal equilibrium distribution is maintained in every replica. This can greatly accelerate the sampling as it allows the lower temperature replica to acquire configurations that are at a comparable energy, but far away (and obstructed by free energy barriers) in phase space from the replica being simulated at a higher temperature.

The parallel tempering method however, is not perfect. One of the well known drawbacks of implementing parallel tempering is the poor scaling of computational effort with system size. As the modelled system gets larger it becomes increasingly unlikely that there will be a configuration of comparable energy in each replica given the temperature of each replica is fixed. This results in a poor chance of

an exchange between replicas, weakening the performance of the algorithm. This can only be overcome by introducing additional replicas at a temperature that lies between the already existing replicas. This of course has a direct cost on the computational effort. A simple analysis [180, 186] suggests that the number of replicas needed will increase as $N^{1/2}$ where N is the number of degrees of freedom in the system. The true cost however, will be somewhat worse than this as there is an additional diffusive cost between replicas. To be most efficient, configurations must be successfully exchanged through all the replicas from the highest in temperature to the lowest. Owing to the stochastic nature of the exchange moves this resembles a diffusive process and as such is expected to scale accordingly.

It has been noted previously that this issue could be overcome by introducing the concept of locality into the parallel tempering algorithm by exchanging parts of the system as opposed to the entire configuration. Some progress has been made in that area by selectively tempering a part of the system [187, 188], however there exists no algorithm which applies these ideas throughout an extended system generally. We go on to demonstrate a new technique which successfully applies parallel tempering locally in this way. As we wish to improve the scaling of computational effort with respect to system size, we consider appropriate systems where energy barriers are distributed throughout the system in a mostly uniform way. We also envisage that the process of transitioning between the numerous metastable regions is therefore an inherently local one. This means that equilibration and sampling in a given local region should not depend on events in other, remote local regions. The amount of computational effort needed to equilibrate the entire system and to perform the sampling necessary to compute local quantities to a given statistical accuracy should then be proportional, in some sense, to the number of atoms in the system. Of course, the effort needed in practice will depend strongly on the algorithm used to calculate the total energy of the system. Obviously, if the total energy is a sum of short-range pair potentials, the effort will be far less than if *ab initio* methods are used.

However, for the size-scaling properties of the thermal sampling algorithm itself, our requirement is that *if* the number of computer operations needed to evaluate the energy of the entire system is proportional to N, *then* the number of operations needed for thermal equilibration and computation of local quantities should also be proportional to N. Normal PT algorithms do not have this property, because they are based on the global swap of configurations. We shall refer to PT algorithms having the property of locality and linear scaling as local parallel-tempering (LPT) algorithms. When we need to contrast these with normal (non-linear-scaling) PT algorithms, we shall call the latter canonical parallel-tempering (CPT) algorithms.

Following a brief overview of well accepted equilibrium sampling techniques the LPT algorithm will be explained by use of several toy models. First the basic ideas will be demonstrated using a model that consists of a one dimensional cyclic chain of double wells with a nearest neighbour coupling. Secondly the method will be extended to a one dimensional periodic potential where particles are free to move throughout the entire system. Lastly it will be demonstrated that correct results can be achieved in higher dimensions. Scaling of computational effort will be discussed along with an assessment of the LPT algorithm's ability to sample the correct equilibrium distribution by comparison with results from CPT.

Chapter 9

Theory and Methodology

9.1 Monte Carlo Methods

9.1.1 Sampling the Equilibrium Distribution

In equilibrium statistical physics we regularly require the average thermodynamic quantities of a system which for a quantity A in a system described by N particles with independent position variables is given by

$$\langle A \rangle = \frac{1}{Z} \int d\mathbf{r}^0 \dots d\mathbf{r}^{N-1} A(\mathbf{r}^0, \dots, \mathbf{r}^{N-1}) \exp\left[-\frac{U(\mathbf{r}^0, \dots, \mathbf{r}^{N-1})}{k_B T} \right] \tag{9.1}$$

where U is the internal energy and where Z is the partition integral

$$Z = \int d\mathbf{r}^0 \dots d\mathbf{r}^{N-1} \exp\left[-\frac{U(\mathbf{r}^0, \dots, \mathbf{r}^{N-1})}{k_B T} \right]. \tag{9.2}$$

This in theory allows us to predict any measurable property from first principles. However, whilst the instantaneous quantity $A(\mathbf{r}^0, \dots, \mathbf{r}^{N-1})$ is typically easy to compute, the integrals over all variables prove to be intractable for all but the most simple systems.

Given that complete enumeration is impossible it becomes necessary for numerical approximations to be employed to calculate these integrals. A first attempt at this is the Monte Carlo method. This involves approximating the integrals by means of choosing a finite and manageable number of random system configurations by use of pseudo-random numbers. The average constituting an approximation to $\langle A \rangle$ is then given by

$$\bar{A} = \frac{1}{n} \sum_i^n A_i \exp\left[-\frac{U(\mathbf{r}_i^0, \dots, \mathbf{r}_i^{N-1})}{k_B T} \right]. \tag{9.3}$$

Owing to the law of large numbers we expect $\bar{A} \to \langle A \rangle$ as $n \to \infty$. This however, is usually unworkable owing to a large variance in the exponential Boltzmann weighting factor. Or rather that we find most relevant configurations that make up the integral in Eq. (9.1) are localised in phase space. The result of which is that too much effort is expended on configurations which contribute very little to the overall average leaving the majority of the integral approximated by only a few random configurations.

What we require is a random sampling of the integral in Eq. (9.1) whereby the number of samples for each region of phase space is proportional to the equilibrium distribution function on that phase space. This is called importance sampling and can be achieved using the well known method, Markov

Chain Monte Carlo (MCMC), which dictates that samples are accessed by performing a random walk, by means of a Markov chain, through configuration space. The relative time that the Markov chain occupies each state is determined by an accept/reject criterion at each step. The accept/reject criterion must be determined using the equilibrium probability of the proposed state which in general is given by

$$p(\mathbf{r}^0, \ldots, \mathbf{r}^{N-1}) = \frac{1}{Z} \exp\left[-\frac{U(\mathbf{r}^0, \ldots, \mathbf{r}^{N-1})}{k_B T}\right]. \tag{9.4}$$

Whilst this still contains the unworkable partition integral, the nature of importance sampling and the Markov property ensures that the accept/reject criterion will be a function of the ratio of the probabilities. This means the partition functions cancel and the equilibrium distribution can be sampled without the costly integral. When carrying out importance sampling the average of a quantity A is now simply given by

$$\bar{A} = \frac{1}{n} \sum_{i}^{n} A_i. \tag{9.5}$$

9.1.2 Reaching the Limiting Distribution with a Markov Chain

We consider, for simplicity, a Markov chain, $X^1, X^2 \ldots X^n$, defined on a finite, discrete state space, $x \in \{x_1, x_2 \ldots x_m\}$. MCMC typically operates in discrete time and so we define the probability distribution for the n^{th} time step

$$P(X^n = x) = P^n. \tag{9.6}$$

At each time step there exists an individual probability of moving from one state to another that is given by a transition probability which we write

$$P(X^n = x_j | X^{n-1} = x_i) = M_{ji}. \tag{9.7}$$

Given a single distribution that we wish to sample, we consider the transition probabilities to be time independent so that the Markov Chain is time homogeneous. These transition probabilities then determine the nature of the Markov Chain and thus the sampling which can be achieved. For example, for a Markov chain to be physical it is required to be stochastic such that it has a conserved probability of transitioning from a given state into any other state equal to unity. As such we can call a Markov chain stochastic if

$$\sum_{j=1}^{m} M_{ji} = 1 \qquad \forall\, i. \tag{9.8}$$

This ensures that the total probability of finding the chain in any state is unity so that we necessarily obtain

$$\sum_{j=1}^{m} P^n(x_j) = 1. \tag{9.9}$$

Considering now the total probability flux into a given state as a sum of contributions from all possible previous states we can construct

$$P^n(x_j) = \sum_{i=1}^{m} M_{ji} P^{n-1}(x_i). \tag{9.10}$$

We now see that by forming a stochastic and homogeneous transfer matrix $M = (M_{ji})$ that acts upon the probability distribution, we can determine the probability distribution at step n from the transition matrix and its initial distribution

$$P^n = M^{n-1} P^1. \tag{9.11}$$

We now introduce the concept of a stationary distribution defined as the probability distribution that is explicitly invariant under any application of the transfer matrix such that

$$P^{\mathrm{st}} = MP^{\mathrm{st}}, \tag{9.12}$$

which may or may not exist for some transition matrix M. We also introduce the concept of a limiting distribution of a Markov chain which is defined as the long term behaviour of the probability distribution in the limit of an infinite number of applications of the transfer matrix, from a given starting distribution P^1, such that

$$P^{\lim} = \lim_{n\to\infty} M^n P^1 \tag{9.13}$$

which similarly may or may not exist, but which depends on both M and the starting distribution. If, however, the limiting distribution exists it is a stationary distribution. Typically the stationary distribution is the probability one wishes to sample in the Markov schemes alluded to previously and one can reach it if the limiting distribution coincides with it. As such one requires certain conditions on M in order for the stationary distribution to exist and be equal to the limiting distribution given any starting distribution. These conditions are that of

- Irreducibility: This is a condition often phrased that all states 'communicate' which amounts to being able to write $(M^n)_{ij} > 0 \quad n \in \mathbb{N} \quad i,j \in \{1,\ldots,m\}$ such that given any state there is a finite probability of transitioning to any other given state (including the same state) in some natural number n transitions which may depend on i and j.

- Aperiodicity: The period of a state, x_i, in a Markov system is the number, d_i, which is the greatest common divisor of all numbers n for which $(M^n)_{ii} > 0$. If $d_i = 1 \quad \forall i \in \{1,\ldots,m\}$ then the chain is said to be aperiodic. Alternatively, one may express this condition as the following: for any given $i \in \{1,\ldots,m\}$, there exists a number n, specific to that state, such that $(M^k)_{ii} > 0 \quad \forall\, k > n$.

A constraint frequently used to ensure convergence to the limiting distribution is microscopic reversibility or detailed balance which is defined here as

$$M_{ji}P^{\mathrm{st}}(x_i) = M_{ij}P^{\mathrm{st}}(x_j) \qquad \forall\, i,j. \tag{9.14}$$

We shall find it useful to note that detailed balance is a strong condition sufficient to lead to the stationary distribution, but not a necessary one. A weaker condition than strict reversibility is that of balance which is defined here as

$$P^{\mathrm{st}}(x_j) = \sum_{i=1}^{m} M_{ji}P^{\mathrm{st}}(x_i) \tag{9.15}$$

which coincides with Eq. (9.12). Within the literature on utilising Markov Chains for Monte Carlo simulation, Manousiouthakis and Deem [189] emphasised this point. In particular, they highlighted that, to obey detailed balance, the manner in which multiple particle systems are updated, according to the relevant transition matrix M, must be performed in ways which are strictly unnecessary to achieve balance and the two key conditions of irreducibility and aperiodicity. They then describe these last two conditions as the single constraint of 'regularity' which is defined as the existence of a *single*, state independent, number n such that $(M^n)_{ij} > 0$ for all i and j. The equivalence of an irreducible, aperiodic chain to a regular one can be recognised by noting that one can select an arbitrarily large and different k, from the definition of aperiodicity above, chosen individually for each state to match the number n from the irreducibility condition such that $n + k$ is the same for each state.

The above has been defined for systems of a finite number of discrete states, however the arguments can be extended to continuous systems. In these cases the probability density $p^n(x_i)$ is incremented according to the transition kernel $\gamma(x_j|x_i)$ such that

$$p^{n+1}(x_j) = \int dx_i \gamma(x_j|x_i) p^n(x_i) \tag{9.16}$$

where the integral is over all state space. The expression of balance in such system is given by

$$p^{\mathrm{st}}(x_j) = \int dx_i \gamma(x_j|x_i) p^{\mathrm{st}}(x_i) \tag{9.17}$$

and similarly detailed balance in such systems is expressed by

$$\gamma(x_i|x_j) p^{\mathrm{st}}(x_j) = \gamma(x_j|x_i) p^{\mathrm{st}}(x_i). \tag{9.18}$$

9.1.3 The Metropolis Algorithm

Proceeding with the consideration of continuous state spaces we discuss the Metropolis algorithm which is one of the most commonly used algorithms for ensuring a Markov chain converges to its stationary distribution. Typically it achieves this by ensuring detailed balance, which, for an n particle system, would constitute explicitly balancing the numbers of transitions of all n particles from a collective configuration i to j with the number of transitions j to i.

We start by considering the action of the transition kernel to consist of the probabilistic proposal of the next state, x_i, from x_j followed by the probabilistic acceptance, or rejection, of this new state. Explicitly, we thus consider the transition kernel to be the product of a proposal density $q(x_i|x_j)$ and the acceptance ratio $\alpha(x_i|x_j)$ such that detailed balance is now expressed

$$q(x_j|x_i)\alpha(x_j|x_i)p^{\mathrm{st}}(x_i) = q(x_i|x_j)\alpha(x_i|x_j)p^{\mathrm{st}}(x_j). \tag{9.19}$$

By considering the ratio of $\alpha(x_j|x_i)$ and $\alpha(x_i|x_j)$ we can see immediately upon substitution that the choice of acceptance ratio

$$\alpha(x_j|x_i) = \min\left(1, \frac{q(x_i|x_j)p^{\mathrm{st}}(x_j)}{q(x_j|x_i)p^{\mathrm{st}}(x_i)}\right) \tag{9.20}$$

leads to detailed balance as expressed in Eq. (9.19). In the case of thermal equilibrium sampling the desired probability density function is simply the Boltzmann distribution such that

$$p^{\mathrm{st}}(x_i) = \frac{1}{Z}\exp\left(-\frac{U(x_i)}{k_B T}\right). \tag{9.21}$$

Importantly the costly partition sum Z cancels rendering the acceptance ratio tractable meaning all one needs to calculate is

$$\alpha(x_j|x_i) = \min\left(1, \frac{q(x_i|x_j)}{q(x_j|x_i)}\exp\left(-\frac{(U(x_j) - U(x_i))}{k_B T}\right)\right). \tag{9.22}$$

One could then implement the above kernel by carrying out the following

- Starting in state $X^n = x_i$, we stochastically propose a new state x_i drawn from $q(x_j|x_i)$
- The acceptance probability $\alpha(x_j|x_i) = \min\left(1, \frac{q(x_i|x_j)}{q(x_j|x_i)}\exp\left(-\frac{(U(x_j)-U(x_i))}{k_B T}\right)\right)$ is calculated

- A random number is drawn uniformly between 0 and 1
- If this random number is less than $\alpha(x_j|x_i)$ the transition is accepted and the new state is $X^{n+1} = x_i$
- If the random number is greater than $\alpha(x_j|x_i)$ the transition is rejected and the new state is $X^{n+1} = x_j$.

Sampling performed in this way ensures the condition of detailed balance and allows us to sample all states reached with equal weighting and achieve the appropriate equilibrium integral since the time spent in each state is proportional to the Boltzmann factor. We note, however, that the Metropolis algorithm can be utilised in more elaborate ways. For example, if in an n particle system we use the algorithm to update one particle at a time, imposing detailed balance in that effective one particle system, we then only achieve detailed balance in the entire system subject to various constraints on the way in which we choose the order of updating those particles. However, since we do not require detailed balance to reach the stationary state, this will ultimately not matter as long as the individual transition kernels leads to the correct distribution utilised in the Metropolis algorithm.

9.1.4 General Properties of Random Walks in Phase Space

In Markov Chain Monte Carlo the usual way of choosing proposal configurations is by randomly suggesting new configurations based upon the current configuration. For example this could be achieved by displacing each particle position by both a random direction and distance. This leads to the Markov property and results in a random walk in phase space. The properties of random walks are well known with the most relevant being the statistics related to their scaling and distribution. Remarkably this theory holds for almost any realisation of a random walk in any number of dimensions allowing us to consider the most simple example, the 1D walker which steps either forwards or backwards one step with equal probability. It is relatively straightforward to demonstrate that the mean end to end distance of this random walk is given as $\sqrt{n}$ where n is the number of steps taken, however this is simply a specific consequence of the central limit theorem from which further properties can be derived. The central limit theorem states that the distribution of a sum of n independently and identically distributed random variables with mean μ and variance σ^2, tends to a Gaussian with mean $n\mu$ and variance $n\sigma^2$. More frequently this is written that given the quantity

$$S = \frac{1}{n}\sum_{i=1}^{n} X_i, \tag{9.23}$$

where each individual X_i is a random variable drawn independently from the same distribution with mean μ and variance σ^2, the random variable $\sqrt{n}(S-\mu)$ converges in distribution to the normal distribution with mean 0 and variance σ^2, $\mathcal{N}(0,\sigma^2)$. Of particular relevance to a discussion on parallel tempering and its scaling is that it demonstrates that the width of the approximate Gaussian can be expected to be proportional to $\sqrt{n}$ or more generally the number of random variables. The effectiveness of parallel tempering is strongly connected with the measure of the overlap between such distributions and as such, in order to quantify the scaling of PT algorithms, this description leads to a useful characterisation of that overlap.

9.2 Parallel Tempering

9.2.1 Reaching the Limiting Distribution with Parallel Tempering

Standard PT methods can be understood in terms of Markov processes on the states of a composite system consisting of a number of replicas of the physical system. If the physical system has a discrete set of states λ, then with two replicas the states of the composite system are specified by the pair $(\lambda_0; \lambda_1)$. For present purposes, CPT is an algorithm for sampling from states of the composite system with an equilibrium probability density distribution $p^{\rm st}(\lambda_0; \lambda_1) = p^{\rm st}_{\beta_0}(\lambda_0) p^{\rm st}_{\beta_1}(\lambda_1)$, which is the product of canonical thermal distributions at different temperatures: $p^{\rm st}_\beta(\lambda) = Z^{-1} \exp(-\beta U(\lambda))$, with $U(\lambda)$ the energy of state λ, β the inverse temperature (using units where $k_B = 1$), and Z the partition function. In practical applications, more than two replicas are often used.

In standard PT, there are two kinds of move: first, transitions between the states of a chosen individual replica, with the other replicas remaining unchanged; and second, swaps between states of chosen pairs of replicas. Moves of the first kind can be any kind of Markov transition that leaves the thermal equilibrium distribution $p^{\rm st}_{\beta_i}(\lambda_i)$ of each replica i invariant. For example, these moves could be standard Metropolis Monte Carlo moves, in which the attempt probability for any transition $\lambda \to \mu$ is identical to that of the reverse move $\mu \to \lambda$, with the acceptance probability $\alpha(\mu|\lambda)$ for the transition $\lambda \to \mu$ being given by the usual Metropolis algorithm:

$$\alpha(\mu|\lambda) = \min\left[1\,,\, \frac{q(\lambda|\mu)}{q(\mu|\lambda)} \exp\left(-\beta(U(\mu) - U(\lambda))\right)\right]. \tag{9.24}$$

By construction, the equilibrium distribution of the composite system remains invariant under moves of the first kind.

In the swap moves, transitions occur from state $(\lambda_0; \lambda_1)$ of the composite system to state $(\lambda_1; \lambda_0)$. The proposal density of a transition is typically state independent and so naturally symmetric and does not feature in the Metropolis algorithm. Consequently, an acceptance ratio that ensures the invariance of the composite thermal distribution is ensured requires, according to Eq. (9.19):

$$\frac{\gamma((\lambda_1;\lambda_0)|(\lambda_0;\lambda_1))}{\gamma((\lambda_0;\lambda_1)|(\lambda_1;\lambda_0))} = \frac{\alpha((\lambda_1;\lambda_0)|(\lambda_0;\lambda_1))}{\alpha((\lambda_0;\lambda_1)|(\lambda_1;\lambda_0))} = \frac{p^{\rm st}_{\beta_0}(\lambda_1) p^{\rm st}_{\beta_1}(\lambda_0)}{p^{\rm st}_{\beta_0}(\lambda_0) p^{\rm st}_{\beta_1}(\lambda_1)} = \exp\left[-(\beta_0 - \beta_1)(U(\lambda_1) - U(\lambda_0))\right]. \tag{9.25}$$

The commonly used acceptance ratio that ensures this condition is:

$$\alpha((\lambda_1;\lambda_0)|(\lambda_0;\lambda_1)) = \min\left[1, \exp\left(-(\beta_0 - \beta_1)(U(\lambda_1) - U(\lambda_0))\right)\right]. \tag{9.26}$$

The overall Markov process for CPT is constructed by specifying a probabilistic rule that decides whether the transition at each step is of the first or second kind. Since each kind of move leaves the required thermal distribution invariant, it is guaranteed that this thermal distribution is obtained as the limiting distribution.

9.2.2 A Consideration of the Number and Arrangement of Replicas in Parallel Tempering

Whilst a determination of the appropriate number of replicas that is optimal for a given system is non-trivial, it is straightforward to consider how the number of replicas we require scales with system size.

Starting with Eq. (9.26) and proceeding on a continuous state space, Kofke [190, 191] demonstrated that we can understand the acceptance probability of exchanges in terms of the distribution functions of both replicas. The resulting mean acceptance probability $\bar{p}$ of a swap is given by:

$$\bar{p} = 2 \int_{U_m}^{\infty} dU_0 \int_{U_m}^{U_0} dU_1 \, p^{\mathrm{st}}_{\beta_0}(U_0) p^{\mathrm{st}}_{\beta_1}(U_1) \tag{9.27}$$

where $\beta_1 < \beta_0$ and U_m denotes the minimum energy of the system. This means that $\bar{p}$ is governed explicitly by the overlap of the probability distributions at the two temperatures. Consequently we can understand how this quantity varies as the probability distributions widen and separate at different system sizes.

As we add terms to the Hamiltonian (i.e. increase the size of the system) the mean system energy increases linearly and as such we expect the mean energy difference between replicas to scale in the same way ($\propto N$). However, as we know from the central limit theorem (see Sect. 9.1.4), when performing a random walk in phase and thus energy space the size of the fluctuations are proportional to the square root of the number of random variables. In a simulation the number of random variables goes with the size of the system and so the fluctuations scale as $N^{1/2}$. Consequently as N increases, the mean distance in energy between replicas increases linearly, but the fluctuations which result in overlapping histograms increase only as $N^{1/2}$. This means that for a given β_0 and β_1, $\bar{p}$ decreases as $N^{-1/2}$. Consequently, if we wish to maintain a constant $\bar{p}$ between neighbouring replicas, the number of replicas should increase as $N^{1/2}$.

There has been considerable discussion in the literature about the best choice of replica temperatures when there are many replicas [190–195]. This is generally a difficult task although some valuable contributions have been offered. Kofke showed that the arrangement should be related to the entropy difference between the replicas which is just a reformulation of Eq. (9.27). Indeed this explains why parallel tempering often struggles with systems that undergo a phase change between the lowest and highest replica as there is a large entropy change for a small temperature gap. However, if we take a simpler approximation for more well behaved systems, an analytical solution can be found. There is a general consensus [194, 196, 197] that the performance of parallel tempering is maximised when the time taken for a round trip in temperature space is minimised. It has then been postulated [198] that this occurs when the acceptance probabilities are equal between all neighbouring replicas. Kofke [190, 191] showed that if the specific heat is independent of T, then the value of $\bar{p}$ is the same between all neighbouring replicas if the values of β_i form a geometric series. This is a specific consequence from finding that the acceptance probabilities follow an incomplete beta law [199]. This way of choosing β_i may not always be optimal owing to the breakdown of this assumption near critical points, but there is evidence that it often works well in practice [181, 182] and is a good starting point for our simple systems. Consequently we shall use this scheme when we compare the efficiency of CPT with our LPT schemes.

9.3 Statistical Errors

In order to assess the scaling of standard CPT techniques alongside our new local method we need some meaningful way of determining the computational effort required to reach some comparable goal. For

our purposes we choose that goal to be the calculation of a relevant physical quantity, that is one that requires sampling from separated metastable regions, to within a given statistical accuracy. To consider how best to construct such a scheme it is instructive to consider some of the well established theory concerning statistical errors for a correlated time series.

9.3.1 Time Series Averages as Random Variables

When we perform a simulation to calculate a variable A at equilibrium we are averaging over a sequence of instantaneous quantities A_i. Inherent in the simulation process, from either Markov Chains for MCMC or atomic trajectories for MD is the idea that each instantaneous quantity is close in phase space to the previous instance. Generally speaking these values are deemed to be correlated in time (where for MCMC each Monte Carlo update is treated like the tick of a clock). Whilst we know that in the infinite limit we can calculate the exact expectation value of A

$$\langle A \rangle = \lim_{\tau \to \infty} \frac{1}{\tau} \int_{t_0}^{t_0+\tau} dt\, A(t), \tag{9.28}$$

in computer simulation τ must be finite and so we can only produce an estimate of A over the sample time and as such consider the estimate

$$A_\tau = \frac{1}{\tau} \int_{t_0}^{t_0+\tau} dt\, A(t). \tag{9.29}$$

itself as a random variable. We mention that when implemented in a Markov Chain, explicitly in discrete time, these time integrals (and all those considered subsequently) necessarily become a summation of time steps or clock ticks.

9.3.2 Correlation Functions

Now, given an estimate for $\langle A \rangle$ we need to consider the reliability of this estimate. A simple way of measuring this is to determine whether all the samples used were from a narrow region of phase space which has similar instantaneous values of A or whether they were distributed throughout all the relevant phase space ensuring a better average. A simple mathematical tool to consider this is the auto-correlation function and is a measure of the degree of correlation of a time dependent variable $A(t)$ with itself given a time interval between samples. The auto-correlation function of A is described as

$$\zeta(t) = \frac{\langle \Delta A(t_0 + t) \Delta A(t_0) \rangle}{\langle \Delta A^2 \rangle}, \tag{9.30}$$

where ΔA is the fluctuating part of A

$$\Delta A = A - \langle A \rangle. \tag{9.31}$$

This means that $\langle \Delta A^2 \rangle$ is $\langle (A - \langle A \rangle)^2 \rangle = \langle A^2 \rangle - \langle A \rangle^2$ which is the variance.

If suitably equilibrated the quantity $\zeta(t)$ will be independent of the time origin t_0 allowing it to be calculated with one realisation by performing the averaging over many time origins. Alternatively one could perform the average over many (independently equilibrated) realisations using the same time origin.

Normalisation dictates that $\zeta(0) = 1$ and $|\zeta(t)| \leq 1$. This value will typically decay to zero. Considering the form of this function we can define a scalar quantity that characterises the time taken on

average for a trajectory $A(t)$ to decorrelate allowing us to consider the accuracy of our estimate A_τ. This quantity is the area bounded by the correlation function and is called the correlation time and is explicitly defined as

$$t_{corr} = \int_0^\infty dt\, |\zeta(t)|. \tag{9.32}$$

9.3.3 Variance of A

If we now consider the typical size of the fluctuations in A_τ we can understand the importance of t_{corr} and understand how to estimate the errors in A_τ. Considering the form of the auto-correlation function and a sampling time τ, the fluctuations expected in A_τ will typically depend on t_{corr}. For example, if $t_{corr} \gg \tau$ all samples used in the computation of A_τ will be strongly correlated meaning the size of the fluctuations in A_τ will be of the same size as for a single measurement in A. If, however, $t_{corr} \ll \tau$ then the samples used in determining A_τ will be largely uncorrelated meaning that the individual fluctuations in A tend to cancel out so that the size of fluctuations in A_τ will be much smaller than in the single measurement of A. Considering the fluctuations of A_τ we write our estimation A_τ as

$$A_\tau = \langle A \rangle + \frac{1}{\tau} \int_{t_0}^{t_0+\tau} dt\, \Delta A(t) \tag{9.33}$$

where

$$\Delta A_\tau = A_\tau - \langle A \rangle. \tag{9.34}$$

Choosing $t_0 = 0$ the variance is consequently defined as in [200]

$$\langle {\Delta A_\tau}^2 \rangle = \frac{1}{\tau^2} \int_0^\tau dt_1 \int_0^\tau dt_2\, \langle \Delta A(t_1) \Delta A(t_2) \rangle. \tag{9.35}$$

Writing this in terms of the auto-correlation function we have

$$\langle {\Delta A_\tau}^2 \rangle = \frac{\langle \Delta A^2 \rangle}{\tau^2} \int_0^\tau dt_1 \int_0^\tau dt_2\, \zeta(t_1 - t_2). \tag{9.36}$$

If $\tau \ll t_{corr}$ then the double integral becomes τ^2 and we have

$$\langle {\Delta A_\tau}^2 \rangle = \langle \Delta A^2 \rangle. \tag{9.37}$$

However, if $\tau \gg t_{corr}$ then

$$\int_0^\tau dt_1 \int_0^\tau dt_2\, \zeta(t_1 - t_2) \to 2 t_{corr} \tau \tag{9.38}$$

so

$$\langle {\Delta A_\tau}^2 \rangle = \frac{2 t_{corr}}{\tau} \langle \Delta A^2 \rangle. \tag{9.39}$$

Crucially we see that our ability to provide reliable estimates of $\langle A \rangle$ requires that $\tau \gg t_{corr}$. Once this condition has been reached one can then expect the error to decay as $\tau^{-\frac{1}{2}}$ and to be scaled according to $\langle \Delta A^2 \rangle^{\frac{1}{2}}$.

9.3.4 Computational Effort

We understand that if we require a fixed number of uncorrelated samples be taken from some Markov chain or trajectory then we find that the time taken to achieve a comparable result is in fact proportional to the correlation time. Indeed it is the ability to reduce this correlation time which gives the

parallel tempering algorithm its power. Consequently we shall use the correlation time as a key means of determining the computational effort. The correlation time has been written in terms of time t, however we consider a unit of time in MCMC calculations to be one Monte Carlo update. As such to obtain a complete computational cost we must also require a measure of the computational cost per Monte Carlo update. So we have the computational effort $= t_{corr} \times$ computational effort per time step.

It then only remains to specify the nature of the quantity for which the correlation time is measured. The exact quantity will be system dependent, but to be relevant it must be a quantity whose correlation function is generally slowly varying in the absence of transitions between metastable regions and significantly faster when transitions are introduced due to exchange moves when parallel tempering is implemented. If this condition is satisfied then differences in correlation time will reflect efficiencies in the parallel tempering algorithm as opposed to counting unnecessary additional work from higher replicas on top of normal decorrelation from standard MCMC steps.

Chapter 10

Local Parallel Tempering

10.1 Our Goals: Linear scaling

We intend to demonstrate with the use of toy models a method which can avoid the scaling associated with normal CPT. The way that this will be achieved is by limiting (and ideally making invariant with respect to system size) the number of replicas required to obtain worthwhile exchanges between sufficiently separated temperatures. By identifying that it is the exchange moves that are the origin of this poor scaling the obvious choice when proposing a new algorithm is to only exchange a limited portion of the system which does not grow as the total system is made larger. This means that as the system grows, more of these 'local' exchanges are required, however as the size of the portion exchanged is kept constant the cost of each exchange should not increase. If the number of 'local' exchanges required grows linearly with system size then we should expect an overall linear scaling.

Whilst this seems straightforward in principle there are several obstacles to achieving this aim. First and foremost we must limit the scope of our algorithm to systems that have well distributed local metastable regions. For example, performing local exchanges on a system that has to perform large global reconfigurations to reach very few relevant minima (e.g. the LJ_{38} cluster [201, 202]) would be highly ineffective. A physical example of a suitable system might be the arrangement of many atoms on a complex surface.

Further issues then arise regarding the specific algorithm of a local technique. A significant obstacle lies in devising a technique where meaningful local exchanges can occur between replicas without necessarily creating prohibitively large surface energies causing very low acceptance probabilities [203]. This arises from the fact that whilst two independent replicas will be exploring the same phase space globally, the phase space being explored locally may be very different. As a result the proposed new configurations are atypical and the exchange is rejected.

10.2 The 1D Chain of Double Wells

We now seek to describe a new method (LPT) which performs these well accepted parallel tempering techniques in a spatially local way such that linear scaling with system size can be achieved. We start with a consideration of a simple system which allows for a clear description of the method and how this scaling is achieved.

10.2.1 The System

Our first model consists of N particles, each of which is acted on by a symmetric double-well potential $V(x) = V_0((x/a)^2-1)^2$, so that the bottoms of the wells are at positions $\pm a$, and the height of the barrier separating the wells is V_0. The particles are labelled $n = 0, 1, \ldots N-1$, and there is a bi-linear coupling between neighbouring particles. Periodic boundary conditions are used, so that particle 0 interacts with particles $N-1$ and 1. The total potential energy U of the system is thus:

$$U = V_0 \sum_{n=0}^{N-1} ((x_n/a)^2 - 1)^2 - \xi \sum_{n=0}^{N-1} x_n x_{n+1} \ , \tag{10.1}$$

where x_n is the position of particle n, and we use the convention that $x_N \equiv x_0$. In the following, we shall use units such that $V_0 = a = 1$. This is a useful model for discussing accelerated sampling, because even in the apparently trivial case $\xi = 0$ the scaling of canonical parallel tempering with number of particles N is poor. Our task is to develop local PT techniques having linear-scaling properties for this model.

10.2.2 The Local Parallel Tempering Algorithm

The defining feature of our proposed algorithm involves the introduction of 'local replicas' for each individual particle in our physical system. The most significant departure from CPT that this takes is the idea that these higher temperature local replicas are formed of frozen configurations in the lowest replica as opposed to independent replicas in their own right. The consequence of this is that the higher replicas are in fact unphysical (in that when combined they do not describe the system at the higher temperature) and so we proceed by describing our algorithm as a series of single particle Markov transitions of particles in the lowest replica which we shall now refer to as the 'primary replica'. Consequently we now denote the positions of the particles in the primary replica by x_0^n $(n = 0, 1, \ldots N-1)$.

Considering the continuous nature of the system we are dealing with a probability density function p and the displacement probabilities of a given particle n are described by the Markov transition kernel, which we denote by

$$\gamma_n(\bar{x}_0^n | x_0^n; \{x_0^n\}_{\neq}) \ . \tag{10.2}$$

The meaning of this is that γ_n specifies the probability distribution of Markov moves of particle n from position x_0^n to $\bar{x}_0^n$, with all other particles held fixed. The notation $\{x_0^n\}_{\neq}$ indicates the set of all positions x_0^m for $m \neq n$. In more detail, if the present probability distribution of particle positions in the primary replica is $p(x_0^0, \ldots x_0^n, \ldots x_0^{N-1})$ then, after a Markov step involving particle n, the new distribution is

$$p'(x_0^0, \ldots \bar{x}_0^n, \ldots x_0^{N-1}) = \int dx_0^n \, \gamma_n(\bar{x}_0^n | x_0^n; \{x_0^n\}_{\neq}) p(x_0^0, \ldots x_0^n, \ldots x_0^{N-1}) \ . \tag{10.3}$$

Since γ_n behaves as a linear operator, we can write this more concisely as

$$p' = \hat{\gamma}_n p \ . \tag{10.4}$$

Note that the probability density $\gamma_n(\bar{x}_0^n | x_0^n; \{x_0^n\}_{\neq})$ depends on the positions x_0^m $(m \neq n)$; however, as indicated by the subscript n on γ and $\hat{\gamma}$, it is only particle n that makes the transition.

Since conventional Metropolis algorithms are often built from single-particle transitions γ_n, the methods we use strongly resemble those conventional algorithms. The major difference is in the kernel γ_n

itself. We will use PT operations involving local exchanges with suitable higher replicas to generate γ_n kernels that give much improved sampling efficiency. However, before specifying what these operations are, we need to comment on the properties of γ_n and how the overall Markov process is built from the individual γ_n kernels.

We must require that the thermal equilibrium distribution $p^{\rm st}(x_0^0, \ldots x_0^{N-1})$ at the temperature of interest is the limiting distribution of the overall process. We ensure this by requiring that $p^{\rm st}$ is invariant under the operation of the single particle moves:

$$\hat{\gamma}_n p^{\rm st} = p^{\rm st} \ , \ \forall \ n. \tag{10.5}$$

For any γ_n satisfying this invariance condition, there are many ways of building the overall Markov process. For example, at each Markov step we could choose a particle at random and perform a displacement γ_n for that particle. Another way would be to run through all the particles in a fixed sequence, performing a displacement γ_n for each one in turn. Some of these procedures do not satisfy overall detailed balance, but if they satisfy balance they will still yield the correct limiting distribution.

In order to exploit the nature of the chain of double wells we separate the particles into two groups. Assuming we have an even number of wells, N, we identify these groups as the even and odd labelled particles. The advantage of this distinction is that particles in the odd group will only ever interact with particles in the even group and vice versa owing to the nearest neighbour interaction. Consequently we choose to update all particles in first one group and then the other. We now have a degree of flexibility in the specific updating sequence of the particles in each group even allowing us to update them simultaneously.

We now turn to the construction of the single-particle transition kernel γ_n and the local replicas. For our 1D chain of double wells, this kernel refers to the transitions of the nth symmetric double well acted on by the field of its surroundings, which for the chain of double wells is simply its nearest neighbours. We construct γ_n with the help of a high temperature local replica of the nth symmetric double well to accelerate the transition. The local replica in question is defined to be a replica of the nth double well acted on by the remainder of the primary replica of which the relevant part is the nearest neighbours. The consequence of this is that both the primary replica and the local replica experience exactly the same external field. The kernel γ_n is then the net result of the following sequence of sub-steps:

- Create a new replica of the nth double well with its surroundings (nearest neighbours) in exactly the same position as in the primary replica. The position x_1^n of the particle in the new replica is initially identical to that in the primary replica, x_0^n. So initially we have $x_0^n = x_1^n$ and $x_1^{m\neq n} = x_0^{m\neq n}$
- With the remaining positions $x_1^{m\neq n} = x_0^{m\neq n}$ held fixed, perform a predetermined number P of conventional Metropolis Monte Carlo steps in x_1^n at temperature T_1 chosen high enough so that equilibration in the new replica is rapid.
- With the positions $x_1^{m\neq n} = x_0^{m\neq n}$ still fixed, we now perform a predetermined number Q of conventional parallel tempering moves on the replica positions x_0^n and x_1^n.

This entire sequence of sub-steps forms a probabilistic algorithm for going from an initial position x_0^n to a final position $\bar{x}_0^n$, with all other positions x_0^m $(m \neq n)$ held fixed. This algorithm specifies the transition kernel γ_n. So long as the overall Markov process is constructed from the single-particle transitions

γ_n so as to satisfy regularity, then it will generate a unique limiting distribution for any values of the parameters P and Q.

To ensure that the correct thermal-equilibrium distribution $p^{\rm st}(x_0^0, \ldots x_0^{N-1})$ remains invariant under the action of the γ_n, it is clearly essential to choose a large enough value of P. The purpose of the P Monte Carlo moves of the replica at temperature T_1 is to ensure that the probability distribution of positions x_1^n is the thermal-equilibrium distribution for the given fixed positions $x_0^m = x_1^m$ $(m \neq n)$. Provided this equilibration is achieved, then the subsequent Q steps of parallel tempering on the two replicas of particle n will not change the probability distribution of the primary replica, provided it has already attained the thermal-equilibrium distribution. We recognise, of course, that no finite value of P can yield the *exact* thermal distribution of the upper replica, but it will approach this exact distribution ever more closely as P is increased. This means that our algorithm is asymptotically exact in the limit $P \to \infty$.

The choice of the number Q of parallel-tempering steps is determined by the need for sampling. While the P equilibration steps are being performed, no sampling can be done. But as soon as the upper replica is well equilibrated and PT is started, positions of the primary replica can be used for sampling. Indeed, all Q steps of LPT can be used for sampling. The choice of Q is a balance between two factors: too small a value will result in inefficient local sampling, but too large a value may result in lengthening the timescale of configuration changes on longer length-scales.

10.2.3 Advantages of Using the Transition Kernel γ_n

The transition kernel that we implement, γ_n, consists of several parts, namely standard update moves, local exchanges and additional procedures which specify the arrangement of atoms in the higher replicas. Much of this is well established and requires no further discussion, however the exact method of our local exchanges given the choice of how to construct the upper replicas and how they preserve the correct limiting distribution in the primary replica warrants an additional explanation.

In order to understand the motivation for the type of local exchange moves implemented in our algorithms it is instructive to start with the most general concept of a local exchange. First we imagine that we have two fully independent replicas between which we wish to exchange only one part of the configurations. This portion of the system, which we label λ_0 and λ_1 for the lower and upper replicas respectively, can be defined in any way we see fit as long as following an exchange the number of particles in each replica is conserved. In defining λ_0 and λ_1 we can immediately define the remainder of the system using the state descriptors η_0 and η_1. Considering the total state in each replica at any given point we see that the composite equilibrium probability is now given as

$$p^{\rm st}(\lambda_0, \eta_0; \lambda_1, \eta_1) = p^{\rm st}_{\beta_0}(\lambda_0, \eta_0) p^{\rm st}_{\beta_1}(\lambda_1, \eta_1) \tag{10.6}$$

Now when we propose a local exchange we simply propose exchanging states λ_0 and λ_1. To ensure invariance in the equilibrium distributions, given again that proposal densities are state independent,

we then require

$$\frac{\gamma((\lambda_1,\eta_0;\lambda_0,\eta_1)|(\lambda_0,\eta_0;\lambda_1,\eta_1))}{\gamma((\lambda_0,\eta_0;\lambda_1,\eta_1)|(\lambda_1,\eta_0;\lambda_0,\eta_1))} = \frac{\alpha((\lambda_1,\eta_0;\lambda_0,\eta_1)|(\lambda_0,\eta_0;\lambda_1,\eta_1))}{\alpha((\lambda_0,\eta_0;\lambda_1,\eta_1)|(\lambda_1,\eta_0;\lambda_0,\eta_1))} = \frac{p^{\mathrm{st}}_{\beta_0}(\lambda_1,\eta_0)p^{\mathrm{st}}_{\beta_1}(\lambda_0,\eta_1)}{p^{\mathrm{st}}_{\beta_0}(\lambda_0,\eta_0)p^{\mathrm{st}}_{\beta_1}(\lambda_1,\eta_1)}$$
$$= \exp\left[\beta_0(U(\lambda_0,\eta_0) - U(\lambda_1,\eta_0)) + \beta_1(U(\lambda_1,\eta_1) - U(\lambda_0,\eta_1))\right]. \quad (10.7)$$

This leads to a general transition kernel which holds for exchanging any two comparable parts of the system without any further constraint

$$\alpha((\lambda_1,\eta_0;\lambda_0,\eta_1)|(\lambda_0,\eta_0;\lambda_1,\eta_1))$$
$$= \min\left[1, \exp(\beta_0(U(\lambda_0,\eta_0) - U(\lambda_1,\eta_0)) + \beta_1(U(\lambda_1,\eta_1) - U(\lambda_0,\eta_1)))\right]. \quad (10.8)$$

Examining the form of this equation we identify the two key terms $U(\lambda_1,\eta_0)$ and $U(\lambda_0,\eta_1)$ that will affect performance of any algorithm built out of local exchanges. We understand that generally these terms will dominate the expression owing to the small probability that a state being exchanged (e.g. λ_0) will form a typical total configuration considering the independent surroundings it will be paired with (η_1). In other words each replica explores two separate regions of phase space resulting in a very small overlap in the energy distributions. Typically this renders the contents of the exponential large and negative resulting in a small acceptance probability. This explicitly is the surface energy problem.

In an attempt to reduce this problem we are proposing a method which performs local exchanges between mutual local regions, the contents of which we denote by the labels λ_0 and λ_1. We then add the constraint that the remainder of the system must be identical between the replicas. As we require thermodynamic accuracy in the lowest replica this is given as $\eta_1 = \eta_0$. Consequently the transition kernel reduces to

$$\alpha((\lambda_1,\eta_0;\lambda_0,\eta_0)|(\lambda_0,\eta_0;\lambda_1,\eta_0)) = \min\left[1, \exp((\beta_0 - \beta_1)(U(\lambda_0,\eta_0) - U(\lambda_1,\eta_0)))\right]. \quad (10.9)$$

We now have a transition kernel in the form originally used in canonical parallel tempering that deals with spatially local exchanges. First by using these spatially local exchanges we are able to remove proposals that will certainly be rejected, that is local exchanges which propose exchanging particles very far from their original position into an unchanged surrounding. Critically, however by demanding that the surroundings of the local region is common between replicas we can ensure that the original terms $U(\lambda_0,\eta_1)$ and $U(\lambda_1,\eta_0)$ do not dominate as we are now insisting that each replica explores a common region of configurational space. An additional benefit of this method is that because the transition kernel now contains only the two terms associated with each replica no further potential calculations are required as would be needed for an unconstrained local exchange.

In practice this allows a well controlled method of local exchanges where the limiting factors on the acceptance rates are not surface energies between independent surroundings. However, we mention that this procedure is at the cost of physical higher replicas and that although both replicas explore the same configurational space a local exchange is still affected by the surrounding particles meaning that there will always be some cost for which there is no analogy in CPT. The ultimate performance of the algorithm then becomes whether two distinct, relevant metastable regions at comparable energies exist within a given local region under the influence of the surroundings. This however, is a practicality of the implementation of the method and we predict this issue could be tuned quite easily with simple

parameters such as the size of the local region.

10.3 1D System of Particles in Periodic Potential

10.3.1 The System

The second model used to illustrate local parallel tempering consists of N particles in one dimension, all acted on by a periodic potential $V(x) = -\frac{1}{2}V_0 \cos(2\pi x/a)$ having a repeat distance a and barrier height V_0. Periodic boundary conditions are applied, the number of periods of the potential in the unit cell being denoted by M. Positions x and $x + pL$, where $L = Ma$ and p is a positive, negative or zero integer, are periodic images of each other. A pair potential $\phi(x)$ also acts between the particles. We impose a cut-off x_c on this potential, so that $\phi(x) = 0$ for $|x| > x_c$, and we require that $x_c \leq \frac{1}{2}L$. This means that each particle interacts with only one image of any other particle. The total potential energy of the system is:

$$U = -\frac{1}{2}V_0 \sum_{n=0}^{N-1} \cos(2\pi x_n/a) + \frac{1}{2} \sum_{n \neq n'} \phi(|x_n - x_{n'}|) \; . \tag{10.10}$$

The major difference between this model and the 1D chain of double wells is that every particle is now free to move throughout the system, and is not tied to any local region. Our task is to develop linear-scaling PT techniques for this model.

10.3.2 The Local Parallel Tempering Algorithm

For the chain of double wells (Sect. 10.2), the overall Markov process was built from the elementary transition kernels γ_n for individual particles. We discussed there how to assemble the γ_n to form the overall process, and we then considered how to use LPT ideas to construct the γ_n. We shall follow a similar plan here. However, there is a crucial difference. For a system of unconfined particles, we think it best to work with an elementary transition kernel γ_ω that operates on chosen regions of space, rather than on individual particles.

Let ω be a local spatial region of our 1D system, consisting of a segment of length l centred at position x_s. (In practice, we shall choose l to be comparable with the repeat length a of the periodic potential.) Then γ_ω represents a probabilistic rule for displacing the positions of *all* the particles in region ω. (If there are no particles in ω, then γ_ω does nothing.) This rule may depend on the positions of particles both outside and inside ω, but it does not result in the displacement of any particles outside ω. We require that the thermal equilibrium distribution $p^{\rm st}(x_0^0, \ldots x_0^{N-1})$ is invariant under the action of $\hat{\gamma}_\omega$:

$$\hat{\gamma}_\omega p^{\rm st} = p^{\rm st} \; , \quad \forall \; \omega \; . \tag{10.11}$$

There are many ways of building the overall Markov process from the elementary transition kernels $\hat{\gamma}_\omega$, but the scheme we adopt here resembles the one we used for the chain of double wells. In order to replicate the properties found in the chain of double wells we divide our system into two sets of transition regions such that the centres of two regions in a given set are separated by at least twice the interaction cut-off distance x_c. This results in two interleaving sets of regions much like the odd and even regions in the chain of double wells. This has the desired result of guaranteeing that there is no interaction between two regions within a given set.

Given this set-up the LPT procedure that we use to form the γ_ω kernels is very similar to the one used for the chain of double wells. For a given initial set of positions $\{x_0^n\}$ and a chosen region ω, we perform the following sequence of sub-steps:

- Create a new replica of the particles in region ω, with the initial positions x_1^n of particles in ω being the same as in the primary replica, and the positions of all particles outside ω also being exactly the same as in the primary replica.
- With the positions of all particles outside ω held fixed, perform a predetermined number P of conventional Metropolis Monte Carlo steps on the positions x_1^n of the particles in ω at a temperature T_1 high enough so that equilibration in the new replica is rapid. Trial moves that take a particle outside ω are rejected, and the configuration is repeated.
- With the positions of all particles outside ω still fixed, perform a predetermined number Q of conventional parallel tempering moves on the replica positions x_0^n and x_1^n inside ω; as before, trial moves that take any position in either replica outside ω are rejected.

This entire set of sub-steps constitutes the kernel γ_ω. We emphasise that in order to perform meaningful exchange moves we must maintain the number of particles contained within each local region between replicas and so require that any proposed moves outside of the region must be rejected.

This kernel is then applied in the same fashion as for the chain of double wells, in that we implement γ_ω upon all regions in the first set (analogous to the odd set) and then all regions in the second set (analogous to the even set). As before because these regions are explicitly independent of each other the order in which we perform these updates is irrelevant and we are free to update them simultaneously.

Finally, we must allow for larger scale movement of particles throughout the system in order to reach the correct limiting distribution. Currently this is not possible whilst ensuring equal numbers of particles in common regions between replicas. To achieve this we then must constantly redefine the boundaries of the local regions throughout the simulation. This could be done in many ways depending on the system, however we choose to shift the boundaries of the local regions by one half of the length of a local region along the x axis. By then implementing this change after γ_ω has acted on all local regions in both sets we form a total probabilistic algorithm built from a elementary transition kernel that maintains balance which also allows particles to move throughout the entire system.

10.4 2D System of Particles in a Periodic Potential

10.4.1 The System

Our 2D model is a straightforward generalisation of the 1D system treated in Sect. 10.3. It consists of N particles all acted on by periodic potential $V(x,y) = -\frac{1}{2}V_0(\cos(2\pi x/a) + \cos(2\pi y/a))$, the periodicity being that of a square lattice of lattice parameter a, and the barrier height being V_0. Periodic boundary conditions are applied, the repeat distance $L = Ma$ (M = positive integer) being the same in the two Cartesian directions, so that position $(x + pL, y + qL)$ is equivalent to (x, y), with p and q positive, negative or zero integers. A pair potential $\phi(r)$ acts between particles, with cut-off distance r_c, so that $\phi(r) = 0$ for $r \equiv (x^2 + y^2)^{1/2} > r_c$. The total potential energy of the system is therefore:

$$U = -\frac{1}{2}V_0 \sum_{n=0}^{N-1} [\cos(2\pi x_n/a) + \cos(2\pi y_n/a)] + \frac{1}{2}\sum_{n\neq n'} \phi(|\mathbf{r}_n - \mathbf{r}_{n'}|) \,, \tag{10.12}$$

with $\mathbf{r}_n = (x_n, y_n)$ being the vector position of the nth particle.

Our LPT algorithm for this system is similar to that for the 1D system, the main difference being the arrangement of the local regions. Many workable schemes could be envisaged, but the one we adopted for numerical tests is a simple chequerboard scheme, with black and white squares being the analogue of the even and odd regions of our 1D model. We take the edge length of each black or white square to be twice a. As for the 1D system the single region transition kernel is performed on each local region in the first set followed by the second set. After updates of this form have been applied to all particles the boundaries of the local sets were shifted in both dimensions by a, half the length of a local region. We envisage that superior schemes could be devised, however we have found that this scheme works correctly and quite efficiently.

Chapter 11

Results

For both the chain of double wells and the 1D periodic potential we have performed extensive tests to illustrate that our local parallel tempering algorithm gives correct results in the limit of large P. We also provide results on the scaling performance of local parallel tempering compared with CPT for the same systems.

11.1 The Chain of Double Wells

11.1.1 Correct Results in the Limit of Large P

To illustrate convergence in the limit of large P we present results for coupling constant $\xi = 0.5$ and temperature of interest $T_0 = 0.5$. We report results for two quantities: first, the single-particle probability distribution $p(x)$ for finding any chosen particle at position x; and second, the 2-particle correlation function $f_m = \langle x_n x_{n+m} \rangle / \langle x^2 \rangle$. Here, $\langle x_n x_{n+m} \rangle$ is the static correlation function for the displacements of particles n and $n + m$ from the origin, and we normalise this by dividing by the mean square displacement $\langle x^2 \rangle$. Clearly, f_m depends only on the separation m and not on n. We have calculated $p(x)$ and f_m first with standard CPT, which following suitable tests of convergence and equilibration gives essentially exact results being a well established technique that leads to the stationary distribution, and then with LPT for different values of the equilibration parameter P. In the LPT calculations, we have deliberately slowed down the equilibration by choosing the maximum displacement step at the high temperature T_1 to have the small value 0.4, so that convergence with respect to P can be examined in detail. The parallel tempering parameter Q was given the small value $Q = 5$, so that we rely entirely on the preceding P steps to achieve equilibration.

Our tests (Figs. 11.1 and 11.2) show that $p(x)$ and f_m obtained by LPT converge systematically to the correct results as P is increased. With the present settings, essentially perfect agreement is obtained with $P = 150$ in both cases. As expected, we find that much faster convergence is obtained with a larger high-temperature step length. For example for a step length of 1.0 good results can be produced with an equilibration time of $P = 5$.

11.1.2 Scaling with System Size

We turn now to the question of the scaling of computational effort with system size. As discussed in Sect. 9.3.4 the meaning of this question is that we have to calculate a chosen physical quantity to within a

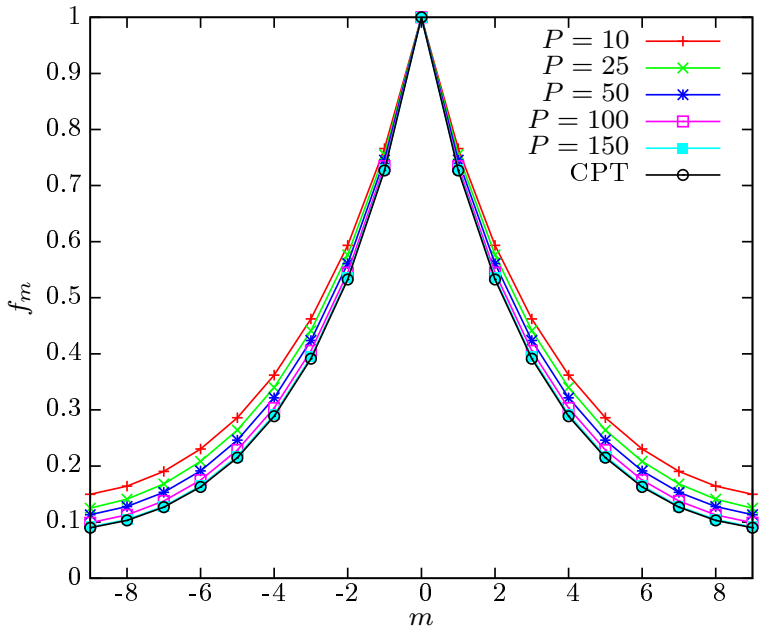

Figure 11.1: Correlation function f_m for different values of equilibration parameter P for the chain of double wells (number of wells $N = 20$, coupling constant $\xi = 0.5$, temperature $= 0.2$).

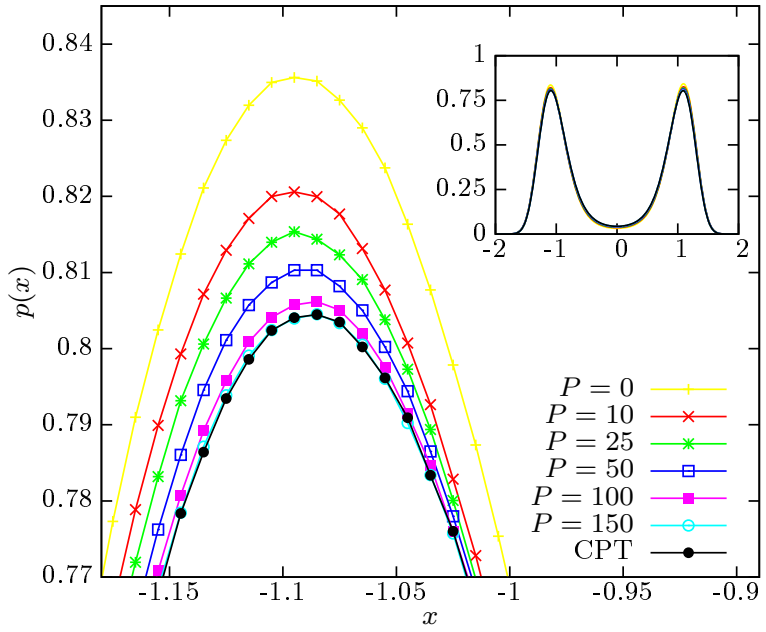

Figure 11.2: A peak of the probability distribution $p(x)$ for different values of equilibration parameter P for the chain of double wells (same parameters as in Fig. 11.1). Inset figure shows the total probability distribution.

specified statistical tolerance, using either CPT or LPT, and we ask for the computational effort needed to do this. We have previously identified that this will be the product of the correlation time and the computational effort per clock tick. We aim to use the approximation that the most computationally intensive procedure is a potential calculation considering the simple nature of the model. As shown in Sect. 10.2.3 no potential calculations are required to perform an exchange move, however they are

required to perform a standard Monte Carlo displacement step. Consequently we regard the computational effort per clock tick to be equal to the total number of displacement steps across all replicas and particles in one full sweep of the composite system, each of which is deemed equally computationally intensive. We then estimate the total computational effort by measuring the correlation time in terms of a number of these full sweeps or clock ticks. Practically, considering that we do not expect any anti-correlated variables in our systems, this is estimated by measuring the number of clock ticks taken for the autocorrelation to have decayed to a predetermined cut off of 0.01.

We note, however, that in the local parallel tempering algorithm the nature of one full Monte Carlo sweep is not exactly the same as for canonical parallel tempering. That is to say it is not absolutely clear how to compare equal clock ticks between CPT and LPT. To proceed we must find a way to compare an equal number of relevant displacement steps in the primary replica of both techniques. The key difference between the two algorithms is that in CPT all individual particles are updated in sequence, each with an analogous update in all higher replicas, whereas in LPT each particle in a given local region is updated a certain prescribed number of times, Q, with analogous updates in each higher replica, preceded by P equilibration updates in the higher replicas only, which is then applied to each local region in turn. As such it is instructive to illustrate how one would determine what would be considered an equal clock tick and how to consider its relative computational effort. We proceed by noting that a given clock tick is comparable when it amounts to the same number of individual updates in all particles in the primary replica. The way to achieve this is to compare the Q sampling steps applied to each local region in LPT to Q full sweeps of the entire system in CPT as both of these situations corresponds to Q total updates per particle in the primary replica. We then consider the computational effort, in individual particle displacements, of these two cases. In the CPT situation the total number of displacement steps required in a system of N particles and M replicas is $Q \times N \times M$. In LPT the equivalent number of displacement steps required for the same number of sampling moves in each particle in the primary replica is $N(MQ+(M-1)P)$. This is because to achieve Q sampling moves for the particles in one local replica we must simulate the $M-1$ higher replicas for the equilibration time P before performing the Q sampling moves in all M replicas. This must then be performed over all local regions or rather over all N particles. Each of these processes would then be a valid measure of a common clock tick each with a different computational cost. Since, however, the ability to compare one individual clock tick in this way depends on both N and Q such a clock tick may be too coarse a time unit practically. As such we can formulate an equivalent comparison by defining one clock tick to occur whenever a single particle in the primary replica is updated. To do so we consider the computational cost per step unity in CPT and $1+(P/Q)\times((M-1)/M)$ in LPT noting that the ratio is the same as for the quantities considered above.

With the computational effort defined we then seek a suitable quantity for which we consider the correlation time. For the chain of double wells the most simple quantity which satisfies the conditions set out in Sect. 9.3.4 is the mean position $\langle x \rangle$ of a given particle. The key question is now whether the "computational effort" scales more favourably with system size in LPT than in CPT.

We now present size scaling comparisons of CPT and LPT for the two example cases $\xi = 0$, $T_0 = 0.1$ and $\xi = 0.25$, $T_0 = 0.5$. The step length in the primary replicas of both CPT and LPT was set to 0.2, to ensure that barrier crossing relies on parallel tempering. In CPT, the temperatures of the higher replicas were set in a geometric sequence between T_0 and the uppermost temperature of 2.0 and the results shown, for all individual system sizes, use the number of higher replicas that resulted in the lowest total computational effort. LPT used only a single higher replica, whose temperature was $T_1 = 2.0$. For the

higher replicas, we used a step length of 1.0, to ensure fast equilibration. In the case of LPT, we used equilibration parameter $P = 5$ and sampling parameter $Q = 5$. (We checked that this value of P gives results in close agreement with CPT.)

Our scaling results (Figs. 11.3 and 11.4) show that the relative efficiency of LPT and CPT depends quite strongly on the value of ξ. For $\xi = 0$, we find that with LPT the computational effort scales linearly with N, but with CPT it increases more rapidly. For this ξ, LPT is always more efficient than CPT, except possibly for very small values of N. For the rather modest value $N = 20$, LPT is over five times more efficient. For the case of $\xi = 0.25$, the more favourable scaling of LPT is again very clear, but now LPT becomes more efficient than CPT only for $N \simeq 20$. Our further tests (not shown here) indicate that the competitive advantage of LPT becomes even less for higher ξ values, though its size scaling remains nearly linear. We will comment on this further in Sect. 12.

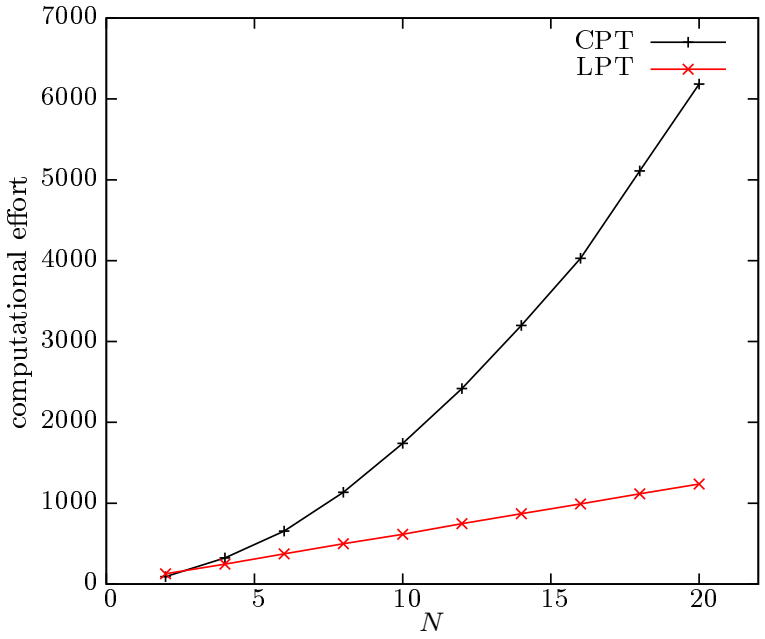

Figure 11.3: Computational effort (see text) for canonical and local parallel tempering applied to 1D chain of double wells as function of number N of double wells (temperature = 0.1, coupling constant $\xi = 0$).

11.2 1D System of Particles in a Periodic Potential

11.2.1 Correct Results in the Limit of Large P

We have performed tests to demonstrate that LPT yields correct thermal averages, as we did for the 1D chain of double wells. We present illustrative results for the case of the repulsive pair potential:

$$\begin{aligned} \phi(x) &= \frac{1}{2}(1-x^2)^2 \quad \text{for} \quad |x| \le 1 \\ &= 0 \qquad \text{for} \quad |x| > 1 \, . \end{aligned} \tag{11.1}$$

The values of the parameters of the periodic potential are chosen to be $a = 1$, $V_0 = 2.0$, and the calculations are for the temperature of interest $T_0 = 0.2$, with a mean particle density of one particle per

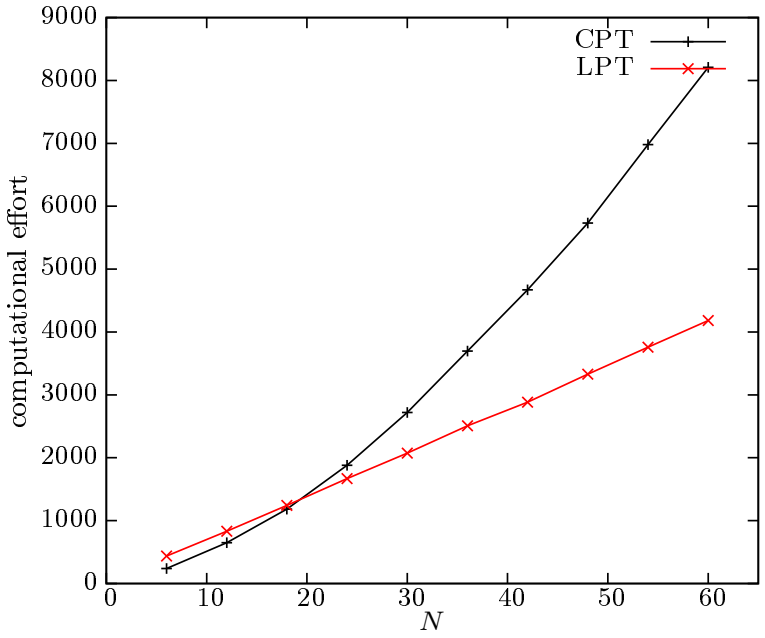

Figure 11.4: Computational effort (see text) for canonical and local parallel tempering applied to 1D chain of double wells as function of number N of double wells (temperature = 0.5, coupling constant $\xi = 0.25$).

periodic repeat a. To demonstrate the correctness of LPT, we compare the radial distribution function $g(x)$, calculated with a range of values of the equilibration parameter P, against essentially exact results from CPT. The sampling parameter Q took the fixed value of 5 and as before the maximum displacement was taken as a very low value, in this case set at 0.1, in order to slow down the equilibration for illustrative purposes. The radial distribution function, $g(r)$, is a pair correlation function and expresses the average density of particles found at a distance r from the particle in question (here all our particles are identical and thus share the same radial distribution function). It is defined as a ratio given by the mean number of particles measured in a region dr a distance r away from the particle in question written $n(r)$ divided by the number of particles found in an equivalent phase space volume given the overall density of particles in the system, $\bar{n} = \rho dV$, so that

$$g(r) = \frac{n(r)}{\bar{n}} = \frac{n(r)}{\rho dV}. \tag{11.2}$$

The specific form of dV would then depend on the dimensionality of the system. For example, dV would be dr, $2\pi r dr$ and $4\pi r^2 dr$ in one, two and three dimensional systems respectively.

The convergence of $g(r)$ as a function of P is reported in Fig. 11.5. The convergence is most easily examined in the region of one of the peaks of $g(r)$, and Fig. 11.5 shows the second peak in detail. We see that the convergence is simple and monotonic, and that almost perfect agreement is attained for $P = 100$. Our tests show that excellent agreement is found for much smaller P values if larger random displacements are used. We note the general form of the radial distribution function reflects the relative likelihood of observing two of the particles in the wells of the periodic potential. The peak at $r = 0$, however, is substantially lower and this reflects the lower probability of observing two particles in the same well because of the repulsive pair potential.

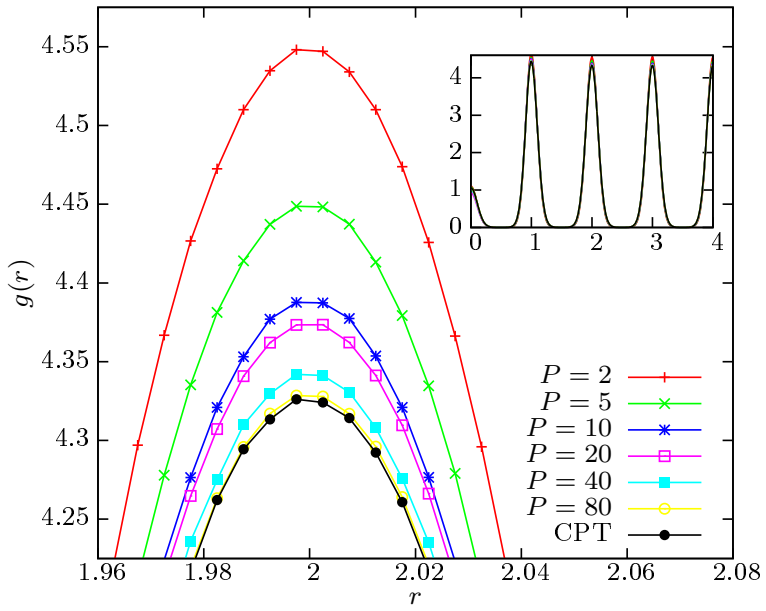

Figure 11.5: A peak of the radial distribution function $g(r)$ as function of equilibration parameter P in local parallel tempering for 1D periodic system of interacting particles. Inset shows the total radial distribution function.

11.2.2 Scaling with System Size

To examine the issue of size-scaling, we consider the computational effort needed to calculate a chosen physical quantity to a specified statistical accuracy with LPT and CPT, as we did in Sect. 11.1.2. We must use a physical quantity whose fluctuations are made slow by the infrequency of barrier crossings, and we choose the quantity:

$$s = \sum_{i=0}^{N-1} \sin(\pi x_i/a) \; . \tag{11.3}$$

Since this involves a wavelength that is twice that of the periodic potential, the average value $\langle s \rangle$ is zero, but the fluctuations about zero will be slow. The correlation time $t_{\rm corr}$ is found from the auto-correlation function:

$$t_{\rm corr} = \int_0^\infty dt \, \langle s(t_0 + t) s(t_0) \rangle / \langle s^2 \rangle \; , \tag{11.4}$$

by analogy with Eq. (9.32). As before, "time" here represents a number of Monte Carlo clock ticks with each tick deemed to occur with the update of a single particle in the primary replica. These ticks then have a different computational effort for LPT and CPT as described in Sect. 11.1.2. In comparing the scaling of computational effort with system size for LPT and CPT, we characterise system size by the number of particles N, whereas the mean density of particles N/L is held fixed.

We present the results of scaling tests for the temperature $T_0 = 0.2$. As before, with CPT we use replicas in a geometric sequence of temperatures with the uppermost temperature being 2.0 and report results for the number of replicas that minimises the computational effort. In contrast, in LPT we use only two replicas for each local region, the uppermost temperature also being 2.0. The step lengths utilised were 0.2 for the lowest replica for both CPT and LPT and 1.0 for all upper replicas. The LPT equilibration and sampling parameters were set to $P = 5$ and $Q = 10$, the adequacy of this P value was checked separately. Scaling comparisons are reported in Fig. 11.6. We see that the scaling of

computational effort with N is close to linear with LPT, but rises more rapidly with CPT, as expected. The cross-over occurs at $N \simeq 10$ in this particular case.

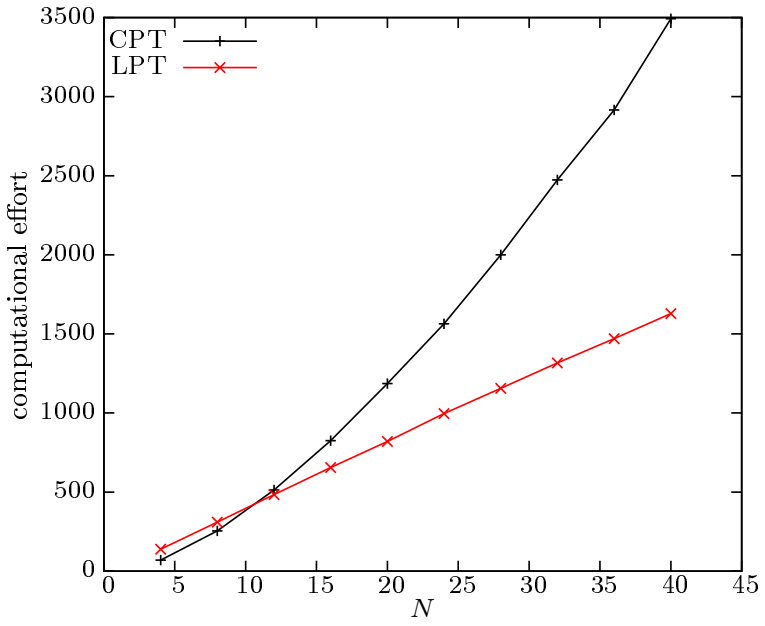

Figure 11.6: Computational effort (see text) with CPT and LPT as function of system size for the 1D periodic system of interacting particles.

11.3 2D System of Particles in a Periodic Potential

We have performed a number of tests to demonstrate that the results produced by the LPT are correct. To illustrate this, we show in Fig. 11.7 the radial distribution function $g(r)$ calculated both by CPT and by LPT for the case $V_0 = 1.0$ and $a = 1.0$ at the temperature $T_0 = 0.1$, the density of particles (average number of particles per potential minimum) being 0.5. Both P and Q were set to 10. The essentially exact agreement between results from the canonical and local methods demonstrates that our general LPT methods work without difficulty for 2D systems, and suggests that the same will be true in 3D.

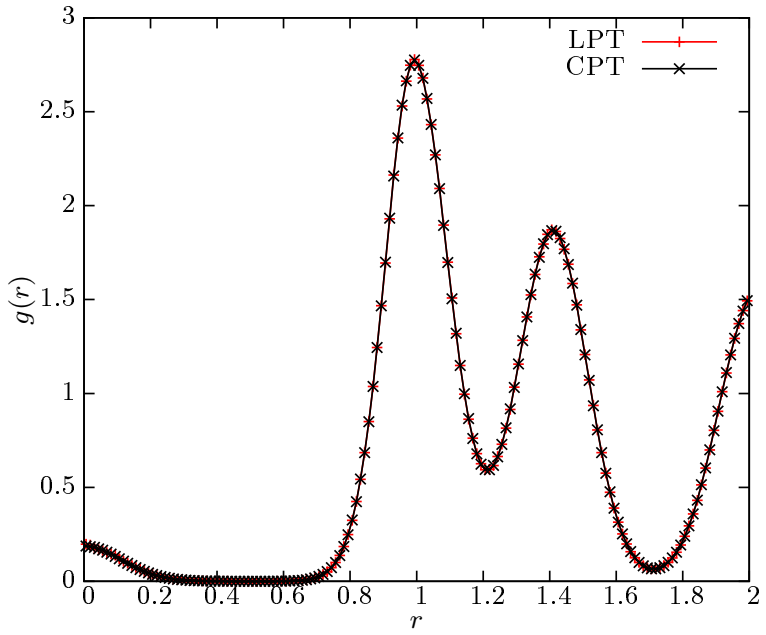

Figure 11.7: Radial distribution function $g(r)$ calculated using canonical and local parallel tempering for the 2D periodic system of interacting particles.

Chapter 12

Discussion and Conclusions

At the start of this section, we outlined the idea of a"linear-scaling" thermal sampling algorithm, for which the computational effort for equilibration and sampling is proportional to the size of the system (the number of atoms, or degrees of freedom). We pointed out that normal, or "canonical" parallel tempering algorithms do not have this property, and we asked whether it is possible to construct local parallel tempering (LPT) algorithms that do have linear-scaling performance. We have shown that this is indeed possible, at least for the simple model systems we have examined. For the 1D linear chain of coupled double wells, and for 1D and 2D systems of interacting particles acted on by periodic potentials, we have described LPT techniques that reproduce the exact results to any required accuracy, and we have shown that the algorithms exhibit practical linear scaling.

Our LPT algorithms use the ideas of normal parallel tempering, but their key feature is that the swaps of configurations are performed only in local regions. To ensure that attempted local swaps have a reasonable probability of being accepted, we require that the configurations of replicas involved in local swaps differ only in the swap region, but are identical outside this region. In order to achieve this, we pay a price: every time the local regions are changed, we need to create new high-temperature replicas, and these new replicas need to be equilibrated before any sampling is done.

The need to perform repeated equilibrations throughout the simulation is a highly unusual feature of our LPT algorithms. In all Monte Carlo sampling methods that we are aware of, the system is first equilibrated, and sampling is performed thereafter. However, in our techniques, periods of equilibration and sampling are interleaved. It might be thought that the need for repeated equilibration would necessarily make our LPT algorithms very inefficient. However, we have shown that this is not the case. Furthermore, the linear-scaling property of the algorithms guarantees that they become more efficient than normal parallel tempering for large enough systems. As we have seen, the crossover occurs at rather moderate particle numbers of a few tens at most.

An important question that we have not addressed concerns the best choice of local region. In our study of the 1D chain of double wells we took the local region to consist of one double well. We saw that the competitive advantage of LPT against CPT is best when there is no coupling between double wells and deteriorates as the coupling increases. This is not surprising because with high coupling equilibration involves collective motion of correlated double wells. This suggests that it would be better to choose the local region for LPT as a group of double wells, since this will facilitate collective equilibration. Similar considerations probably apply to our systems of diffusing particles.

In conclusion, we have shown that it is possible to construct local parallel tempering techniques, for which the computational effort needed for thermal equilibration and sampling is proportional to the size of the system. For simple models, we have shown that the new techniques give correct results and are more efficient than standard parallel tempering for large systems.

Appendix A

Numerical Solution to Stochastic Differential Equations

Approximation Procedure

Despite the existence of some analytical results, explicit expressions for many quantities, in particular the distribution of thermodynamic quantities, are very hard if not impossible to determine. Consequently we must investigate the behaviour of these quantities numerically in addition to the need to confirm the analytically derived results. The quantities are all intrinsically based on the underlying stochastic differential equation and its solution which written in the form of a general Itō SDE is

$$dx = a(x,t)dt + b(x,t)dW \tag{A.1}$$

where W is the Wiener process. This is however just shorthand for the integral equation

$$x(t) - x(t_0) = \int_{t_0}^{t} a(x,t')dt' + \int_{t_0}^{t} b(x,t')dW \tag{A.2}$$

which under the Itō integration convention can be written as the limit of the sum

$$x(t) - x(t_0) = \lim_{n\to\infty} \sum_{i=0}^{n-1} a(x_i,t_i)(t_{i+1} - t_i) + \sum_{i=0}^{n-1} b(x_i,t_i)(W_{i+1} - W_i). \tag{A.3}$$

If we take the time $t - t_0$ in the original integral equation to be small, such that $t - t_0 \to \Delta t$ we can rewrite this as

$$x(t + \Delta t) - x(t) = a(x,t)\Delta t + b(x,t)\Delta W. \tag{A.4}$$

where $\Delta t = t_{i+1} - t_i$ and ΔW is an increment of the Wiener process over the time Δt. Consequently, because of the Itō construction we are able to construct a stochastic forward Euler solution for x known formally as an Euler Maruyama scheme [204]

$$x(t + \Delta t) = x(t) + a(x,t)\Delta t + b(x,t)\Delta W \tag{A.5}$$

because $b(x,t)$ is statistically independent of the increment of the Wiener process $\Delta W = W_{i+1} - W_i$. We can now use this formalism to calculate solutions to the stochastic differential equation choosing Δt small enough to accurately represent the integral in Eq. (A.2) by the sum in Eq. (A.3) and by drawing

the stochastic variable ΔW from a suitable distribution using a pseudo random number generator.

One can also implement higher order schemes, with their importance arising typically when the SDEs are multiplicative. A common higher order scheme is known as the Milstein method [204] and is summarised as

$$x(t+\Delta t) = x(t) + a(x,t)\Delta t + b(x,t)\Delta W + \frac{1}{2}\frac{\partial b(x,t)}{\partial x} b(x,t)\left((\Delta W)^2 - \Delta t\right). \tag{A.6}$$

Despite the small time step Δt approximation to the integrals in Eq. (A.2) we must understand that the increment in the Wiener process is the result of an infinite number of vanishingly small contributions independent of how small we choose Δt such that the increment in the Wiener process is always described by its statistical properties. These properties are simple consequences of the central limit theorem and state that the distribution in the increment of the Wiener process is Gaussian with mean zero and variance equal to $\Delta t^{1/2}$. This then finally specifies the complete numerical algorithm

$$x(t+\Delta t) = x(t) + a(x,t)\Delta t + b(x,t)\Gamma + \frac{1}{2}\frac{\partial b(x,t)}{\partial x} b(x,t)\left(\Gamma^2 - \Delta t\right) \tag{A.7}$$

where

$$p(\Gamma) = N(0, \Delta t^{\frac{1}{2}}). \tag{A.8}$$

Box-Muller Algorithm

In order to realise the algorithm of Eq. (A.7) we must be able to draw random numbers from a normal distribution. The majority of pseudo random number generators provide random numbers with a uniform distribution and so a transforming algorithm is required. The most common is known as the Box-Muller algorithm [205]. The basic principle behind the algorithm is the fundamental transformation law of probabilities such that if we have a random variable a distributed by $p(a)$, the function $b(a)$, distributed by $p(b)$ is given by

$$|p(b)db| = |p(a)da|. \tag{A.9}$$

This then allows us to construct the desired distribution given uniform distribution of a ($p(a) = 1$) by

$$p(b) = \left|\frac{da}{db}\right| \tag{A.10}$$

with solution

$$a = \int p(b)db = F(b) \tag{A.11}$$

which relates the uniformly distributed a with the cumulative probability distribution of b. Considering now that we wish to draw two normally distributed random numbers such that

$$p(x) = \sqrt{\frac{1}{2\pi}} e^{-\frac{x^2}{2}} \tag{A.12}$$

and

$$p(y) = \sqrt{\frac{1}{2\pi}} e^{-\frac{y^2}{2}} \tag{A.13}$$

then the joint distribution is

$$p(x,y) = \frac{1}{2\pi} e^{-\frac{x^2+y^2}{2}} \tag{A.14}$$

such that we can generate x and y from a polar representation $x = R\cos(\Theta)$ and $y = R\sin(\Theta)$ where we randomly generate R on the interval $[0, \infty]$ and Θ on the interval $[0, 2\pi]$. The equivalent distribution

would then be

$$p(R,\Theta) = \frac{1}{2\pi} e^{-\frac{R^2}{2}}. \tag{A.15}$$

We can then relate uniformly distributed variables z_1 and z_2 to R and Θ using Eq. (A.11). The cumulative distribution function with respect to Θ is

$$F(\Theta) = p(\theta < \Theta) = \int_0^{\infty} \int_0^{\Theta} \frac{1}{2\pi} e^{-\frac{R^2}{2}} R dR d\theta = \int_0^{\Theta} \frac{1}{2\pi} = \frac{\Theta}{2\pi} \tag{A.16}$$

Comparison again with Eq. (A.11), using a uniform random number z_1 drawn from $[0,1]$, gives

$$\Theta = 2\pi z_1. \tag{A.17}$$

Similarly to calculate the cumulative distribution function of R we find

$$F(R) = p(r < R) = \int_0^{R} \int_0^{2\pi} \frac{1}{2\pi} e^{-\frac{r^2}{2}} r dr d\Theta = \int_0^{R} e^{-\frac{r^2}{2}} r dr \tag{A.18}$$

which after a change of variable is

$$p(r < R) = \left[-e^{-s}\right]_0^{\frac{R^2}{2}} = 1 - e^{-\frac{R^2}{2}} \tag{A.19}$$

allowing us to write R in terms of a new uniformly distributed random variable z_2

$$1 - z_2 = 1 - e^{-\frac{R^2}{2}} \tag{A.20}$$

because both z_2 and $1 - z_2$ are uniformly distributed giving solution

$$R = \sqrt{-2 \ln z_2}. \tag{A.21}$$

After substitution we then obtain two normally distributed random numbers

$$x = \sqrt{-2 \ln z_2} \cos(2\pi z_1) \tag{A.22}$$

$$y = \sqrt{-2 \ln z_2} \sin(2\pi z_1). \tag{A.23}$$

This method although simple can be quite inefficient owing to the use of trigonometric functions in their calculation. A more efficient method is known as the polar Box-Muller transform and uses two uniformly randomly distributed numbers w_1 and w_2 on the interval $[-1,1]$ to generate random numbers uniformly within circle of radius 1 by defining

$$p = w_1^2 + w_2^2. \tag{A.24}$$

This is achieved by discarding any combination w_1 and w_2 where $p = 0$ or $p \geq 1$ resulting in p being uniformly distributed between $[0,1]$. Considering this uniform circle in polar form we identify $r = \sqrt{p}$ and θ for which $\theta/2\pi$ is uniformly distributed between $[0,1]$. Since p is uniformly distributed we can identify it with the uniform distribution of the cumulative distribution function in Eq. (A.20) so that

$$p = z_2 \tag{A.25}$$

and the uniform variable z_1 with the new uniform variable $\theta/2\pi$

$$\frac{\theta}{2\pi} = z_1 \tag{A.26}$$

so that

$$x = \sqrt{-2\ln p}\cos(\theta) \tag{A.27}$$

and

$$y = \sqrt{-2\ln p}\sin(\theta). \tag{A.28}$$

However we now identify from the polar representation that $\cos(\theta) = w_1/\sqrt{p}$ and $\sin(\theta) = w_2/\sqrt{p}$ finally giving

$$x = w_1\sqrt{\frac{-2\ln p}{p}} \tag{A.29}$$

and

$$y = w_2\sqrt{\frac{-2\ln p}{p}}. \tag{A.30}$$

Given a variance σ and a mean μ this can then be transformed to a normal distribution $N(\mu, \sigma)$ by using $\sigma x + \mu$. Its implementation as a function in C++ is given in Fig. A.1 where the use of static variables exploits the fact that two normally distributed random numbers are generated with each implementation.

```
double gauss_rand(double mu, double sigma){ // box muler transform (radial)

  double x,y,z1,R,r,a;
  static double z2 = 0.0; // BM transform makes 2 random numbers z1 and z2
  static int cached = 0;  // so "remember" z2 by making it static


  if (!cached){ // If there isn't a random number saved from last time
                // generate two using the algorithm

      do{
        x = 2*ran2(&iseed)-1;  //produce numbers in box from -1 to +1
        y = 2*ran2(&iseed)-1;

        R=x*x + y*y;

      }while((R>=1.0)||(R==0.0)); //only take numbers from in the circle
                                  //we don't want R=0 --> log(0) undefined

      a=sqrt(-2*log(R)/R);

      z1=x*a;  //the random variables are z1=x*sqrt(-2ln(R)/R) & z2=y*sqrt(-2ln(R)/R)
      z2=y*a;

      cached = 1;

      return z1*sigma +mu; //multiply by stddev and add mean --> value is returned

    }
  else{  // if two numbers were produced previously we don't need to generate new ones

      cached = 0;
      return z2*sigma + mu;

    }
}
```

Figure A.1: C++ code for a function which performs the polar Box-Muller transform.

Bibliography

[1] R. Balescu, *Statistical Dynamics: Matter out of Equilibrium* (Imperial College Press, London, 1997).

[2] D. J. Evans, E. G. D. Cohen, and G. P. Morriss, Phys. Rev. Lett. **71**, 2401 (1993).

[3] D. J. Evans and D. J. Searles, Phys. Rev. E **50**, 1645 (1994).

[4] D. J. Evans and D. J. Searles, Phys. Rev. E **52**, 5839 (1995).

[5] D. J. Evans and D. J. Searles, Phys. Rev. E **53**, 5808 (1996).

[6] D. J. Evans and D. J. Searles, Adv. Phys. **51**, 1529 (2002).

[7] G. E. Crooks, J. Stat. Phys. **90**, 1481 (1998).

[8] G. E. Crooks, Phys. Rev. E **60**, 2721 (1999).

[9] G. E. Crooks, Phys. Rev. E **61**, 2361 (2000).

[10] C. Jarzynski, Phys. Rev. Lett. **78**, 2690 (1997).

[11] C. Jarzynski, J. Stat. Mech. P09005 (2004).

[12] U. Seifert, Phys. Rev. Lett. **95**, 040602 (2005).

[13] E. Sevick, R. Prabhakar, S. R. Williams, and D. J. Searles, Ann. Rev. Phys. Chem. **59**, 603 (2008).

[14] G. M. Wang, E. M. Sevick, E. Mittag, D. J. Searles, and D. J. Evans, Phys. Rev. Lett. **89**, 050601 (2002).

[15] C. Bustamante, J. Liphardt, and F. Ritort, Physics Today **58**, 43 (2005).

[16] C. Jarzynski, Eur. Phys. J. B **64**, 331 (2008).

[17] C. Jarzynski, Ann. Rev. Cond. Mat. Phys. **2**, 329 (2011).

[18] C. Van den Broeck, J. Stat. Mech. P10009 (2010).

[19] H. Qian, J. Phys. Chem. B **110**, 15063 (2006).

[20] J. Kurchan, J. Stat. Mech. P07005 (2007).

[21] S. Ciliberto, S. Joubaud, and A. Petrosyan, J. Stat. Mech. P12003 (2010).

[22] L. P. A. Imparato, C. R. Physique **8**, 556 (2007).

[23] F. Ritort, Seminare Poincaré **2**, 195 (2003).

[24] F. Ritort, Adv. Chem. Phys. **137**, 31 (2008).

[25] R. J. Harris and G. M. Schütz, J. Stat. Mech. P07020 (2007).

[26] C. Jarzynski, J. Stat. Phys. **98**, 77 (2000).

[27] J. Kurchan, J. Phys. A: Math. Gen. **31**, 3719 (1998).

[28] J. L. Lebowitz and H. Spohn, J. Stat. Phys. **95**, 333 (1999).

[29] M. Esposito, U. Harbola, and S. Mukamel, Rev. Mod. Phys. **81**, 1665 (2009).

[30] M. Campisi, P. Hänggi, and P. Talkner, Rev. Mod. Phys. **83**, 771 (2011).

[31] G. Gallavotti and E. G. D. Cohen, Phys. Rev. Lett. **74**, 2694 (1995).

[32] G. Gallavotti and E. G. D. Cohen, J. Stat. Phys. **80**, 931 (1995).

[33] U. Seifert, Eur. Phys. J. B **64**, 423 (2008).

[34] K. Sekimoto, Prog. Theor. Phys. Suppl. **130**, 17 (1998).

[35] Y. Oono and M. Paniconi, Prog. Theor. Phys. Suppl. **130**, 29 (1998).

[36] M. Esposito and C. Van den Broeck, Phys. Rev. Lett. **104**, 090601 (2010).

[37] R. F. Pawula, Phys. Rev. **162**, 186 (1967).

[38] C. Gardiner, *Stochastic Methods: A Handbook for the Natural and Social Sciences* (Springer, 2009).

[39] P. Langevin, C. R. Acad. Sci. (Paris) **146**, 530 (1908).

[40] N. van Kampen, *Stochastic Processes in Physics and Chemistry* (North-Holland, 2007).

[41] T. C. Gard, *Introduction to Stochastic Differential Equations*, 1st ed. (Marcel Dekker, New York, 1988).

[42] J. L. Doob, Ann. Math. **43**, 351 (1942).

[43] E. Wong and M. Zakai, Ann. Math. Stat. **36**, 1560 (1965).

[44] R. Kupferman, G. A. Pavliotis, and A. M. Stuart, Phys. Rev. E **70**, 036120 (2004).

[45] K. Sekimoto, *Stochastic Energetics*, Lecture Notes in Physics Vol. 799 (Springer, Berlin Heidelberg, 2010).

[46] H. Risken, *The Fokker-Planck equation: methods of solution and applications, 2nd edition* (Springer, 1989).

[47] C. Wissel, Z. Physik B **35**, 185 (1979).

[48] L. Onsager and S. Machlup, Phys. Rev. **91**, 1505 (1953).

[49] C. Maes and K. Netočný, J. Stat. Phys. **110**, 269 (2003).

[50] R. Chetrite and K. Gawędzki, Comm. Math. Phys. **282**, 469 (2008).

[51] H. Ge and D.-Q. Jiang, J. Stat. Phys. **131**, 675 (2008).

[52] F. Liu and Z. Ou-Yang, Phys. Rev. E **79**, 060107 (2009).

[53] Y. Sughiyama and M. Ohzeki, Physica E **43**, 790 (2011).

[54] F. Liu, Y. Luo, M. Huang, and Z. Ou-Yang, J. Phys. A **42**, 332003 (2009).

[55] R. García-García, D. Domínguez, V. Lecomte, and A. B. Kolton, Phys. Rev. E **82**, 030104 (2010).

[56] R. García-García, V. Lecomte, A. B. Kolton, and D. Domínguez, J. Stat. Mech. P02009 (2012).

[57] T. Schmiedl, T. Speck, and U. Seifert, J. Stat. Phys. **128**, 77 (2007).

[58] J. Schnakenberg, Rev. Mod. Phys. **48**, 571 (1976).

[59] L. Jiu-li, C. Van den Broeck, and G. Nicolis, Z. Physik B **56**, 165 (1984).

[60] C. Y. Mou, J. li Luo, and G. Nicolis, J. Chem. Phys. **84**, 7011 (1986).

[61] P. Gaspard, J. Stat. Phys. **117**, 599 (2004).

[62] R. Kawai, J. M. R. Parrondo, and C. V. den Broeck, Phys. Rev. Lett. **98**, 080602 (2007).

[63] A. Gomez-Marin, J. M. R. Parrondo, and C. V. den Broeck, Eur. Phys. Lett. **82**, 50002 (2008).

[64] J. M. R. Parrondo, C. V. den Broeck, and R. Kawai, New Journal of Physics **11**, 073008 (2009).

[65] E. H. Feng and G. E. Crooks, Phys. Rev. Lett. **101**, 090602 (2008).

[66] E. H. Feng and G. E. Crooks, Phys. Rev. E **79**, 012104 (2009).

[67] C. Maes, Seminare Poincaré **2**, 29 (2003).

[68] J. Farago, J. Stat. Phys. **107**, 781 (2002).

[69] J. Farago, Physica A **331**, 69 (2004).

[70] R. van Zon and E. G. D. Cohen, Phys. Rev. Lett. **91**, 110601 (2003).

[71] R. van Zon and E. G. D. Cohen, Phys. Rev. E **69**, 056121 (2004).

[72] M. Baiesi, T. Jacobs, C. Maes, and N. S. Skantzos, Phys. Rev. E **74**, 021111 (2006).

[73] H. Qian, Phys. Rev. E **65**, 016102 (2001).

[74] K. Sekimoto, J. Phys. Soc. Jpn. **66**, 1234 (1997).

[75] L. Y. Chen, J. Chem. Phys. **129**, 144113 (2008).

[76] L. Y. Chen, J. Chem. Phys. **129**, 091101 (2008).

[77] A. B. Adib, J. Chem. Phys. **130**, 247101 (2009).

[78] G. E. Crooks, J. Chem. Phys. **130**, 107101 (2009).

[79] L. Y. Chen, J. Chem. Phys. **130**, 107102 (2009).

[80] A. Imparato and L. Peliti, Phys. Rev. E **74**, 026106 (2006).

[81] C. Jarzynski, Phys. Rev. E **56**, 5018 (1997).

[82] J. Horowitz and C. Jarzynski, J. Stat. Mech. P11002 (2007).

[83] C. Jarzynski, C. R. Physique **8**, 495 (2007).

[84] G. Bochkov and Y. Kuzovlev, Physica A **106**, 443 (1981).

[85] G. Bochkov and Y. Kuzovlev, Physica A **106**, 480 (1981).

[86] J. M. G. Vilar and J. M. Rubi, Phys. Rev. Lett. **100**, 020601 (2008).

[87] L. Peliti, Phys. Rev. Lett. **101**, 098903 (2008).

[88] J. Horowitz and C. Jarzynski, Phys. Rev. Lett. **101**, 098901 (2008).

[89] L. Peliti, J. Stat. Mech. P05002 (2008).

[90] F. Zamponi, F. Bonetto, L. F. Cugliandolo, and J. Kurchan, J. Stat. Mech. P09013 (2005).

[91] D. J. Searles and D. J. Evans, J. Chem. Phys. **112**, 9727 (2000).

[92] R. Zwanzig, *Nonequilibrium Statistical Mechanics* (Oxford University Press, New York, 2001).

[93] T. Hatano and S. Sasa, Phys. Rev. Lett. **86**, 3463 (2001).

[94] T. Speck and U. Seifert, J. Phys. A: Math. Gen. **38**, L581 (2005).

[95] M. Esposito, U. Harbola, and S. Mukamel, Phys. Rev. E **76**, 031132 (2007).

[96] H. Ge, Phys. Rev. E **80**, 021137 (2009).

[97] H. Ge and H. Qian, Phys. Rev. E **81**, 051133 (2010).

[98] C. Van den Broeck and M. Esposito, Phys. Rev. E **82**, 011144 (2010).

[99] M. Esposito and C. Van den Broeck, Phys. Rev. E **82**, 011143 (2010).

[100] K. H. Kim and H. Qian, Phys. Rev. Lett. **93**, 120602 (2004).

[101] K. H. Kim and H. Qian, Phys. Rev. E **75**, 022102 (2007).

[102] T. Sagawa and M. Ueda, Phys. Rev. Lett. **104**, 090602 (2010).

[103] T. Sagawa and M. Ueda, Phys. Rev. E **85**, 021104 (2012).

[104] D. Abreu and U. Seifert, Phys. Rev. Lett. **108**, 030601 (2012).

[105] J. M. Horowitz and S. Vaikuntanathan, Phys. Rev. E **82**, 061120 (2010).

[106] M. Ponmurugan, Phys. Rev. E **82**, 031129 (2010).

[107] S. Lahiri, S. Rana, and A. M. Jayannavar, J. Phys. A **45**, 065002 (2012).

[108] M. Bauer, D. Abreu, and U. Seifert, J. Phys. A **45**, 162001 (2012).

[109] F. J. Cao and M. Feito, Phys. Rev. E **79**, 041118 (2009).

[110] F. Liu and H. Lei, J. Phys. A **45**, 125004 (2012).

[111] V. Y. Chernyak, M. Chertkov, and C. Jarzynski, J. Stat. Mech. P08001 (2006).

[112] T. Hatano, Phys. Rev. E **60**, R5017 (1999).

[113] C. Kwon, J. D. Noh, and H. Park, Phys. Rev. E **83**, 061145 (2011).

[114] B. H. Shargel, J. Phys. A: Math. Gen. **43**, 135002 (2010).

[115] T. Tomé and M. J. de Oliveira, Phys. Rev. E **82**, 021120 (2010).

[116] D. Daems and G. Nicolis, Phys. Rev. E **59**, 4000 (1999).

[117] W. Ebeling, Cond. Mat. Phys. **7**, 539 (2004).

[118] B. Lindner, New J. Phys. **9**, 136 (2007).

[119] S. Toyabe, T. Sagawa, M. Ueda, E. Muneyuki, and M. Sano, Nat. Phys. **6**, 988 (2010).

[120] O. Mazonka and C. Jarzynski, cond-mat/9912121 (1999).

[121] R. van Zon and E. G. D. Cohen, Phys. Rev. E **67**, 046102 (2003).

[122] A. Imparato, L. Peliti, G. Pesce, G. Rusciano, and A. Sasso, Phys. Rev. E **76**, 050101 (2007).

[123] A. Saha, S. Lahiri, and A. M. Jayannavar, Phys. Rev. E **80**, 011117 (2009).

[124] D. Chatterjee and B. J. Cherayil, Phys. Rev. E **82**, 051104 (2010).

[125] E. H. Trepagnier, C. Jarzynski, F. Ritort, G. E. Crooks, C. J. Bustamante, and J. Liphardt, Proc. Natl. Acad. Sci. U.S.A. **101**, 15038 (2004).

[126] A. M. Jayannavar and M. Sahoo, Phys. Rev. E **75**, 032102 (2007).

[127] J. I. Jiménez-Aquino, R. M. Velasco, and F. J. Uribe, Phys. Rev. E **79**, 061109 (2009).

[128] J. I. Jiménez-Aquino, F. J. Uribe, and R. M. Velasco, J. Phys. A **43**, 255001 (2010).

[129] J. I. Jiménez-Aquino, Phys. Rev. E **82**, 051118 (2010).

[130] J. I. Jiménez-Aquino, J. Phys. A **44**, 295002 (2011).

[131] A. Saha and A. M. Jayannavar, Phys. Rev. E **77**, 022105 (2008).

[132] D. Chatterjee and B. J. Cherayil, J. Stat. Mech. P03010 (2011).

[133] R. van Zon, S. Ciliberto, and E. G. D. Cohen, Phys. Rev. Lett. **92**, 130601 (2004).

[134] N. Garnier and S. Ciliberto, Phys. Rev. E **71**, 060101 (2005).

[135] S. Joubaud, N. B. Garnier, and S. Ciliberto, Europhys. Lett. **82**, 30007 (2008).

[136] C. Falcón and E. Falcon, Phys. Rev. E **79**, 041110 (2009).

[137] V. Blickle, T. Speck, L. Helden, U. Seifert, and C. Bechinger, Phys. Rev. Lett. **96**, 070603 (2006).

[138] T. Speck, V. Blickle, C. Bechinger, and U. Seifert, Europhys. Lett. **79**, 30002 (2007).

[139] A. Gomez-Marin and I. Pagonabarraga, Phys. Rev. E **74**, 061113 (2006).

[140] J. Mehl, T. Speck, and U. Seifert, Phys. Rev. E **78**, 011123 (2008).

[141] T. Nemoto and S.-i. Sasa, Phys. Rev. E **83**, 030105 (2011).

[142] U. Seifert, Europhys. Lett. **70**, 36 (2005).

[143] N. Kumar, C. Van den Broeck, M. Esposito, and K. Lindenberg, Phys. Rev. E **84**, 051134 (2011).

[144] M. Einax and P. Maass, Phys. Rev. E **80**, 020101 (2009).

[145] L. Crochik and T. Tomé, Phys. Rev. E **72**, 057103 (2005).

[146] T. Tomé and M. J. de Oliveira, Phys. Rev. Lett. **108**, 020601 (2012).

[147] A. Piscitelli, F. Corberi, G. Gonnella, and A. Pelizzola, J. Stat. Mech. P01053 (2009).

[148] T. Taniguchi and E. Cohen, J. Stat. Phys. **126**, 1 (2007).

[149] T. Taniguchi and E. G. D. Cohen, J. Stat. Phys. **130**, 1 (2008).

[150] E. G. D. Cohen, J. Stat. Mech. P07014 (2008).

[151] D. D. L. Minh and A. B. Adib, Phys. Rev. E **79**, 021122 (2009).

[152] B. I. Lev and A. D. Kiselev, Phys. Rev. E **82**, 031101 (2010).

[153] T. Speck, V. Blickle, C. Bechinger, and U. Seifert, Europhys. Lett. **79**, 30002 (2007).

[154] P. de Gennes, J. Stat. Phys. **119**, 953 (2005).

[155] H. Touchette, E. V. der Straeten, and W. Just, J. Phys. A **43**, 445002 (2010).

[156] M. Abramowitz and I. A. Stegun, *Handbook of Mathematical Functions* (Dover, New York, 1970).

[157] J. Dunkel and P. Hänggi, Phys. Rev. E **71**, 016124 (2005).

[158] J. Dunkel and P. Hänggi, Phys. Rev. E **72**, 036106 (2005).

[159] A. Fingerle, C. R. Physique **8**, 696 (2007).

[160] K. S. Fa, Braz. J. Phys. **36**, 777 (2006).

[161] P. Visco, J. Stat. Mech. P06006 (2006).

[162] H. C. Fogedby and A. Imparato, J. Stat. Mech. P05015 (2011).

[163] M. Matsuo and S. Sasa, Physica A **276**, 188 (2000).

[164] N. V. Kampen, J. Phys. Chem. Sol. **49**, 673 (1988).

[165] G. Stolovitzky, Phys. Lett. A **241**, 240 (1998).

[166] M. Widder and U. Titulaer, Physica A **154**, 452 (1989).

[167] P. M. Chaikin and T. C. Lubensky, *Principles of Condensed Matter Physics* (Cambridge University Press, Cambridge, 1995).

[168] L. Brenig and C. Van den Broeck, Phys. Rev. A **21**, 1039 (1980).

[169] N. G. van Kampen, AIP Conference Proceedings **27**, 153 (1976).

[170] M. Kac, Rocky Mountain J. Math. **4**, 497 (1974).

[171] G. M. Torrie and J. P. Valleau, J. Comp. Phys. **23**, 187 (1977).

[172] D. D. Frantz, D. L. Freeman, and J. D. Doll, J. Chem. Phys. **93**, 2769 (1990).

[173] B. A. Berg and T. Neuhaus, Phys. Lett. B **267**, 249 (1991).

[174] E. Marinari and G. Parisi, Europhys. Lett. **19**, 451 (1992).

[175] F. Wang and D. P. Landau, Euro. Phys. Lett. **86**, 2050 (2000).

[176] R. H. Swendsen and J.-S. Wang, Phys. Rev. Lett. **57**, 2607 (1986).

[177] K. Hukushima and K. Nemoto, J. Phys. Soc. Jpn. **65**, 1604 (1996).

[178] U. H. E. Hansmann, Chem. Phys. Lett. **281**, 140 (1997).

[179] U. H. E. Hansmann, Y. Okamoto, and F. Eisenmenger, Chem. Phys. Lett. **259**, 321 (1996).

[180] M. Falcioni and M. W. Deem, J. Chem. Phys. **110**, 1754 (1999).

[181] Y. Sugita and Y. Okamoto, Chem. Phys. Lett. **314**, 141 (1999).

[182] Y. Sugita, A. Kitao, and Y. Okamoto, J. Chem. Phys. **113**, 6042 (2000).

[183] M. K. Fenwick and F. A. Escobedo, J. Chem. Phys. **119**, 11998 (2003).

[184] H. Liu and K. D. Jordan, J. Phys. Chem. A **109**, 5203 (2005).

[185] Y. Ishikawa, Y. Sugita, T. Nishikawa, and Y. Okamoto, Chem. Phys. Lett. **333**, 199 (2001).

[186] H. Fukunishi, O. Watanabe, and S. Takada, J. Chem. Phys. **116**, 9058 (2002).

[187] P. Liu, B. Kim, R. A. Friesner, and B. J. Berne, P. Natl. Acad. Sci. USA **102**, 13749 (2005).

[188] X. Cheng, G. Cui, V. Hornak, and C. Simmerling, J. Phys. Chem. B **109**, 8220 (2005).

[189] V. I. Manousiouthakis and M. W. Deem, J. Chem. Phys. **110**, 2753 (1999).

[190] D. A. Kofke, J. Chem. Phys. **117**, 6911 (2002).

[191] D. A. Kofke, J. Chem. Phys. **120**, 10852 (2004).

[192] N. Rathore, M. Chopra, and J. J. de Pablo, J. Chem. Phys. **122**, 024111 (2005).

[193] A. Kone and D. A. Kofke, J. Chem. Phys. **122**, 206101 (2005).

[194] S. Trebst, M. Troyer, and U. H. E. Hansmann, J. Chem. Phys. **124**, 174903 (2006).

[195] C. Predescu, M. Predescu, and C. V. Ciobanu, J. Phys. Chem. B **109**, 4189 (2005).

[196] H. G. Katzgraber, S. Trebst, D. A. Huse, and M. Troyer, J. Stat. Mech. P03018 (2006).

[197] E. Bittner, A. Nußbaumer, and W. Janke, Phys. Rev. Lett. **101**, 130603 (2008).

[198] N. Rathore, M. Chopra, and J. J. de Pablo, J. Chem. Phys. **122**, 024111 (2005).

[199] C. Predescu, M. Predescu, and C. V. Ciobanu, J. Chem. Phys. **120**, 4119 (2004).

[200] D. Frenkel and B. Smit, *Understanding Molecular Simulation - From Algorithms to Applications*, 2nd ed. (Academic Press, San Diego, 2002).

[201] J. P. Neirotti, F. Calvo, D. L. Freeman, and J. D. Doll, J. Chem. Phys. **112**, 10340 (2000).

[202] F. Calvo, J. P. Neirotti, D. L. Freeman, and J. D. Doll, J. Chem. Phys. **112**, 10350 (2000).

[203] D. J. Earl and M. W. Deem, Phys. Chem. Chem. Phys. **7**, 3910 (2005).

[204] P. E. Kloeden, E. Platen, and H. Schurz, *Numerical Solution of SDE Through Computer Experiments* (Springer-Verlag, Berlin Heidelberg, 2000).

[205] W. H. Press, S. A. Teukolsky, W. T. Vetterling, and B. P. Flannery, *Numerical Recipes in C: The Art of Scientific Computing*, 2nd ed. (Cambridge University Press, New York, 1995).