Nonequilibrium phase transitions and dynamical scaling regimes

Richard Alexander Blythe

Doctor of Philosophy
The University of Edinburgh
September 2001

Abstract

A general problem in physics is the prediction of a system's macroscopic behaviour given a microscopic model. The established procedure—known as statistical physics—is to try and determine the probability distribution of microscopic configurations. Once this is known, one can evaluate macroscopic quantities such as the average energy, volume or pressure of the system.

In recent years, the application of statistical mechanics to nonequilibrium systems, and quite specifically the probabilistic modelling of nonequilibrium microscopic dynamics, has become a major research topic. However, in contrast to the equilibrium case, there is currently no general framework within which nonequilibrium systems are understood. Hence the aim of this thesis is to improve our understanding of nonequilibrium systems through the study of a range of systems with probabilistic microscopic dynamics and the collective phenomena—notably phase transitions and the onset of scaling regimes—that arise.

In this thesis I briefly review general aspects of mathematical models of probabilistic dynamics (stochastic processes), with a particular emphasis on steady-state properties and the origin of phase transitions. Then I consider separately four specific types of nonequilibrium dynamics.

Firstly, I introduce and solve exactly a model of a particle reaction system. The solution, which employs commutation properties of the q-deformed harmonic oscillator algebra, reveals that phase transitions in the analytic form of the particle density as a function of time arise as a direct consequence of randomness in the reaction dynamics. I also use similar mathematical techniques to solve the partially asymmetric exclusion process, an important prototype of a physical system that is driven by its environment. This model is also found to exhibit phase transitions, although in this case their origin lies in the nonequilibrium interactions between the system and its surroundings.

Then I examine the scaling behaviour associated with the nonequilibrium directed percolation continuous phase transition. This transition is related to the presence of an absorbing state and I provide evidence for such a transition in a wetting model that does not possess an absorbing state. Finally, I generalise the wetting model to two dimensions and study its interfacial scaling behaviour. This is found to belong to the Kardar-Parisi-Zhang universality class, although there are strong crossover effects—which I quantify—that obscure the scaling regime.

Acknowledgements

My warmest thanks go to my principal supervisor, Martin Evans, for motivation, inspiration, guidance and for taking the time to explain it all so patiently.

I also thank my second supervisor, Mike Cates, for general support, encouragement and illuminating conversations, particularly those that took place in the Old Bell.

Additional thanks are due to all those people—both from within and outside the department here—who gave me the opportunity to discuss my work with them and responded with helpful suggestions.

Thanks also to all those people who put up with me during my stay in Edinburgh. This includes the various incumbents of JCMB Room 5202, my friends, flatmates and the Edinburgh University Music Society for giving me something to do on Saturday mornings and during the Easter holidays. Special thanks to those who provided distraction during the writing-up period, most notably the various Edinburgh summer festivals and the *Ex-C3-Flurbewohner* for their hospitality during a most welcome break.

Finally, I am especially grateful to my family for their support and for understanding the idea, if not the details.

Contents

Declaration v

Acknowledgements vii

1 Introduction 1

2 Mathematics of stochastic processes 7

2.1 Discrete-time Markov process 7

2.2 Continuous-time Markov process 13

2.3 Detailed balance and nonequilibrium steady states 16

2.4 The thermodynamic limit 22

2.5 How to solve a stochastic process 25

2.6 Summary 31

3 Annihilation and coalescence reactions 33

3.1 Particle reaction systems 33

3.2 The combinatorics of sorting 39

3.3 Combinatorics of ballistic reaction systems 46

3.4 Survival probabilities in the SBAC model 50

3.5 Properties of the infinite system 62

3.6 Summary of results and open questions 66

4 Exclusion processes 69

4.1 Review of the driven lattice gas 69

4.2 Partially asymmetric exclusion process (PASEP) 73

4.3 Exact solution of the PASEP 78

4.4 Exact expressions for the normalisation 83

4.5 Phase diagram for the current 91
4.6 Density profiles and correlations 97
4.7 Summary of results 101
4.8 Other studies of exclusion processes: a review 102
4.9 Outlook 106

5 Models with an absorbing state **109**

5.1 The absorbing state phase transition 109
5.2 Driven asymmetric contact process (DACP) 117
5.3 Steady state of the driven asymmetric contact process 118
5.4 Solvability of models with an absorbing state 125
5.5 Summary of results 128

6 Dynamics of wetting **129**

6.1 The wetting model 129
6.2 Theory of interfacial growth: a review 136
6.3 Interfacial growth in the wetting model 143
6.4 Crossover to KPZ scaling in noncompact structures 146
6.5 Summary of results 149

7 Conclusion and outlook **151**

Bibliography **155**

Chapter 1

Introduction

A puzzle that has occupied physicists' minds for centuries is how the macroscopically observable properties of a system arise from the behaviour of its microscopic constituents. Particularly intriguing is the fact that whilst an everyday sample of, say, a gas comprises $\sim 10^{23}$ particles, its equilibrium thermodynamic state is entirely specified by just a few macroscopic variables, e.g. pressure, volume and temperature, which in turn are related through an equation of state.

The reduction of a large number of internal degrees of freedom to a small set of macroscopic thermodynamic variables is well understood in the framework of equilibrium statistical physics. This unified approach to the prediction of macroscopic properties of many-body systems that are at equilibrium with their environment is most elegantly formulated as follows[1].

Firstly, in recognition of the fact that a large number of coupled equations of motion cannot be solved, and even if they could one would only be able to treat the solution usefully by computing some form of average, one introduces a *probabilistic* description of the system at the microscopic level. That is, one introduces a probability distribution $P(\omega)$ where ω labels a microscopic configuration that the system can adopt.

Then, one wishes to minimise the amount of information required to describe a system, given any equilibrium constraints. One example of such a constraint is a heat bath that allows energy to flow in and out of the system in such a way that the temperatures of the system and the heat bath are equal. The equilibrium probability distribution of configurations $P_{eq}(\omega)$ is then obtained by varying the probabilities $P_{eq}(\omega)$ until a suitably defined indicator of the information content of the system[2] is minimised whilst respecting the constraints that $P_{eq}(\omega)$ is properly normalised and that the average energy is fixed.

[1] For more details, see e.g. [5]

[2] This indicator is the Gibbs entropy, albeit negated so that high entropy (or greater disorder) corresponds to low information content.

The resulting equilibrium distribution is the Boltzmann distribution

$$P_{eq}(\omega) = \frac{\exp(-\mathcal{H}(\omega)/k_B T)}{Z_{eq}} \tag{1.1}$$

in which the Hamiltonian $\mathcal{H}(\omega)$ is the energy of the microscopic configuration ω, k_B is Boltzmann's constant, T the temperature and Z_{eq} the *partition function* which is required to normalise $P_{eq}(\omega)$. In fact, it turns out that the partition function $Z_{eq} = \sum_\omega \exp(-\mathcal{H}(\omega)/k_B T)$ is a fundamental statistical signature in its own right and simply related to the free energy of the system and thence through differentiation to thermodynamic quantities such as the average energy. Thus every microscopic model of an equilibrium system can be solved by evaluating the partition function—although for nearly all systems with nontrivial interactions this is a difficult task.

This thesis is concerned with the contrasting scenario of a physical system that is *far* from equilibrium. To explain the meaning of this term, I provide two concrete examples.

Consider first the case of a thermal conductor, e.g. a metal rod, coupled at opposite ends to heat baths at different temperatures. Clearly there will be a flow of energy from the warmer bath to the cooler and we understand there to be a steady state in which the mean heat flow is constant. Such a system then exhibits a *nonequilibrium* steady state, i.e. one which is not described through a probability distribution of the Boltzmann type (1.1).

Now consider a ferromagnetic sample that initially is at a high enough temperature that its equilibrium state has no overall magnetisation. Then, quench it rapidly to a temperature below that at which the system spontaneously magnetises. Somehow, the magnet must relax from the high-temperature disordered state to the equilibrium ordered state in which the mean magnetisation is nonzero. In this instance, we would wish to understand the dynamics of this relaxation.

In both of these nonequilibrium scenarios it is unclear whether the considerations that led to the important result (1.1) are at all applicable. However we should still like to understand the relationship between a microscopic model of a nonequilibrium system and the macroscopic phenomena exhibited, particularly because the vast majority of real physical systems are not at equilibrium.

To treat nonequilibrium systems theoretically it is necessary to consider explicitly the microscopic dynamics. To understand this, consider again the above example systems. For the conducting rod, one must clearly stipulate a means for the energy to be transported across the system; similarly for the ferromagnet, one must supply a description of the relaxation mechanism. Given that the microscopic dynamics of a system expressed in terms of a set of equations of motion cannot be solved, it is once again appropriate to use instead a probabilistic

description of the dynamics. In other words, one prescribes the probability that the system moves from one microscopic configuration to another over a specified time period.

Probabilistic models of microscopic dynamics are known mathematically as *stochastic processes* and I discuss aspects of these in chapter 2 of this thesis. In particular I indicate how the probabilistic modelling of a system's dynamics amounts to the definition of a *master equation* that describes how probability flows through the space of microscopic configurations. By solving this equation for the probability distribution of configurations, one can—as with equilibrium systems—perform averages to obtain predictions for macroscopic quantities. However, as will become evident in chapter 2, there is currently no general approach to the solution of a master equation that is tractable for macroscopically large systems. Hence, there is no unified framework within which nonequilibrium systems can be understood. For example, it is not currently known if there is something akin to a partition function for a nonequilibrium system that encodes all the important statistical information about the macroscopic behaviour.

So that our understanding of nonequilibrium systems can be improved, it is necessary to consider separately systems with different types of stochastic dynamics to find out what macroscopic phenomena emerge. This is the strategy I follow in chapters 3 to 6, each of which is concerned with a different type of microscopic dynamics.

Exact solutions provide arguably the greatest insight into a model of a physical system and in chapter 3 I present an exact solution of a particle reaction system. The basic picture of a reaction system is one of a large number of particles exploring a region of space and which react on contact. This leads to a time-dependent density decay as the system relaxes to a steady state in which no more reactions are possible. Although one can devise arbitrarily complicated reaction dynamics, it will be seen in chapter 3 that even systems comprising only pairwise reactions between particles of a single species exhibit a range of nontrivial macroscopic phenomena.

As will be discussed in more detail in chapter 3, there are a number of established results that hold for single-species reaction dynamics with diffusive particle motion. Most notably, the density decay exhibited by both pairwise annihilating and coalescing reactants is the same and depends on the spatial dimensionality of the system until it is sufficiently large that a *mean-field* description (in this case, one that assumes a uniform spatial distribution of the particles at all times) is valid. Here we recognise two properties of equilibrium critical phenomena—i.e. universality and the existence of an upper critical dimension below which mean-field results do not apply [6, 7].

By solving exactly a one-dimensional single-species reaction system that has ballistic (deterministic) rather than diffusive reactant motion, I show that it exhibits another phenomenon familiar from equilibrium statistical physics and thermodynamics, namely a phase transition.

This transition is manifested in different analytical forms of the density decay in the system and, as will be shown in chapter 3, it arises as a direct consequence of introducing stochasticity into the reaction dynamics.

I obtain the exact solution of the particle reaction system by recognising the one-dimensional ballistic reaction system as a combinatorial problem related to the reordering of operator products. As I explain in chapter 4, similar techniques can be used to find an exact solution in the markedly different context of the steady state of the partially asymmetric exclusion process. This latter system is a prototype of a model that is driven by its environment rather than being at equilibrium with it. Specifically it comprises a one-dimensional lattice along which particles are transported in a preferred direction. At one end of the lattice particles are inserted and at the other they are removed, and so one can view this model as being in the spirit of the thermal conductor coupled to two heat reservoirs mentioned above. Additionally the model can be directly mapped onto a wide range of other systems including transport across membranes, the kinetics of biopolymerisation and traffic flow. These applications are described in chapter 4 along with a detailed discussion of the model's exact solution and the rich phase behaviour that results.

It is not always possible to solve a model exactly and in chapters 5 and 6 I investigate a wetting process using approximate and numerical methods. In this model there is a transition from a regime in which only part of the system becomes wet to one in which the wet zone increases in size indefinitely. In chapter 5 I show that the transition between these two regimes can be characterised by a steady-state *order parameter* and that the transition is continuous. In common with equilibrium continuous phase transitions, there is a *scaling regime* around the transition point where length- and timescales diverge as power-laws. In chapter 5 I provide evidence for the exponents of these power-laws to be in agreement with those of directed percolation which defines a well-studied and ubiquitous nonequilibrium universality class.

Invariance under rescaling is also exhibited in the time-dependent properties of the wetting process, and this is the subject of chapter 6. Specifically I consider the statistics of the interface that separates the advancing wet zone from the dry region in front of it. These statistics obey a *dynamical scaling relation* in which the characteristic length- and timescales of the interface scale as powers of the system size. For reasons that are discussed in chapter 6, the most common type of interfacial scaling is that of the Kardar-Parisi-Zhang universality class and I also provide evidence for the wetting model to belong to this class. However, complications arise due the presence of a slow crossover to the true asymptotic scaling regime. This is a common situation and in chapter 6 I detail the ways in which this crossover can be identified and quantified.

Already in this brief overview I have noted a number of macroscopic phenomena present in

the simple models of this thesis and how some are reminiscent of those seen in equilibrium systems. In the final chapter of this thesis, I comment on these recurring features and contrast with their equilibrium counterparts and also pose some open questions regarding our general understanding of nonequilibrium systems.

Chapter 2

Mathematics of stochastic processes: an informal introduction

In the previous chapter I motivated the modelling of nonequilibrium systems through a probabilistic description of microscopic dynamics. In this chapter I discuss how such models, known as stochastic processes, are constructed and consider specifically Markov processes in discrete and continuous time. Of particular interest are steady states of such processes, and I discuss in detail the situation where a process exhibits a single, unique steady state (which I refer to as ergodicity). Furthermore, in section 2.3, I attempt to understand the distinction between equilibrium and nonequilibrium steady states and how this is related to the concept of detailed balance. Additionally I discuss how one calculates macroscopic properties that arise from a microscopic model, focussing in section 2.4 on how phase transitions may arise in systems with an infinite number of configurations.

The discussion will touch upon various established aspects of probability, matrix and graph theory. For more thorough accounts of these subjects, the reader is advised to consult standard textbooks, such as [8, 9, 10] (for probability theory), [11, 12] (matrix theory) and [13, 14] (graph theory).

2.1 Discrete-time Markov process

A number of important notions associated with stochastic processes are conveniently introduced using the specific example of the discrete-time Markov process. This is defined as follows. At a given time t it is assumed that the physical configuration adopted by a system, denoted ω, is a known quantity. Then, instantaneously at time $t+1$, a *transition* to a new configuration ω' occurs with a prescribed probability $\mathcal{M}(\omega \to \omega')$. Alternatively, the system may remain in the same configuration ω with a probability $\mathcal{M}(\omega \to \omega) = 1 - \sum_{\omega' \neq \omega} \mathcal{M}(\omega \to \omega')$.

This set of rules then generates a sequence of configurations $\omega_0, \omega_1, \ldots, \omega_t$ which together form a *realisation* of the stochastic process. Note that the choice of configuration ω_{t+1} at time $t+1$ depends *only* on the configuration adopted at time t in the realisation. This is the Markov property of the process. Additionally it will be assumed that the *transition probabilities* $\mathcal{M}(\omega \to \omega')$ are time-independent—this is true of all the systems studied in this thesis. In this section I also consider only the case of a finite number n of configurations accessible by the system—extension to the limit $n \to \infty$ will be considered in section 2.4.

2.1.1 Graphical representation of the model, directed paths and connectedness

The Markov process defined through a set of transition probabilities is conveniently represented using a *directed graph* (or digraph). This is constructed by drawing a vertex for each configuration accessible by the system, and a directed edge (arrow) drawn from one configuration ω to another ω' wherever the transition rate $\mathcal{M}(\omega \to \omega')$ is nonzero. A graph for a system with five (numbered) configurations is shown in figure 2.1(i).

Note that the existence of a directed edge (transition) $\omega \to \omega'$ does not imply the existence of the transition $\omega' \to \omega$. More generally, a realisation of the process that begins at configuration ω_0 and reaches configuration ω_t at time t is represented by a sequence of directed edges placed 'end-to-end' on the graph—see figure 2.1(ii). Together this sequence of edges forms a *directed path* $\omega_0 \to \omega_1, \omega_1 \to \omega_2, \ldots \omega_{t-1} \to \omega_t$. In this chapter I will use the notation $\omega \twoheadrightarrow \omega'$ to denote a path originating at configuration ω and ending at ω'.

As with single edges, the existence of a path $\omega \twoheadrightarrow \omega'$ does not necessarily imply the existence of the path $\omega' \twoheadrightarrow \omega$. This notion is associated with the graphical property of *connectedness*

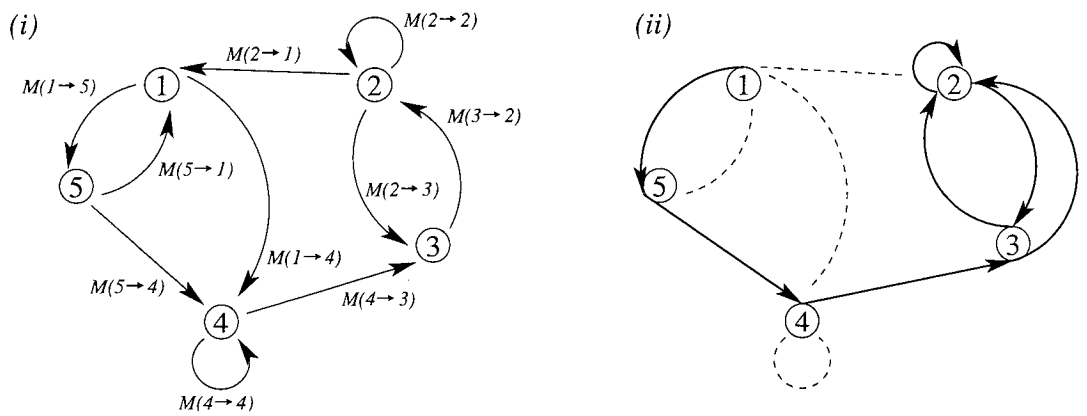

Figure 2.1: (i) Directed graph showing the allowed transitions in a system which may be in one of five configurations, numbered. Only transitions for which the probability $\mathcal{M}(\omega \to \omega') \neq 0$ are included. (ii) A realisation of the Markov process represented as a directed path comprising 7 edges; note that the same edge may be included in the path more than once.

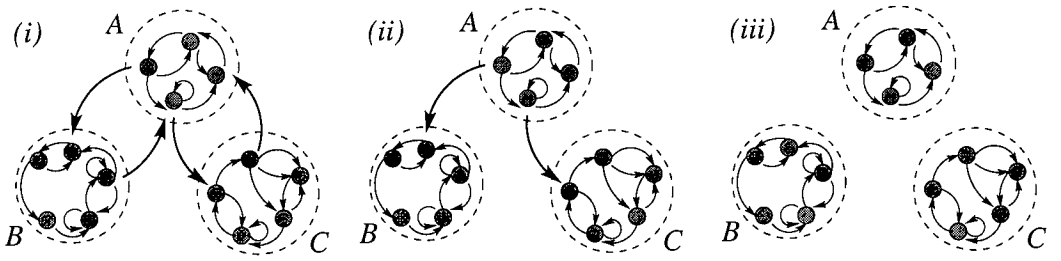

Figure 2.2: Each of these three graphs comprises three subgraphs (regions) A, B and C which are strongly connected. Graph (i) is strongly connected; graph (ii) is connected but not strongly connected due to the absence of transitions from region B to A and from C to A; graph (iii) is not connected due to the absence of transitions between any of the three regions.

which will be central to the discussion in section 2.1.4 below and which is defined as follows.

A graph is called *strongly connected* if for every pair of vertices ω and ω' there exists a directed path $\omega \to\!\!\!\!\to \omega'$ and $\omega' \to\!\!\!\!\to \omega$. Such a graph is shown in figure 2.2(i). More generally, a graph is *connected* if the vertices cannot be grouped into subsets (regions) between which there are no directed paths (possible transitions) in either direction. Figure 2.2(ii) provides an example of a graph that is connected but not strongly connected and figure 2.2(iii) an example of a disconnected graph. The latter corresponds to a set of noninteracting stochastic processes, each of which can be treated separately.

2.1.2 Directed paths, probability distributions and ensemble averaging

For a general set of transition rates $\mathcal{M}(\omega \to \omega')$ there may be more than one sequence of transitions (directed path) that leads from an initial configuration ω_0 to a specified configuration ω_t at time t. Each of these realisations of the stochastic process occurs with a probability given by the product of the transition rates $\mathcal{M}(\omega \to \omega')$ associated with each component edge of the path. The total probability that a transition is made from ω_0 to ω_t over a time t is then the sum of the probabilities associated with all possible paths $\omega_0 \to\!\!\!\!\to \omega_t$ of length t.

By enumerating all paths $\omega_0 \to\!\!\!\!\to \omega_t$ comprising t edges, and summing over the associated probabilities, one obtains the conditional *probability distribution* $P_t(\omega|\omega_0)$ that configuration ω is reached after time t given that the process began in configuration ω_0. This distribution contains complete information about the statistical properties of the process.

For example, if with every configuration ω one can define an observable quantity $\mathcal{O}(\omega)$ (such as particle density or current) then its *expectation value* at time t is given by

$$\langle \mathcal{O}_t \rangle = \sum_{\omega} P_t(\omega|\omega_0)\mathcal{O}(\omega) \; . \tag{2.1.1}$$

Generally it is this type of average that is associated with a system's macroscopic properties, and hence it is by using (2.1.1) that the collective behaviour of physical systems evolving

under stochastic dynamics is determined. In this thesis, angular brackets are used exclusively to denote the average over the ensemble of all possible paths $\omega_0 \twoheadrightarrow \omega$ over a time t as indicated in equation (2.1.1). I refer to this type of average as an *ensemble average* and in section 2.1.4 I concentrate on the case where the steady state of a process does not depend on the initial condition ω_0. There I shall also briefly discuss the contrasting case of a time average.

In general, the construction of a probability distribution by enumerating paths between configurations is tractable only for simple processes. A specific example, to be found in nearly all probability textbooks, is that of the drunkard making his way home from an evening's entertainment by lurching randomly from one lamp-post to another. The resulting probability distribution is then the well-known binomial distribution. In chapter 3 I solve a stochastic particle reaction system in one dimension by considering all possible paths from one sequence of particles to another.

2.1.3 Matrix formulation and master equation

It is useful to encode the probabilities $\mathcal{M}(\omega \to \omega')$ into a *transition probability matrix* M as follows

$$[M]_{ij} = \mathcal{M}(\omega(j) \to \omega(i)) \tag{2.1.2}$$

in which $\omega(i)$ is the configuration associated with matrix index $1 \leq i \leq n$ with n the number of configurations. A probability distribution $P_t(\omega|\omega_0)$ can similarly be cast into vector form through

$$\boldsymbol{P}_t = \sum_{i=1}^{n} P_t(\omega(i)|\omega(j))\, \boldsymbol{e}_i \tag{2.1.3}$$

in which $\{\boldsymbol{e}_i\}$ are basis vectors in the n-dimensional vector space.

The matrix element $[M^t]_{ij}$ (in which the superscript t is a power) corresponds to the sum of probabilities associated with all paths of length t originating at configuration $\omega(j)$ and ending at $\omega(i)$. Hence, from the previous subsection, the probability distribution at time t is given by

$$\boldsymbol{P}_t = M^t \boldsymbol{P}_0 \tag{2.1.4}$$

where $\boldsymbol{P}_0 = \boldsymbol{e}_j$, the initial condition. Then one finds that the probability distribution at time $t+1$ is related to that at time t through the *master equation*

$$\boldsymbol{P}_{t+1} = M\boldsymbol{P}_t\,. \tag{2.1.5}$$

Usually this is the equation one must solve in order to understand a stochastic process.

An important matrix property in the context of Markov processes is that of *reducibility*. This is closely related to the connectedness of graph that accompanies the set of transition rates

contained within the matrix M. Specifically, if the graph is strongly connected, it follows that $[M^\ell]_{ij} > 0 \, \forall \, i, j$ and some positive integer ℓ which may depend on the indices i and j. This is one condition [12] for a matrix to be irreducible.

A stronger property, relevant in the following, is that of *primitivity*. A matrix is primitive if there exists at least one common integer ℓ for which $[M^\ell]_{ij} > 0$ for all i and j. It may not always be straightforward to test the primitivity of a matrix directly. However it is usually simple to apply the equivalent graphical condition that there exists a directed path of fixed length ℓ between every pair of configurations ω, ω' (including $\omega = \omega'$).

2.1.4 Convergence to a stationary state and the property of ergodicity

Often the probability distribution that is of greatest interest is the limiting distribution

$$\lim_{t \to \infty} P_t(\omega|\omega_0) = P^*(\omega) \, . \tag{2.1.6}$$

When this limit exists and is unique (i.e. independent of the initial condition ω_0) the stochastic process is called *ergodic*[1]. As will be seen below, neither the existence nor the uniqueness of the limit (2.1.6) is guaranteed for a general process.

Ergodic processes warrant attention because they exhibit a number of special features. Firstly (if the number of configurations is finite) the probabilities $P^*(\omega)$ are all nonzero. Secondly, $P^*(\omega)$ is a *stationary* (or *steady-state*) distribution in that $P^*(\omega) = P_{t+1}(\omega) = P_t(\omega)$. Finally, an ensemble average computed with respect to this distribution using (2.1.1) can be shown [9] to be equal to a time average evaluated over a single, infinitely long realisation of the stochastic process. This property is of particular importance for comparison of results with computer simulation or experiment in which such a time average is usually performed.

Sufficient criterion for ergodicity

I now present a sketch[2] of the proof of the sufficient (but not necessary) condition for ergodicity which states that every process for which the transition probability matrix M is primitive (see section 2.1.3) is ergodic. Recall that the primitivity property implies that the graph representing the process is strongly connected and that there exists some common path length ℓ that connects every pair of configurations ω and ω'.

To understand why a process represented by a graph that is not strongly connected is not necessarily ergodic, consider again figure 2.2(ii). If the initial configuration ω_0 is in region B

[1] This present definition of ergodicity is not universally accepted, although it is commonplace in the mathematical (as opposed to the physics) literature.

[2] But see also [11, 12].

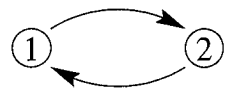

Figure 2.3: The simplest strongly-connected graph which does not correspond to an ergodic process. All paths of even length connect $1 \nrightarrow 1$ and $2 \nrightarrow 2$ whereas all odd-length paths connect $1 \nrightarrow 2$ and $2 \nrightarrow 1$. Hence there is no common path length ℓ that connects every possible pair of configurations.

of that system, there are no paths from ω_0 to any configurations in regions A, implying that $P_t(\omega|\omega_0)$ is nonzero only for ω in region B. Similarly, if ω_0 is in region C, $P_t(\omega|\omega_0)$ is nonzero only for ω in region C. Hence the limiting probability distribution is not unique, and this particular process nonergodic.

If the graph is strongly connected, but the transition probability matrix M is not primitive (i.e. it is only irreducible), the limiting probability distribution (2.1.6) does not exist. In fact, the distribution exhibits some cyclic behaviour as $t \to \infty$ as is illustrated by the simplest system for which (2.1.6) does not exist, figure 2.3. The state of this system alternates deterministically between configurations 1 and 2 no matter what the starting configuration is.

To show that the transition probability matrix M must be primitive to describe an ergodic process one invokes the Perron-Frobenius theorem [11]. This states that a primitive nonnegative matrix M has a real eigenvalue whose magnitude is larger than that of any other eigenvalue. Furthermore, the right eigenvector corresponding to this eigenvalue has only positive elements. I now demonstrate that this right eigenvector corresponds to a stationary probability distribution and that all initial probability distributions converge to it.

First, it is assumed that the matrix M can be diagonalised[3]. Then there exists a set of eigenvalues $\{\lambda_i\}$ and pairs of left and right eigenvectors $\boldsymbol{u}_i, \boldsymbol{v}_i$ that satisfy

$$\boldsymbol{u}_i M = \lambda_i \boldsymbol{u}_i \ , \quad M\boldsymbol{v}_i = \lambda_i \boldsymbol{v}_i \quad \text{with} \quad \boldsymbol{u}_i \cdot \boldsymbol{v}_j = \delta_{i,j} \ . \tag{2.1.7}$$

This last relationship expresses the *biorthogonality* of the left and right eigenvectors. Any transition probability matrix M always has a left eigenvector $\boldsymbol{u}_1 = (1, 1, 1, \ldots, 1)$ because the columns of M add up to one (recall that this reflects conservation of probability). Hence $\lambda_1 = 1$ is an eigenvalue of M because $\boldsymbol{u}_1 M = \boldsymbol{u}_1$. The right eigenvector $\boldsymbol{v}_1$ corresponding to this eigenvalue is then a stationary probability distribution because of the master equation (2.1.5)

$$\boldsymbol{P}_{t+1} = M\boldsymbol{P}_t \tag{2.1.8}$$

and so if $\boldsymbol{P}_t = \boldsymbol{v}_1$, $\boldsymbol{P}_{t+1} = \boldsymbol{P}_t = \boldsymbol{v}_1$ because $M\boldsymbol{v}_1 = \boldsymbol{v}_1$.

[3]This is not generally true; however M can always be reduced to Jordon normal form, and a similar, although slightly more cumbersome, argument holds [12].

To show that this is the only probability distribution invariant under the master equation (2.1.5) one must ensure that there are no other eigenvalues of M of magnitude one. This is seen by expanding an initial condition $\boldsymbol{P}_0$ as

$$\boldsymbol{P}_0 = \boldsymbol{v}_1 + \sum_{i=2}^{n} (\boldsymbol{u}_i \cdot \boldsymbol{P}_0)\boldsymbol{v}_i \tag{2.1.9}$$

in which the fact that $\boldsymbol{u}_i \cdot \boldsymbol{v}_j = \delta_{i,j}$ and $\boldsymbol{u}_1 = (1, 1, 1 \ldots, 1)$ has been used. Then, (2.1.4) implies that after time t

$$\boldsymbol{P}_t = M^t \boldsymbol{P}_0 = \boldsymbol{v}_1 + \sum_{i=2}^{n} (\boldsymbol{u}_i \cdot \boldsymbol{P}_0)\, \lambda_i^t\, \boldsymbol{v}_i \,. \tag{2.1.10}$$

Now if $\boldsymbol{P}_t$ is to be a properly normalised probability vector for all t, no element may ever exceed 1, and so $|\lambda_i| \leq 1$. However, the Perron-Frobenius theorem for the primitive matrix M states that only one eigenvalue may satisfy the equality $|\lambda_i| = 1$ and this has already been identified as $\lambda_1 = 1$. Hence, $\boldsymbol{v}_1$ is not just the only stationary distribution of the process, but also the limiting distribution as $t \to \infty$. That is

$$\boldsymbol{P}^* = \lim_{t\to\infty} \boldsymbol{P}_t = \boldsymbol{v}_1 \,. \tag{2.1.11}$$

If the matrix were nonprimitive, there could be other eigenvalues of magnitude 1, and hence the above limit would not exist. Note also that the Perron-Frobenius theorem guarantees that every element of $\boldsymbol{v}_1$ (and hence $\boldsymbol{P}^*$) is greater than zero, i.e. every configuration has some probability of being realised in the steady state of an ergodic process.

2.2 Continuous-time Markov process

In the foregoing, I examined the construction and some properties of a Markov process evolving in discrete time. I now consider the (physically more reasonable) case of a Markov process evolving in continuous time and establish how it is related to a discrete-time Markov process. Then, the results obtained above that concern ergodicity and the steady state of a discrete-time process can be directly transferred to the case of continuous time.

2.2.1 Transition rates and master equation

A process evolving in continuous time is defined in terms of *transition rates* $\mathcal{W}(\omega \to \omega')$ at which a system can migrate from configuration ω to ω'. We understand this to mean that, in the limit of an infinitesimal time interval $\mathrm{d}t$ going to zero, only one transition may occur, and that the probability of this transition is $\mathcal{W}(\omega \to \omega')\mathrm{d}t$. Again, the Markov property is assumed.

This implies that if the continuous-time process is discretised using infinitesimal time intervals $\mathrm{d}t$, the corresponding transition probabilities $M_{\mathrm{d}t}(\omega \to \omega')$ are given by

$$\mathcal{M}_{\mathrm{d}t}(\omega \to \omega') = \delta(\omega, \omega') + \mathcal{W}(\omega \to \omega')\mathrm{d}t + \mathrm{O}([\mathrm{d}t]^2) \; . \tag{2.2.1}$$

In this equation, the first term reflects the fact that if one chooses a zero timestep, no transitions are possible. Recalling that probability is conserved, i.e. $\sum_{\omega'} \mathcal{M}_{\mathrm{d}t}(\omega \to \omega') = 1$, we note that the previous equation implies $\sum_{\omega'} \mathcal{W}(\omega \to \omega') = 0$ and so

$$\mathcal{W}(\omega \to \omega) = - \sum_{\omega' \neq \omega} \mathcal{W}(\omega \to \omega') \; . \tag{2.2.2}$$

In words, the 'self-transition' rates $\mathcal{W}(\omega \to \omega)$ express the total rate at which configuration ω is exited.

Using the master equation for a discrete-time process (2.1.5) we find

$$P_{t+\mathrm{d}t}(\omega) = \sum_{\omega'} \mathcal{M}_{\mathrm{d}t}(\omega' \to \omega) P_t(\omega') \; . \tag{2.2.3}$$

Taking now the limit $\mathrm{d}t \to 0$ and using (2.2.1) one obtains

$$\frac{\mathrm{d}}{\mathrm{d}t} P_t(\omega) = \sum_{\omega'} \mathcal{W}(\omega' \to \omega) P_t(\omega') \tag{2.2.4}$$

which is one form of the master equation for a continuous-time process. An alternative and more revealing form of this equation is obtained after using (2.2.2). It reads

$$\frac{\mathrm{d}}{\mathrm{d}t} P_t(\omega) = \sum_{\omega' \neq \omega} \left(P_t(\omega') \mathcal{W}(\omega' \to \omega) - P_t(\omega) \mathcal{W}(\omega \to \omega') \right) \tag{2.2.5}$$

which can be interpreted as a flow of probability in configuration space. The first term in the summand describes that rate at which probability flows into configuration ω from any other configuration and the second term describes the outflow of probability to other configurations.

As in the discrete-time case, it is convenient to encapsulate the transition rates in a matrix W through

$$[W]_{ij} = \mathcal{W}(\omega(j) \to \omega(i)) \tag{2.2.6}$$

which implies W has nonnegative off-diagonal elements and nonpositive diagonal elements. Then the master equation (2.2.5) can also be written as

$$\frac{\mathrm{d}}{\mathrm{d}t} \boldsymbol{P}_t = W \boldsymbol{P}_t \tag{2.2.7}$$

where $\boldsymbol{P}_t$ is the vector of probabilities at time t introduced in section 2.1.3. Given an initial condition $\boldsymbol{P}_0$ one can write the formal solution to (2.2.7) as

$$\boldsymbol{P}_t = \exp(Wt) \boldsymbol{P}_0 \; . \tag{2.2.8}$$

This last equation also indicates how one replaces a continuous-time Markov process with an equivalent discrete-time process with a macroscopically large timestep t (recall that equation (2.2.1) is valid only in the limit of this timestep going to zero). Specifically, the transition probability matrix M_t for the discrete-time process is related to the continuous-time matrix W through

$$M_t = \exp(Wt) \; . \tag{2.2.9}$$

I now use this correspondence to apply the results obtained above for discrete-time processes to continuous-time processes.

2.2.2 Ergodicity and the steady state

Recall from section 2.1.4 that if the discrete-time transition probability matrix M_t is primitive, the stochastic process it describes is ergodic. I now show that if a transition rate matrix W is irreducible (which can be determined by investigating the connectedness of a graph as discussed in section 2.1.3) the matrix M_t is primitive for all time intervals t, and hence the continuous-time process described by W is ergodic.

To do this one rewrites W as $W = A - a\mathbb{I}$ where $\mathbb{I}$ is the identity matrix and a is suitably chosen so that all elements of A are nonnegative—recall that the off-diagonal elements of W are zero or positive, and the diagonal elements of W are zero or negative. Then

$$M_t = \exp(Wt) = \exp(-a\mathbb{I}t)\exp(At) = \exp(-at)\sum_{\ell=0}^{\infty}\frac{(At)^\ell}{\ell!} \; . \tag{2.2.10}$$

Now, if W is irreducible, then so is A. Hence, from the definition of irreducibility, one can always find an integer ℓ such that $[A^\ell]_{ij} > 0$ for fixed i, j. Since all positive integer powers of A appear in the above expression for M_t, it follows that $[M_t]_{ij} > 0 \, \forall \, i, j$. This implies that the matrix M_t is primitive and hence the stochastic process is ergodic.

Recall that ergodicity implies that the probability vector

$$\boldsymbol{P}^* = \lim_{t\to\infty} \exp(Wt)\boldsymbol{P}_0 \tag{2.2.11}$$

exists, is strictly positive and independent of the initial condition $\boldsymbol{P}_0$. Additionally, this probability vector $\boldsymbol{P}^*$ is a stationary distribution, in the sense that

$$\frac{\mathrm{d}}{\mathrm{d}t}\boldsymbol{P}^* = W\boldsymbol{P}^* = 0 \; . \tag{2.2.12}$$

Then, equation (2.2.9) reveals that eigenvectors of W are the same as those of M_t, and that the eigenvalues are simply related. In particular, an eigenvalue $\lambda = 1$ of M_t corresponds to an eigenvalue of $\lambda = 0$ of W, and so from the arguments of section 2.1.4 we know that (2.2.12) has only one solution if the stochastic process is ergodic (or equivalently if W is irreducible).

2.3 Detailed balance and nonequilibrium steady states

The concept of detailed balance often crops up in discussions of equilibrium and nonequilibrium steady states, and I review some of the ideas here.

2.3.1 Definition of detailed balance

First consider equation (2.2.12) satisfied by the steady-state (stationary) probability distribution $\boldsymbol{P}^*$ rewritten in terms of the transition rates $\mathcal{W}(\omega \to \omega')$. This reads

$$\sum_{\omega'} \left(P^*(\omega)\mathcal{W}(\omega \to \omega') - P^*(\omega')\mathcal{W}(\omega' \to \omega) \right) = 0 \quad \forall \quad \omega \,. \tag{2.3.1}$$

A stochastic process is said to obey detailed balance if each term in the summation is identically zero, i.e.

$$P^*(\omega)\mathcal{W}(\omega \to \omega') = P^*(\omega')\mathcal{W}(\omega' \to \omega) \,. \tag{2.3.2}$$

Often one wishes to know if a stochastic process obeys detailed balance or not; in particular (2.3.2) gives a method through which the steady-state probability distribution $P^*(\omega)$ can be determined efficiently from the transition rates. It is first of all evident that only ergodic processes can satisfy detailed balance: more than one steady state would imply more than one solution in $P^*(\omega)$ of (2.3.2) for the same set $\mathcal{W}(\omega \to \omega')$ which is not possible. However, for general ergodic processes, there appears to be something of a chicken-and-egg problem. One can only tell from (2.3.2) whether a process obeys detailed balance if $P^*(\omega)$ has already been calculated, at which point the construction of $P^*(\omega)$ from (2.3.2) becomes unnecessary.

To resolve this problem, one calls upon the following equivalent detailed balance condition that is expressed in terms of transition rates $\mathcal{W}(\omega \to \omega')$ alone [15, 16].

Equivalent condition for detailed balance: An ergodic process satisfies detailed balance if and only if the transition rates satisfy

$$\mathcal{W}(\omega_1 \to \omega_2)\mathcal{W}(\omega_2 \to \omega_3)\ldots\mathcal{W}(\omega_n \to \omega_1) = \mathcal{W}(\omega_1 \to \omega_n)\mathcal{W}(\omega_n \to \omega_{n-1})\ldots\mathcal{W}(\omega_2 \to \omega_1) \tag{2.3.3}$$

for *every* possible sequence in configuration space $\omega_1, \omega_2, \omega_3, \ldots, \omega_n$. This condition is sometimes called a Kolmogorov criterion [15].

That (2.3.2) implies (2.3.3) is very simple to show. To show that the converse is also true is most simply achieved [16] by using (2.3.2) to determine $P^*(\omega)$. Then, one finds that $P^*(\omega)$ is unique if and only if (2.3.3) is satisfied.

Here I shall take a longer route whose starting point is to assume only (2.3.3). In the course of obtaining (2.3.2) directly from this assumption, I shall uncover a couple of less widely-known results concerning the structure of the steady-state probability distribution $P^*(\omega)$. The first step from (2.3.3) to (2.3.2) is to solve formally the steady-state probability distribution as I now describe.

2.3.2 Formal solution for the steady-state probability distribution

As it stands, the set of linear equations (2.2.12) is underdetermined because the determinant of W is zero (one of its eigenvalues is zero). However, we know that a probability distribution must be properly normalised, so consider instead the equation

$$\widetilde{W}^{(i)}\boldsymbol{P}^* = \boldsymbol{e}_i \tag{2.3.4}$$

in which $\widetilde{W}^{(i)}$ is the matrix obtained from $-W$ (the minus sign for convenience below) by replacing the i^{th} row with all ones, i.e. the vector $(1, 1, 1, \ldots 1)$. This set of equations can now be solved using Cramer's rule. This states that the probability of the system being in configuration ω_j in the steady-state is

$$P_j^* = \frac{\det \widetilde{W}^{(i;j)}}{\det \widetilde{W}^{(i)}} \tag{2.3.5}$$

in which the matrix $\widetilde{W}^{(i;j)}$ is obtained from $\widetilde{W}^{(i)}$ by replacing column j with the vector $\boldsymbol{e}_j$. For simplicity I choose $i = j$, and to be clear I now write out $\widetilde{W}^{(j;j)}$ explicitly.

$$\widetilde{W}^{(j;j)} = \begin{bmatrix} -W_{1,1} & \cdots & -W_{1,j-1} & 0 & -W_{1,j+1} & \cdots & -W_{1,n} \\ \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots \\ -W_{j-1,1} & \cdots & -W_{j-1,j-1} & 0 & -W_{j-1,j+1} & \cdots & -W_{j-1,n} \\ 1 & \cdots & 1 & 1 & 1 & \cdots & 1 \\ -W_{j+1,1} & \cdots & -W_{j+1,j-1} & 0 & -W_{j+1,j+1} & \cdots & -W_{j+1,n} \\ \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots \\ -W_{n,1} & \cdots & -W_{n,j-1} & 0 & -W_{n,j+1} & \cdots & -W_{n,n} \end{bmatrix} . \tag{2.3.6}$$

It can be seen that $\det \widetilde{W}^{(j;j)}$ is equal to the determinant of the matrix obtained from $\widetilde{W} = -W$ by removing row j and column j. This determinant is called a cofactor of $\widetilde{W}$ and will be written $f_j \equiv f(\omega(j))$. Furthermore, the determinant in the denominator of (2.3.5) is the sum of all these cofactors. That is

$$Z = \sum_{j=1}^{n} f_j \tag{2.3.7}$$

so that

$$P_j^* = \frac{f_j}{Z} . \tag{2.3.8}$$

We can thus identify the cofactor f_j with the *steady-state weight* of configuration $\omega(j)$ and Z as a *normalisation* for the stochastic process. In an equilibrium system, Z is related to the partition function (as given by e.g. $Z_{eq} = \sum_\omega \exp(-\mathcal{H}(\omega)/k_B T)$). However, the normalisation Z obtained using the above procedure is not guaranteed to be identical to Z_{eq}. For example, one could multiply all the transition rates by a function of, e.g., temperature which would not affect the steady-state probability distribution but would give rise to common factors present in all the weights f_j and hence Z. However, if one takes care to eliminate these one has a means to extend the notion of a partition function to nonequilibrium systems. I return to this idea below in section 2.4.

2.3.3 The steady-state normalisation and eigenvalues of the transition-rate matrix

Before I proceed with the formal solution of the master equation and proof of the equivalence of (2.3.2) and (2.3.3) I note the interesting and (as far as I am aware) little-known result that the normalisation Z as defined above is also given by the equation

$$Z = \prod_{\lambda_j \neq 0} (-\lambda_j) \tag{2.3.9}$$

in which the set $\{\lambda_j\}$ comprises the eigenvalues of the transition rate W.

This result is obtained by expanding the characteristic polynomial $\det(\lambda\mathbb{I} - W)$ in powers of λ. One finds [12]

$$\det(\lambda\mathbb{I} - W) = (-)^n \left(\det W + (-)^n \lambda \sum_{j=1}^{n} f_j + \mathrm{O}(\lambda^2) \right) \tag{2.3.10}$$

with f_j as defined above. Given that $\det W = 0$ and $Z = \sum_j f_j$ one finds that

$$Z = \lim_{\lambda \to 0} \frac{\det(\lambda\mathbb{I} - W)}{\lambda} = \lim_{\lambda \to 0} \frac{\prod_j (\lambda - \lambda_j)}{\lambda} = \prod_{\lambda_j \neq 0} (-\lambda_j) \tag{2.3.11}$$

where the last step follows because one, and only one, eigenvalue of W is zero for an ergodic process.

This gives a further means to calculate the normalisation Z for any ergodic process, although one should heed the warnings of the previous subsection when trying to relate it to an equilibrium partition function Z_{eq}. Nevertheless (2.3.9) will be useful in the discussion of phase transitions in section 2.4.1 below.

2.3.4 Graphical solution for the steady-state weights

I now return to the main thread—showing the equivalence of (2.3.2) and (2.3.3)—by indicating how the steady-state weights $f(\omega(j))$ may be obtained graphically.

Recall the graphs associated with discrete-time processes discussed in section 2.1. There, a vertex was drawn for each configuration ω in the system and a directed edge for each nonzero transition probability $\mathcal{M}(\omega \to \omega')$. The graphs associated with continuous-time processes are constructed in a similar way, i.e. one draws directed edges $\omega \to \omega'$ for each nonzero transition rate $\mathcal{W}(\omega \to \omega')$ except when $\omega = \omega'$. By doing this, one obtains a graph (denoted G) such as that shown in figure 2.4(i).

The steady-state weights f_j can be obtained by considering particular types of subgraphs T of G, a subgraph being a subset of the vertices ω and edges $\omega \to \omega'$ of G. The type of subgraph T that is relevant here is a *spanning in-tree* $T(\omega)$ with sink ω. This is defined as a subgraph that contains (i) all vertices of G; (ii) exactly $n-1$ edges of G; and (iii) exactly one directed path from every vertex $\omega' \neq \omega$ to ω.

Note that this definition implies that a spanning in-tree is connected. In figure 2.4 a graph G and examples of subgraphs that are, and are not, spanning in-trees $T(\omega)$ are illustrated.

As is evident from figure 2.4, there may be more than one spanning in-tree with a common sink ω. In the following it is necessary to consider all trees with the same sink ω, and this set is denoted as $\mathcal{T}(\omega)$. Furthermore, we will call upon a weight $w(T)$ for each tree $T \in \mathcal{T}(\omega)$ which is defined as the product of all the transition rates associated with the edges of T.

The weights of spanning in-trees and matrix determinants are related through the matrix tree theorem [13, 14]. This states that the sum of the weights of all trees $T \in \mathcal{T}(\omega(j))$ is equal to the cofactor f_j of the matrix $\widetilde{W}$. That is

$$f_j = \widetilde{W}^{(j;j)} = \sum_{T \in \mathcal{T}(\omega(j))} w(T) \tag{2.3.12}$$

which indicates that both the steady-state weight f_j and normalisation $Z = \sum_j f_j$ are positive.

In principle, the enumeration of spanning in-trees gives a method to compute the steady-state weights, and thence the normalisation, graphically. In practice, the size of the set $\mathcal{T}(\omega)$ becomes

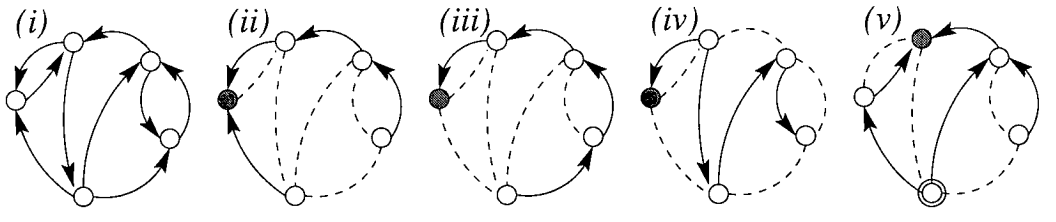

Figure 2.4: (i) A directed graph G associated with a stochastic process in continuous time. (ii)–(v) Subgraphs of G. Graphs (ii) and (iii) are spanning in-trees with the sink shaded. Graph (iv) is not a spanning in-tree with the indicated sink because there are no paths directed towards the sink from some of the other vertices. Graph (v) is not a spanning in-tree because it contains more than one path from the encircled configuration to the shaded sink.

very large as graphs reach a modest size which makes such computations difficult, although some methods have been suggested in this context [17].

2.3.5 Equivalence of the two detailed balance conditions

I now call upon the matrix tree theorem to derive the first detailed balance condition (2.3.2) from the second (2.3.3). Firstly, the special case $n = 2$ in (2.3.3) implies that for every directed edge $\omega_1 \to \omega_2$ in the graph G, the reverse edge $\omega_2 \to \omega_1$ is also present. In turn, this implies that the number of spanning trees with sink ω is the same for all ω, as one can verify by drawing the appropriate pictures.

Furthermore, there exists a one-to-one mapping $K : \mathcal{T}(\omega) \to \mathcal{T}(\omega')$ on every tree $T \in \mathcal{T}(\omega)$ that transforms it to a tree $T' \in \mathcal{T}(\omega')$. The effect of this mapping K on a tree T is to replace every directed edge $\omega_i \to \omega_j$ in that tree on the directed path from $\omega' \twoheadrightarrow \omega$ with the reverse edge $\omega_j \to \omega_i$ from G. Figure 2.5 illustrates this transformation.

The effect of this transformation on the weight of the tree is

$$w(T') = w(KT) = \frac{\mathcal{W}(\omega_1 \to \omega_2)\mathcal{W}(\omega_2 \to \omega_3)\ldots\mathcal{W}(\omega_{n-1} \to \omega_n)}{\mathcal{W}(\omega_n \to \omega_{n-1})\mathcal{W}(\omega_{n-1} \to \omega_{n-2})\ldots\mathcal{W}(\omega_2 \to \omega_1)} w(T) \tag{2.3.13}$$

where $\omega_1 = \omega$, $\omega_n = \omega'$ and ω_i, $1 < i < n$ are intermediate vertices along the path from ω to ω'.

To show that (2.3.3) implies (2.3.2) I consider the case where ω and ω' are adjacent (i.e. the transition $\mathcal{W}(\omega \to \omega') \neq 0$). Then for the path from $\omega \twoheadrightarrow \omega'$ in the tree $T \in \mathcal{T}(\omega)$ one can construct a loop by adding the single edge $\omega' \to \omega$. Using the relationship between the loops in each direction from the condition (2.3.3) one finds, for all trees $T \in \mathcal{T}(\omega)$, that

$$w(T') = \frac{\mathcal{W}(\omega_1 \to \omega_n)}{\mathcal{W}(\omega_n \to \omega_1)} w(T) = \frac{\mathcal{W}(\omega \to \omega')}{\mathcal{W}(\omega' \to \omega)} w(T) \,. \tag{2.3.14}$$

Thus

$$f(\omega') = \sum_{T \in \mathcal{T}(\omega')} w(T') = \frac{\mathcal{W}(\omega \to \omega')}{\mathcal{W}(\omega' \to \omega)} \sum_{T \in \mathcal{T}(\omega)} w(T) = \frac{\mathcal{W}(\omega \to \omega')}{\mathcal{W}(\omega' \to \omega)} f(\omega) \tag{2.3.15}$$

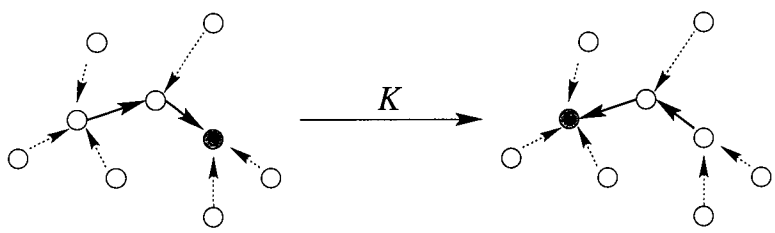

Figure 2.5: Transformation of a spanning tree from one sink to another (sinks shown shaded). Note the reversal of all edges on the path between the two sinks (solid lines).

or, because $P^*(\omega) \propto f(\omega)$,

$$P(\omega)\mathcal{W}(\omega \to \omega') = P(\omega')\mathcal{W}(\omega' \to \omega) \; . \tag{2.3.16}$$

This is none other than the detailed balance condition (2.3.2) and thus completes the proof that (2.3.2) and (2.3.3) are equivalent.

2.3.6 Equilibrium versus nonequilibrium steady states

The notion of detailed balance often arises in discussions of nonequilibrium steady states. Specifically, it is often stated that a lack of detailed balance implies a nonequilibrium steady state. However, the true situation is not quite so clear-cut, particularly because a precise definition of what constitutes an equilibrium distribution is not easily formulated.

For example, the equilibrium distribution for a system in contact with a heat bath (as described in the introduction, chapter 1) is the Boltzmann distribution

$$P^*(\omega) = \frac{\exp(-\mathcal{H}(\omega)/k_B T)}{Z_{eq}} \; . \tag{2.3.17}$$

Recall that the partition function $Z_{eq} = \sum_\omega \exp(-\mathcal{H}(\omega)/k_B T)$ and the Hamiltonian $\mathcal{H}(\omega)$ expresses the energy of the state ω. Naturally if the energy is well-defined for a system, one can use that to define $\mathcal{H}$. In a general stochastic process it is not always clear from a microscopic viewpoint what the energy is. Indeed, one could simply *define* the energy of a configuration as the logarithm of its steady-state probability. However, the resulting expression may not seem physically reasonable when compared with the transition rates $\mathcal{W}(\omega \to \omega')$ one prescribes when modelling a physical system.

It is, however, evident from (2.3.2) that the lack of detailed balance does imply a circulation of probability in configuration space in the steady state. That is, the probability being transported around a loop in one direction is not necessary equal to that being transported in the other direction. As this circulation feels at odds with one's intuitive idea of an equilibrium, it is perhaps justified to define a nonequilibrium system as one whose microscopic transition rates do not satisfy detailed balance, as was done in [16]. Certainly in the absence of detailed balance one does not have a general means to calculate the steady-state probability distribution (see section 2.5), although it is not impossible that the resulting expression could be written using a Hamiltonian which can be readily interpreted using physical arguments.

On the other hand, the detailed balance condition is of great utility when devising transition rates that are intended to converge to a known distribution, e.g. the Boltzmann distribution for an equilibrium system. These can then be used to study the relaxation of a model system to equilibrium, or in a Monte Carlo simulation to sample steady-state properties of a system which

cannot be treated analytically. In particular, because a set of rates that satisfy detailed balance implies that the process is ergodic, microscopic configurations are generated by a computer with frequencies proportional to their steady-state probability $P^*(\omega)$ and hence equilibrium properties can be effectively sampled, at least in the limit of infinite run times[4].

2.4 The thermodynamic limit

I now relax the assumption that a model system comprises only a finite number n of configurations. As I argue below, the reason for considering $n \to \infty$ (the thermodynamic limit) is because only in this limit can *phase transitions* arise. These dramatic changes in collective behaviour are manifested mathematically as singularities in a macroscopic variable as one varies a control parameter.

Recall equation (2.1.1) that relates a macroscopic observable $\mathcal{O}$ to a probability distribution $P_t(\omega)$. I now introduce a generic control parameter β which enters the microscopic model through the transition rates $\mathcal{W}(\omega \to \omega';\beta)$. Then, generalising the steady-state average of an observable (2.1.1),

$$\langle \mathcal{O}^*(\beta) \rangle = \sum_{\omega} P^*(\omega;\beta)\mathcal{O}(\omega) \; . \tag{2.4.1}$$

For a finite system we have seen that the steady-state probability distribution can be expressed as

$$P^*(\omega;\beta) = \frac{f(\omega;\beta)}{Z(\beta)} \tag{2.4.2}$$

where $Z(\beta) = \sum_\omega f(\omega;\beta)$ and $f(\omega;\beta)$ is some finite product of the transition rates $\mathcal{W}(\omega \to \omega';\beta)$. If the transition rates are analytic functions of β (which we assume to be the case), the probability distribution $P^*(\omega;\beta)$ and thus $\langle \mathcal{O}^*(\beta) \rangle$ are also analytic functions of β.

In order for any singularities to arise, one must take the thermodynamic limit, i.e. the system size to infinity. Then, although for each finite system size $\langle \mathcal{O}(\beta) \rangle$ is analytic, there is no requirement for the limit of this sequence of functions to be analytic.

Our intuition tells us that we need a singularity in $P^*(\omega;\beta)$ as one varies β. Such a singularity could arise if it so happened that $Z(\beta) \to 0$ as the system size is taken to infinity. The rigorous results of Yang and Lee make this idea more precise in the context of equilibrium phase transitions.

[4]For the real-world constraint of a finite run time, simulations need a certain amount of prodding to sample configuration space effectively. This, however, is the remit of an entirely different thesis [18].

2.4.1 Yang-Lee theory of equilibrium phase transitions

In two papers, Yang and Lee [19, 20] related phase transitions and singularities in an equation of state to the zeros of a grand canonical partition function $\mathcal{Z}(z, V)$ where z is a fugacity and V the volume of the system. The fugacity expansion $\mathcal{Z}(z, V)$ reads

$$\mathcal{Z}(z, V) = \sum_{N=0}^{\infty} z^N Z_N(V) \tag{2.4.3}$$

in which $Z_N(V)$ is the canonical partition function for fixed particle number N. Yang and Lee focussed specifically on systems for which the number of particles is bounded at a finite volume, e.g. as a consequence of hard-core exclusion. Then clearly, $\mathcal{Z}(z, V)$ is analytic in z for finite V.

Yang and Lee then went on to prove two theorems. The first states that the pressure in the infinite system

$$P(z) = \lim_{V \to \infty} \frac{kT}{V} \ln \mathcal{Z}(z, V) \tag{2.4.4}$$

is a continuous nondecreasing function of z providing the surface area of V increases no faster than $V^{2/3}$ (which will, of course, usually be satisfied). Their second theorem states that if one can find a region R in the complex-z plane that includes a portion of the positive real axis and contains no zeros of the grand partition function, then $P(z)$ converges uniformly to its limit as $V \to \infty$ for all $z \in R$, and this limit is analytic.

In other words, one can locate a transition point as follows. Consider a region R as defined above including a point z_0 on the positive real axis (recalling that a region is an open subset of $\mathbb{C}$). If one finds that, no matter how small R is, the grand partition function $\mathcal{Z}(z, V)$ contains at least one zero as $V \to \infty$, the point z_0 is a transition point between two phases. Alternatively, the zeros of the grand partition function converge onto the transition point z_0 on the real axis as $V \to \infty$.

Either side of this transition point, one can find regions which contain some portion of the real axis on which the equation of state is analytic. At the transition point itself, the equation of state may exhibit a discontinuity in one of its derivatives (which correspond to different orders of phase transition). See figure 2.6 for an illustration.

A natural question to ask at this point is whether these ideas are extendable to more general stochastic processes, and in particular those that do not satisfy detailed balance. As was shown in section 2.3.2 one can define a normalisation that is analogous to a partition function for any ergodic process, and so it seems plausible that zeros of this normalisation could exhibit similar properties to those of the equilibrium partition function. I do not speculate more on this subject at this point, but later present some evidence for extension to nonequilibrium

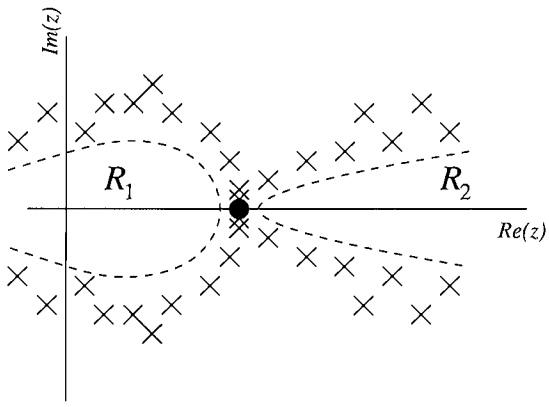

Figure 2.6: Zeros (× symbols) of a (hypothetical) grand partition function in the complex fugacity plane. Regions R_1 and R_2 (which contain no zeros) correspond to phases in which the equation of state is analytic; the zeros converge to the point marked by the filled circle, which is a phase transition point in the fugacity z.

systems in the context of exclusion processes (section 4.8.4) and models with absorbing states (section 5.4).

Note, however, that the convergence of the normalisation Z to zero implies through (2.3.11) the simultaneous convergence to zero of at least one eigenvalue of the transition rate matrix W (in addition to the zero eigenvalue that is always present). Intuitively one could perhaps view this eigenvalue behaviour as corresponding either to diverging timescales ('critical slowing down') as a continuous phase transition or to phase coexistence at a first-order phase transition.

2.4.2 Stationary states in the thermodynamic limit

I now discuss the properties of stationary states in stochastic processes with an infinite number of states. Here the situation becomes more complicated than that described earlier in this chapter, as it involves infinite-dimensional matrices and measure theory. I shall therefore only make a few general observations.

Considering the case of a countably infinite set of configurations (as would be appropriate for a lattice model), one classifies configurations according to the probability that they are returned to at some subsequent time. This probability is related to the existence of paths (sequences of transitions) between configurations. If for every path that exists from a state ω to a second state ω' there also exists a path from ω' to ω, the probability of returning to the state ω is unity and the state is called *recurrent*. The mean time taken to return is denoted $\tau(\omega)$ and may be infinite, in which case the state is called *null recurrent*. When the return probability is less than one, the state ω is called *transient*.

The main result here is that if all states are recurrent, there is a unique steady state given by

$$P^*(\omega) = \lim_{t\to\infty} P_t(\omega) = \frac{1}{\tau(\omega)} \tag{2.4.5}$$

which can be interpreted as the frequency with which the state ω is visited in the steady state [11]. This fact is the foundation of the ergodic theorem, which states that an infinite time average is equivalent to an ensemble average in the steady state. Furthermore, it can be shown [11] that finite 'truncations' of the infinite system exhibit steady-state properties indicative of those of the infinite system.

For the more general class of systems with transient states there are fewer strong statements that can be made. In chapter 5 I consider models with an *absorbing state*, i.e. a configuration that once entered cannot be left. Such systems can exhibit two types of behaviour in the thermodynamic limit: either the absorbing state is reached with certainty, or there is some probability that the process remains *active* forever (i.e. the absorbing state is not reached). As will be seen in chapter 5, the infinite-time behaviour of the finite-size and infinite-size systems differ, and one should be aware that the limits of infinite system size and infinite time need not always commute.

2.5 How to solve a stochastic process

In this chapter I have concentrated mainly on general aspects of stochastic processes with particular emphasis on steady states and detailed balance. I have, however, largely avoided the subject of solving a stochastic process, i.e. obtaining some analytical solution for a quantity of interest. Given that we are particularly interested in dynamics that do not satisfy detailed balance, we do not know *a priori* even the steady-state probability distribution of configurations. In this final section of the chapter, I take a brief look at some of the methods one can use to determine the behaviour of stochastic processes.

2.5.1 Exact solution of the master equation

Exact solutions are highly sought after in physics as they yield complete information about a system in hand. In the course of the present investigation of stochastic processes we have already come across a couple of methods that, in principle, allow one to determine an exact solution for a process. Firstly, one can attempt a graphical solution, e.g. the enumeration of paths from an initial to a final configuration (as described in section 2.1.2), or of spanning in-trees (section 2.3.4). Generally these approaches are tractable for large system sizes only if one can identify some convenient structure in the graph.

Secondly, one can try and find the eigenvectors of a transition matrix. Again, standard techniques (such as Cramer's rule, Gauss-Jordan elimination etc.) become harder to apply as the system size becomes large. More efficient approaches usually rely on the identification of a suitable ansatz. Two particular ansätze are applicable to a class of one-dimensional lattice-based processes in which the number of particles occupying a single lattice site is restricted. Exclusion processes, the subject of chapter 4, fall into this class.

The first of these techniques is the matrix product method for the steady state. This method has been shown to be equivalent to finding the solution of the linear set of equations defined by a master equation (2.2.7) in the steady state [21]. The way it works is as follows. For each site on a one-dimensional lattice one associates a "state" (occupation) number, drawn from a discrete set. Then for each of these states τ one introduces a matrix A_τ. The steady-state probability $P^*(\omega)$, where $\omega = (\tau_1, \tau_2, \ldots, \tau_N)$ with N the lattice size, is then a scalar derived from the ordered matrix product $A_{\tau_1} A_{\tau_2} \ldots A_{\tau_N}$. The matrices themselves satisfy relations which can be obtained from the master equation. I do not discuss this method further here, as chapter 4 provides a detailed example yielding new results.

Alternatively, one can try a Bethe ansatz, which also admits solution for nonstationary states. This technique was originally used in the context of a spin-chain model for magnetism and is suitable for a one-dimensional lattice system in which the number of particles is conserved. One then proposes the following form for eigenvectors ϕ of the transition rate matrix W:

$$\phi(x_1, x_2, \ldots, x_M) = \sum_Q A(Q) z_1^{x_{Q(1)}} z_2^{x_{Q(2)}} \cdots z_M^{x_{Q(M)}} \tag{2.5.1}$$

in which x_i is the position of particle i on a lattice, and $1 \leq x_1 < x_2 < x_3 \ldots < x_M \leq N$ with N the lattice size once again. The sum is over all permutations Q of $(1, 2, \ldots, M)$ and $A(Q)$ is some amplitude. Again one uses the master equation to determine relations that must be satisfied by $\phi(x_1, \ldots, x_M)$ in its unphysical regions, e.g. where two particles occupy the same site and $x_i = x_{i+1}$. After some manipulation, one arrives at a set of nonlinear equations that must be simultaneously satisfied by all the 'momenta' z_i. An early account of this method in the context of stochastic processes is given in [22] and it is also described in [23]. More sophisticated variants of the Bethe ansatz also exist, and an introduction to these can be found in [24] and references therein.

2.5.2 Alternative formulations

When no obvious means to solve the master equation (2.2.7) can be found, it can be more fruitful to derive from it an alternative exact equation which one can attempt to solve instead. One such equation is a differential equation for a macroscopic observable. Recall that if one

can define an observable $\mathcal{O}(\omega)$ for each configuration ω, its expectation value is

$$\langle \mathcal{O}_t \rangle = \sum_{\omega} P_t(\omega)\mathcal{O}(\omega) \; . \tag{2.5.2}$$

The master equation (2.2.7) implies

$$\frac{\mathrm{d}}{\mathrm{d}t}\langle \mathcal{O}_t \rangle = \sum_{\omega}\sum_{\omega'} P_t(\omega')\mathcal{W}(\omega' \to \omega)\mathcal{O}(\omega) \tag{2.5.3}$$

in which one should be aware that the self-transition rate $\mathcal{W}(\omega \to \omega)$ is negative and whose magnitude is the total rate at which it is possible to leave the configuration ω.

As an example, let $\omega \equiv x \in \mathbb{Z}$, where x is the position of a random walker on an infinite lattice with lattice spacing one. Also let the random walker hop one site to the right $x \to x+1$ at rate p and one site to the left $x \to x-1$ at rate q. Then

$$\mathcal{W}(x' \to x) = \begin{cases} p & x' = x-1 \\ q & x' = x+1 \\ -(p+q) & x' = x \\ 0 & \text{otherwise} \end{cases} \tag{2.5.4}$$

Additionally, the observable $\tau(y)$ is defined as the occupancy of site y, i.e. $\mathcal{O}(x) = \tau(y;x) = \delta(x,y)$. Inserting these definitions into (2.5.3) one obtains

$$\begin{aligned} \frac{\mathrm{d}}{\mathrm{d}t}\langle \tau_t(y) \rangle &= p\sum_x P_t(x)\tau(y-1;x) + q\sum_x P_t(x)\tau(y+1;x) - (p+q)\sum_x P_t(x)\tau(y;x) \\ &= p\langle \tau_t(y-1) \rangle + q\langle \tau_t(y+1) \rangle - (p+q)\langle \tau_t(y) \rangle \end{aligned} \tag{2.5.5}$$

which is precisely the equation one would immediately write down for such a simple process. However the method generalises and using it one can derive *exact* differential equations for any desired macroscopic quantity. This procedure will be followed a number of times in the course of this thesis.

Simple-minded mean-field approximation

Often one finds that the time-derivative of a one-point quantity (e.g. the density at a lattice site) derived in the manner described above depends on two-point quantities (e.g. density correlations). Then repeating the procedure for the two-point quantities one finds dependence on three-point quantities and so on. That each equation derived depends on a new unknown means that one cannot *close* the set of equations.

One way in which to deal with this situation approximately is to replace two-point function with the product of two one-point functions. For example, in the case of densities

$$\langle \tau_t(y)\tau_t(y') \rangle \to \langle \tau_t(y) \rangle \langle \tau_t(y') \rangle \; . \tag{2.5.6}$$

Clearly this is an approximation which neglects correlations and fluctuations of a random variable and, as such, is often called a *mean-field approximation.*

Approximations of this type can generally be used to make qualitative predictions as to the macroscopic behaviour of a model, but often fail to make correct quantitive predictions. However, there are situations in which the mean-field approximations can be shown to be exact. An example is provided in section 6.1.1 in which I show that a lattice-based model is governed by mean-field equations of motion in a specific limit.

Taking fluctuations into account

Rather than simply ignore fluctuations and correlations, as was done above, it is more useful to derive differential equations that explicitly include them. One scheme for doing this where a random variable x_t is continuous (e.g. the off-lattice position of a particle) is to Taylor expand the master equation (2.2.7) about the 'current' particle position x_t. The expansion involves the set of *jump moments*

$$a_n(x) = \int \mathrm{d}x' \, \mathcal{W}(x \to x') \, (x' - x)^n \tag{2.5.7}$$

in which $\mathcal{W}(x \to x')\mathrm{d}x'$ is the rate at which a particle hops from a point x to a point in the interval $[x', x' + \mathrm{d}x']$.

The resulting equation, with the Taylor expansion truncated to second order, is the Fokker-Planck equation

$$\frac{\partial}{\partial t} P_t(x) = -\frac{\partial}{\partial x} \left[a_1(x) P_t(x) \right] + \frac{1}{2} \frac{\partial^2}{\partial x^2} \left[a_2(x) P_t(x) \right] \, . \tag{2.5.8}$$

It should be noted that this is an approximation to the original master equation, and that full details of its derivation can be found in e.g. [25] or [26].

It can also be shown [25] that (2.5.8) is equivalent to a Langevin equation

$$\frac{\mathrm{d}}{\mathrm{d}t} x_t = \left[a_1(x_t) - \frac{1}{4} a_2'(x_t) \right] + \left[a_2(x_t) \right]^{1/2} \eta_t \tag{2.5.9}$$

in which $a_2'(x)$ is the derivative of $a_2(x)$ with respect to x and η_t is a random variable that has mean of zero, unit variance and is uncorrelated. That is

$$\langle \eta_t \rangle = 0 \quad \text{and} \quad \langle \eta_t \eta_{t'} \rangle = \delta(t - t') \, . \tag{2.5.10}$$

There are a couple of points to note about the Langevin equation (2.5.9). Firstly, both η_t and x_t are random variables and so the Langevin equation expresses the relation between two stochastic processes. The usual interpretation of a Langevin equation is that the first term in square brackets is the deterministic part of the evolution of x_t, and the noise term

(containing η_t) represents fluctuations about the mean behaviour. The presence of a function of x_t multiplying the noise η_t means that the differential equation is meaningless as it stands: it must be supplemented by a scheme for integrating the right-hand side. Two commonly used schemes are Ito and Stratonovich integration. Their relevance to stochastic processes is detailed in [27]; I simply state here that equations (2.5.8) and (2.5.9) are equivalent under Stratonovich integration. In this case, one integrates a random function by dividing time into N small intervals Δt_i, and using the values of function at the midpoints of these intervals to approximate the integral with a sum. Finally, the integral is obtained by simultaneously taking the limits $N \to \infty$ and $\Delta t_i \to 0$ in the summation.

In this thesis, a Langevin equation will be used to model interfacial growth (see chapter 6). Other examples and a more general treatment of Langevin equations can be found in textbooks such as [25, 26, 27].

The master equation and the Schrödinger equation

The final alternative form of the master equation (2.2.7) that I wish to discuss is not so much a reformulation as a reinterpretation. Consider again the master equation in matrix form

$$\frac{\mathrm{d}}{\mathrm{d}t}\boldsymbol{P}_t = W\boldsymbol{P}_t \tag{2.5.11}$$

and compare with the Schrödinger equation

$$i\hbar\frac{\mathrm{d}}{\mathrm{d}t}|\psi(t)\rangle = \mathcal{H}|\psi(t)\rangle \ . \tag{2.5.12}$$

Note that, up to notational differences, the two have the same basic structure. In quantum mechanics the time-dependence of the wavefunction $|\psi\rangle$ is governed by the Hamiltonian (energy of the system in terms of operators) $\mathcal{H}$. Therefore, the transition rate matrix W is often referred to as a Hamiltonian; however this use of the term does not imply any relationship to the energy of a process, even if such a thing can be defined.

Additionally, some differences between the nature of the Hilbert space spanned by the wavefunctions $|\psi\rangle$ and the configuration space of $\boldsymbol{P}_t$ should be noted. Firstly, in quantum mechanics the elements of the vector $|\psi\rangle$ are complex probability amplitudes, and it is the squares of their absolute values which are probabilities. In stochastic processes, the elements of $\boldsymbol{P}_t$ themselves are probabilities, and hence real and nonnegative. Further, in quantum mechanics, each of the eigenvectors of $\mathcal{H}$ is a stationary state; in stochastic processes, only those eigenvectors with zero eigenvalue are stationary. Finally, in quantum mechanics, $\mathcal{H}$ is Hermitian, and so the vector dual to $|\psi\rangle$ is just its complex conjugate transpose. In stochastic processes, W is not necessarily symmetric, and thus the vectors $\boldsymbol{P}_t$ are not simply related to their duals. In

particular, the elements of the left and right eigenvectors of W need not take similar forms in a general representation.

Nevertheless, an operator formulation of the master equation, particularly for lattice-based processes where particle moves are conveniently written using creation and annihilation operators, can be beneficial. Firstly, connections can be made to other fields e.g. that of quantum spin chains [28]. Additionally, the use of a 'second-quantised' description (i.e. creation and annihilation operators) allows one to recast the master equation as a path integral which can then be analysed using the methods of quantum field theory. For descriptions of this procedure in the context of stochastic processes, see e.g. [29, 30, 31].

2.5.3 Numerical methods and simulation

Even if one writes the master equation for a process in a different form, it may, and probably still will, be unyielding to an analytical solution. In this instance one may wish to turn to numerical methods in order to e.g. diagonalise the transition rate matrix W or solve a Langevin equation (2.5.9). This in itself is a nontrivial exercise, especially in the case of numerical diagonalisation as the system need not comprise many configurations before there is insufficient memory to hold a probability vector, let alone the entire matrix W. Here one must start to use optimisation techniques, e.g. those appropriate to sparse matrices (i.e. where many transition rates are zero), or approximation schemes, such as the density matrix renormalisation group [32] in which a subset of the entire system is diagonalised, and a subset of the eigenvectors found retained and used to construct eigenvectors for a larger system.

Finally, much can be gleaned by performing computer simulations of a process. In principle these are quite easy to set up, as all one needs do is generate a random walk through configuration space according to the transition rates $\mathcal{W}(\omega \to \omega')$. Again, however, simulation is easier said than done. Usually one needs large system sizes to see the desired physical effects; often one also needs long run times (especially when correlation timescales begin to diverge). Also, as is the case when we are studying nonequilibrium probability distributions, we do not know in advance what the steady-state distribution $P^*(\omega)$ is, so it becomes more difficult to devise schemes for sampling this distribution efficiently. This is in contrast to the case of an equilibrium system where one can vary the transition rates $\mathcal{W}(\omega \to \omega')$ so that detailed balance with respect to the desired steady-state distribution $P^*(\omega)$ is obeyed (2.3.2), but with transitions to the region of configuration space of interest that are greatly enhanced.

2.6 Summary

In this chapter I have reviewed how one devises a mathematical model of a stochastic process and some general properties such as the existence of steady states, ergodicity and detailed balance. In respect of the latter, I have adopted the loose definition of a nonequilibrium system as one whose microscopic dynamics do not satisfy detailed balance. Such systems form the basis of the models that will be studied in greater depth in the remainder of this thesis.

Meanwhile, it should be apparent that there is no single, foolproof method for solving a nonequilibrium system. Thus in the following chapters, the methods described in the previous section will be called upon in a rather *ad hoc* manner. Nevertheless, this approach will allow us to achieve the stated aim of gaining some understanding of collective behaviour exhibited by nonequilibrium systems, at least for the restricted class that is considered in this thesis.

Chapter 3

Annihilation and coalescence reactions

The first type of stochastic dynamics I consider is that of particle reactions. Specifically I discuss single-species annihilation and coalescence reactions in which the collision of two particles may result in one or both reactants being removed from the system. I begin by reviewing the results known for the case where the reactant motion is diffusive, concentrating on the equivalence of the two types of reaction. I then move on to the case of ballistic (deterministic) reactant motion, and find from an original exact solution that annihilation and coalescence are equivalent in this case too. These results are obtained by exploiting a connection to the combinatorics of reordering (sorting) processes, which is described in detail.

3.1 Particle reaction systems

3.1.1 Overview

I set the scene with a physically reasonable picture of a particle reaction system. Consider some kind of container in which there are particles of a number of different species, labelled $A, B, C \ldots$. Now let them explore the box according to some suitably defined equations of motion until two collide and react with some prescribed probability that depends on the species involved in the collision. For example, if species A and B combine upon meeting to create a product C with probability p, one would write

$$A + B \xrightarrow{p} C \,. \tag{3.1.1}$$

A two-dimensional reaction system in which there are two reactants A and B which undergo the reaction $A + A \xrightarrow{p} B$ is illustrated in figure 3.1.

Although this picture is extremely simple, a few remarks are already in order. Clearly the number of particles of a given species will change with time through the reactions and often it

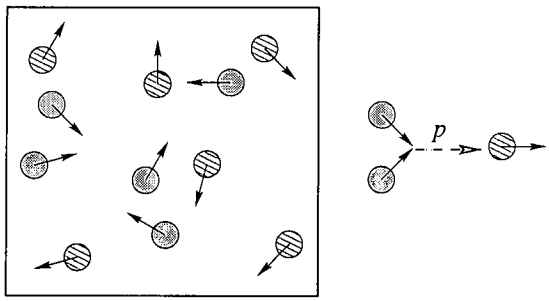

Figure 3.1: The standard picture of a reaction system. In this case, type A reactants (shaded) combine to give rise to type B reactants (hatched) with probability p on meeting.

is this time-dependence of a particle species' density which is most easily computed for such systems. Furthermore, the reaction dynamics may be irreversible: in the above example once A and B particles have combined to form a C particle, the original reactants are lost forever. Although I will not go so far as to formulate an explicit master equation for a particle reaction system, it can easily be seen from condition (2.3.3) in the previous chapter that an irreversible reaction implies the stochastic reaction process does not satisfy detailed balance. Hence, particle reaction dynamics belong to the loosely-defined class of nonequilibrium dynamics of the previous chapter (see section 2.3.6).

By far the most studied reactions are single-species annihilation $A + A \to \emptyset$ and coalescence $A + A \to A$ reactions, and only these reactions are considered in this chapter. Before I outline some of the established results for this pair of reactions, I briefly illustrate a physical context in which they arise, namely the dynamics of the q-state Potts model.

The Potts model is a generalisation of the Ising model in which the magnetic moments (spins) s_i may point in one of q directions. In the absence of a field, the system is described by a Hamiltonian

$$\mathcal{H} = -J \sum_i (\delta(s_i, s_{i+1}) - 1). \tag{3.1.2}$$

which simply states that there is an energy cost J associated with each neighbouring pair of nonaligned spins. Thus, the ground state will comprise a system in which all spins are aligned, although the Hamiltonian does not favour a particular direction. In order to map the model onto a reacting particle system, one must examine the approach under zero-temperature Glauber dynamics [33] to a ground state from some random initial configuration of spins.

These dynamics are implemented as follows. First one randomly picks a spin, and also one of its neighbours; then, if the two spins are differently oriented, the first spin is flipped so that it becomes aligned with its neighbour—see figure 3.2. In one dimension, this process can be seen always to minimise (or leave unchanged) the local nearest-neighbour interaction energy due to the flip of one spin alone. Alternatively, one can view each spin as an entity which may have

one of a fixed number q of opinions. Not wishing to stand out of the crowd, each entity chooses a neighbour's opinion on the matter in hand entirely at random. Such a scenario describes a voter model, see e.g. [34] for further discussion.

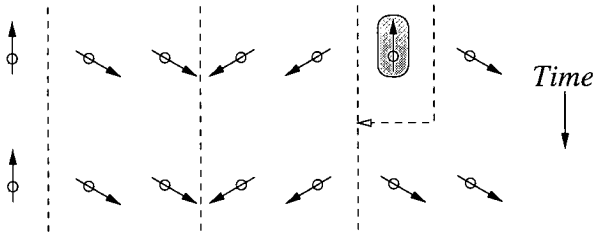

Figure 3.2: A $q = 3$ states Potts model evolving under zero-temperature Glauber dynamics. The highlighted spin is flipped so that it becomes aligned with its right nearest neighbour. This causes two domain walls (shown as dashed lines) to coalesce. To map to a particle reaction system, one interprets the domain walls as particles.

In figure 3.2 the evolution of the domain walls—lines between nonaligned spins—as a result of a single spin-flip is shown. In the figure, a randomly chosen spin has been made to align with its right neighbour, causing the domain wall between them to move one site to the left where it has coalesced with the domain wall already present there. Had the chosen spin's left neighbour also pointed the same way as that to the right, the two domain walls would have annihilated rather than coalesced. By interpreting the domain walls as particles, one obtains a one-dimensional annihilation and coalescence model.

The motion of the particles is diffusive: that is the domain walls hop with equal probability to the left or right because either spin next to a domain wall may be selected and flipped with equal probability. One can also calculate the probability that two particles annihilate or coalesce on contact under the assumption that each possible initial spin configuration was chosen with equal likelihood. Then the orientations of the spins on either side of a domain wall are statistically independent at all times (apart from the fact that they must, necessarily, be different). Hence, the probability for two next-nearest neighbouring domains to have their spins pointing in the same direction is $1/(q-1)$ and therefore to point in different directions is $(q-2)/(q-1)$. In the former case, two approaching domain walls annihilate whereas in the latter they coalesce. This implies that the one-dimensional q-state Potts model evolving under zero-temperature Glauber dynamics can be expressed as a single species particle reaction system described by

$$A + A \xrightarrow{p} \emptyset \quad \text{and} \quad A + A \xrightarrow{1-p} A \tag{3.1.3}$$

with $p = 1/(q-1)$.

3.1.2 Universality of diffusive annihilation and coalescence

In the foregoing we have seen how a combination of particle diffusion combined with annihilation and coalescence reactions emerges from the q-state Potts model. Intensive study of these two reactions (see [35, 36] for reviews) has shown that the mean (in the sense of space and ensemble averaged) density of particles, as well as all n-point density-density correlation functions [37], are identical for the two reactions, and also any combination thereof.

Consider in particular the case of the mean density as a function of time. If one assumes that the reactants are well-mixed at all times (i.e. the density $\rho(x,t)$ is the same everywhere) and further neglects fluctuations and correlations (i.e. a mean-field approximation, see section 2.5.2) one quickly arrives at the rate equation

$$\frac{\mathrm{d}}{\mathrm{d}t}\rho(t) = -\lambda\rho(t)^2 \,. \tag{3.1.4}$$

for the density. This equation simply expresses the fact that, under the mean-field approximation, the probability that two particles are at the same point is $\rho(t)^2$. Then, the rate at which a particle is removed from the system is proportional to λ (which in turn depends on the annihilation and coalescence rates). The solution to (3.1.4) is

$$\rho(t) = \frac{1}{\lambda t} \,. \tag{3.1.5}$$

This result provides the first (albeit rather simple) example of a dynamical scaling regime in this thesis. Consider the typical distance between two particles $\ell \sim \rho^{-1/d}$ where d is the number of spatial dimensions in the model. Then the above implies that $\ell \sim t^{1/d}$ and so the system is invariant under the rescaling transformation $t \to bt$ and $\ell \to b^{1/d}\ell$. Note however that this transformation is trivial in the sense that a homogeneous spatial distribution of reactants has been assumed.

The rate equation (3.1.4) is not expected to describe the system accurately in low dimensions because reaction events induce density fluctuations. As the dimensionality is decreased, the dynamics become more sensitive to these fluctuations and indeed from exact solutions for the one-dimensional case (see e.g. [38, 39, 40]) it has been found that

$$\rho(t) \sim \frac{1}{\sqrt{8\pi D t}} \tag{3.1.6}$$

for large times given $\rho(0) = 1$. Note that this result applies for *both* annihilation and coalescence reactions in the presence of diffusive motion with diffusion constant D.

Information about the relationship between the two reaction processes can be gleaned through a field-theoretic treatment. As discussed in section 2.5.2 one can interpret the master equation for a stochastic process as a Schrödinger equation and this allows one to express expectation

values of observables as path integrals—see e.g. [29, 30, 31]. Peliti [41] was able to show that this description of diffusive annihilation and coalescence reactions can be renormalised exactly to all orders in perturbation theory and confirmed the conjecture that the two types of reaction belong to the same universality class. Furthermore, the upper critical dimension, above which the mean-field result (3.1.5) holds, was shown to be two. At this upper critical dimension one obtains a density decay proportional to $\ln(t)/t$, again identical for both processes. The decay exponent of $\frac{1}{2}$ turns out to hold also for 'higher-order' single species processes i.e. $kA \to \emptyset$ [42] (although here the upper critical dimension depends on k) but not for the two-species process $A + B \to \emptyset$ [43].

Another way in which the correspondence between the reaction processes $A + A \to A$ and $A + A \to \emptyset$ is understood is through a similarity transformation between their master equations. That is, there exists a one-to-one mapping (given explicitly in [44, 45]) from any probability distribution of configurations in the annihilation process to a probability distribution in the coalescence process. Although such a transformation does not imply a simple relationship between macroscopic observables (as defined by equation (2.1.1)) in the two systems, it turns out that the density and density-correlation functions are closely related in the case of annihilation and coalescence reactions [37].

3.1.3 The case of ballistic motion

Until now, I have described only reactants that move diffusively between reactions. This is the case if the number of collisions a particle suffers before reacting is large, thereby randomising its velocity. If, on the other hand, the mean-free path of the reactants is much greater than the typical interparticle spacing, then ballistic (straight-line) motion is deemed appropriate.

Ballistic coalescence and annihilation reactions can also be mapped onto a range of other processes, notably an interfacial growth model (see [46] and also section 6.2.1) as well as certain limits of traffic flow [47] and competition between warring factions [48]. Here I do not consider these applications but instead the more fundamental question of whether there is any relationship between annihilation and coalescence reactions when the particle motion is ballistic rather than diffusive.

This question is made harder to answer due to the fact that, as yet, no general approach to the solution of ballistic reaction systems has emerged. Principally this is because any model one devises must explicitly keep track of all particle velocities and positions in the system. A seminal model was introduced and solved exactly by Elskens and Frisch [49] (also it was independently studied in [46]) and describes deterministic pairwise annihilation of oppositely moving particles in one dimension—see figure 3.3. In that case it was found that if the initial fractions of the particles moving in each direction are equal, a power-law density decay

proportional to $t^{-1/2}$ is obtained, whereas for unequal initial densities an exponential density decay modulated by a power-law $t^{-3/2}$ applies. Hence, even in this simple model, one finds a power-law decay *only* for particular initial conditions, a fact that contrasts starkly with the case of diffusive motion.

Generalisations of the Elskens-Frisch (EF) model have been many and varied, as have the theoretical approaches employed. The extension to three and four velocities has been considered using a mean-field theory, geometrical arguments and computer simulation [50] and a general feature is that the different particle species decay at different rates which depend on the initial conditions. Continuous velocity distributions have also been treated, most commonly using a Boltzmann equation formalism (see e.g. [51]) and also through scaling arguments [52]. Again it emerges that the long-time behaviour (e.g. the exponents of the power-laws) is strongly dependent on the initial distribution of velocities. This behaviour can be intuitively understood from the fact that the use of deterministic equations of motion implies that there will be a long-term memory of initial particle velocities.

Since Elskens and Frisch's work in 1985, very few other exact results had been obtained for ballistic reaction systems. In this chapter I obtain new exact results for a generalisation of the EF model which includes any combination of annihilation and coalescence reactions, arbitrary initial densities of the two particle species and *stochastic* reaction dynamics (in the EF model two approaching particles always annihilate one another). This stochasticity models the physically reasonable idea of a finite reaction rate, as in figure 3.1.

The key feature of the model that allows it to be solved is that it is equivalent to a sorting (reordering) problem. To understand this connection consider figure 3.4 in which a realisation of particle trajectories in a two-species stochastic reaction system is shown. It is clear that the system evolves by two particles with opposite (positive and negative) velocities approaching and then either exchanging places (if no reaction occurs) or reacting to cause at least one of the particles to be removed from the sequence. Ultimately, a configuration is reached in which no

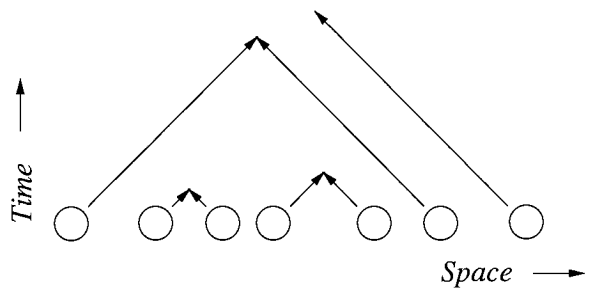

Figure 3.3: The ballistic reaction model of Elskens and Frisch [49]. In it particles have one of two velocities with their relative initial proportion a parameter of the model. Whenever two particles meet, they always react causing both to be removed from the system.

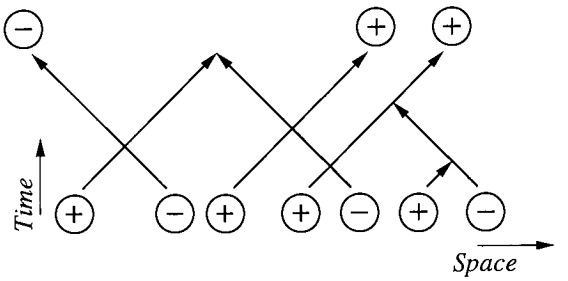

Figure 3.4: The model of stochastic ballistic annihilation and coalescence solved in this chapter. The model comprises particles with positive ⊕ and negative ⊖ velocities. When two meet both, one or neither particle may be removed from the system, each with a prescribed probability. No more reactions are possible once all particles are receding.

more reactions can occur because all particles with negative velocity are receding from those with positive velocity.

Hence the one-dimensional reaction system is equivalent to the sorting of particles into ascending order of velocities, albeit with the possibility of particle removal *en route* to the final configuration. In the next section I review some of the ideas associated with reordering problems, and this will serve as an introduction to mathematical entities known as q-series which arise in a range of contexts including number theory and combinatorics. It had already been noted [53, 54] that q-series arise in special cases of the stochastic ballistic reaction system that I solve in the general case here. The interpretation of a reaction system as a sorting process allows us to understand why q-series are important in the context of ballistic reaction dynamics.

3.2 The combinatorics of sorting

3.2.1 Sorting a set of distinct objects

The combinatorics of sorting is concerned with questions of the following type. Given an arbitrary permutation of N distinct objects $K_1, K_2, \ldots K_N$ which possess an ordering relation (i.e. the assertion $K_i < K_j$ is always either true or false for $i \neq j$ and that $K_i < K_j, K_j < K_k \Rightarrow K_i < K_k$) how much rearrangement is required to bring these objects into a prescribed order (e.g. one in which the smallest K appears first, the next smallest second and so on)?

Such a question is of particular interest to computer scientists wishing to compare the efficiency of different sorting algorithms—indeed a good account of some of the material described here can be found in the standard work on sorting and searching algorithms by Knuth [55]. To answer this question, two points must first be clarified. Firstly one must define quantitively the amount of rearrangement required to bring the initial sequence of objects into the prescribed

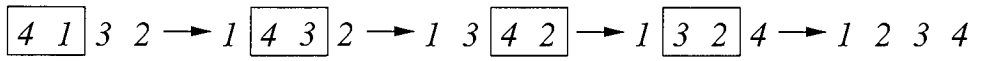

Figure 3.5: Reordering a permutation of four objects into ascending order through a sequence of nearest-neighbour swaps. In each step, the pair about to be exchanged is boxed.

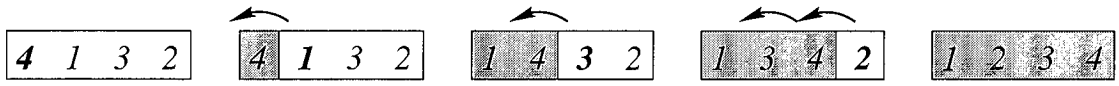

Figure 3.6: Systematic iterative algorithm for reordering a sequence of objects using the smallest number of nearest-neighbour exchanges. The shaded box indicates the objects that are already correctly sorted amongst themselves; only the first object after this (shown bold) is moved at each step of the process.

order: a convenient choice is the smallest number of nearest-neighbour exchanges required to obtain the desired order, see figure 3.5. Secondly this number of swaps varies according to the initial ordering of the objects and so one considers the possibilities that arise from considering all $N!$ initial permutations at once.

One way in which an initial sequence of N objects can be ordered using only the minimum number n of nearest-neighbour swaps is through an incremental algorithm as follows. First one notes that if the first $n-1$ objects in the sequence are already correctly ordered, the n^{th} object needs to be exchanged only with those 'larger' objects that immediately precede it. Hence the sequence can be sorted by iterating n from 1 to N. This process[1] is illustrated in figure 3.6.

A convenient way to keep track of the number of swaps is through the generating function

$$R_N(q) = \sum_n a_n \, q^n \tag{3.2.1}$$

in which a_n is the number of the initial $N!$ sequences requiring exactly n swaps to become ordered. The above sorting algorithm implies that

$$R_N(q) = (1 + q + q^2 + \cdots + q^{N-1}) R_{N-1}(q) = \frac{1 - q^N}{1 - q} \, R_{N-1}(q) \tag{3.2.2}$$

because the N^{th} object may have to exchange places with its left neighbour between 0 and $N-1$ times to reach its rightful location, and each of these possibilities is represented in the generating function by an additional term in q to the power of the number of swaps required. As a sequence comprising $N = 1$ objects is always correctly ordered, we have $R_1(q) = 1$. Hence the previous equation implies

$$R_N(q) = \frac{1}{(1-q)^N} \prod_{k=1}^{N} (1 - q^k) \tag{3.2.3}$$

[1]This algorithm is known to computer scientists as a bubble sort and is widely regarded as the least efficient sorting algorithm; however more sophisticated algorithms that employ long-range swaps do not have a phyisical interpretation in terms of reacting particles.

from which, in principle, one can obtain the number a_n of initial permutations of N objects that require n swaps to be brought into the right order by reading off the coefficient of q^n.

In practice, information about the statistics of the sorting process can be determined directly from the generating function $R_N(q)$ without having to determine explicitly the individual coefficients a_n. As an example, I show how one can manipulate (3.2.3) to find the mean number of exchanges required to order an initial sequence chosen with equal probability from the $N!$ possibilities. This mean is given from (3.2.3) as

$$\langle n \rangle = \frac{1}{N!} \sum_n n\, a_n \,. \tag{3.2.4}$$

Noting that

$$\sum_n a_n \equiv \lim_{q \to 1} R_N(q) = N! \,, \tag{3.2.5}$$

the mean number of swaps can be recast as

$$\langle n \rangle = \lim_{q \to 1} \frac{1}{R_N(q)} \frac{\mathrm{d}}{\mathrm{d}q} R_N(q) = \lim_{q \to 1} \frac{\mathrm{d}}{\mathrm{d}q} \ln R_N(q) \,. \tag{3.2.6}$$

Inserting (3.2.3) into this expression one finds

$$\langle n \rangle = \lim_{q \to 1} \sum_{k=1}^{n} \frac{1 - q^k - k(1-q)q^{k-1}}{(1-q)(1-q^k)} = \frac{1}{2} \sum_{k=1}^{N} (k-1) = \frac{1}{4} N(N-1) \tag{3.2.7}$$

in which the limit is computed using two iterations of L'Hôpital's rule. In actual fact, this particular quantity can be calculated more easily from the definition of the iterative sorting process given above; however this example illustrates the fact that a generating function can be a useful object in its own right.

Relationship to the q-factorial

I now make a few general observations about the structure of generating function $R_N(q)$ given by (3.2.3). We have already seen that in the limit $q \to 1$, $R_N(q) \to N!$ and so for general q this generating function can be thought of as a continuous deformation of the factorial. Indeed, the product over k on the right-hand side of (3.2.3) forms the conventional definition of the *q-factorial* (see, e.g. [56, ch. 1] or [57, ch. 10]) sometimes denoted $N!_q$ or $(q;q)_N$. That is, the q-factorial defined as

$$N!_q \equiv (q;q)_N \equiv \prod_{k=1}^{N} (1 - q^k) \,, \tag{3.2.8}$$

is equivalent to the ordinary factorial in the limit

$$\lim_{q \to 1} \frac{N!_q}{(1-q)^N} = N! \,. \tag{3.2.9}$$

As an aside, we note that the q-factorial is ubiquitous in the generalisation of functions defined through (e.g. hypergeometric) series. These q-series are obtained from the 'classical' series by replacing any ordinary factorials with corresponding q-factorials. For example, the q-exponential is the generalisation

$$\exp_q(x) = \sum_{n=0}^{\infty} \frac{x^n}{n!_q} \tag{3.2.10}$$

which reverts to the ordinary exponential in the limit

$$\lim_{q\to 1} \exp_q([1-q]x) = \lim_{q\to 1} \sum_{n=0}^{\infty} \frac{(1-q)^n}{n!_q} x^n = \sum_{n=0}^{\infty} \frac{x^n}{n!} = \exp(x)\ . \tag{3.2.11}$$

There is a wealth of known results concerning such q-series, many of which can be found in [56, 58].

Sorting process as a noninteracting ballistic particle system

To conclude this section, I make more precise the mapping of a sorting problem to a ballistic particle system in the absence of reactions. For each element K_i in the unsorted set one ascribes an initial position x_i which has $x_i < x_i + 1, i < N$ and a velocity c_i which has $c_i < c_j, K_i < K_j$. After following their straight-line trajectories for a sufficiently large time the particles will have reordered themselves correctly—see figure 3.7 for the particle-trajectory formulation of the reordering process illustrated in figures 3.5 and 3.6.

Note that in figure 3.7 there are a number of vertices, each one corresponding to a nearest-neighbour particle exchange that occurs in the reordering process, and that this number is unaltered by the numerical values chosen for x_i, c_i (as long as they fulfil the inequalities given above). This corresponds to the fact that there may be more than one way to reorder a particular initial sequence of objects using the same number of swaps. Each vertex has also been labelled with a q, the idea being that the diagram contributes a term equal to the product

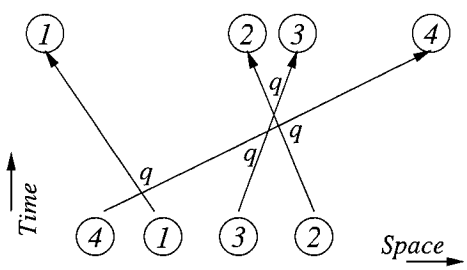

Figure 3.7: The reordering process of figures 3.5 and 3.6 reformulated as particle trajectories in a space-time diagram. Each vertex indicates a nearest-neighbour particle exchange; this particular diagram contributes a term q^4 to the generating function (3.2.1) which can be thought of as the product of the labels associated with each vertex.

of these vertex weights to the generating function (3.2.1). The generating function is then the sum over the $N!$ diagrams that correspond to each initial sequence of objects. It is this vertex counting that will later be used to track the development of a reacting particle system.

3.2.2 Sorting a set of indistinct objects

In a particle reaction system, one expects more than one particle to have the same initial velocity, as in figure 3.8. Here it is clear that the noninteracting particle system corresponds to a sorting process in which some of the N objects may be treated as being equivalent. Consequently any two sequences which differ only by reordering of objects of the same type are also equivalent. In this section I describe a formalism for treating such systems which involves the reordering of operator products which obey q-commutation relations.

Generating function and q-binomial

Consider now a system in which there are only two different particle species and let us pose the following question. Assuming that there are k objects of type 1 (and hence $N-k$ of type 2), how many of the $\binom{N}{k}$ initial sequences require exactly n nearest-neighbour swaps to be brought into the ordered form where all objects of type 1 appear first?

In the previous section, a similar question in which there were $N!$ initial sequences was addressed and it was found that the generating function of nearest-neighbour exchanges was given by the q-deformation of this factorial. We introduce for the present case a similar generating function

$$S_{N,k}(q) = \sum_n c_n \, q^n \tag{3.2.12}$$

in which the coefficient c_n is the number of the $\binom{N}{k}$ initial sequences of objects requiring exactly n nearest-neighbour swaps to become reordered. Based on the results of the previous section,

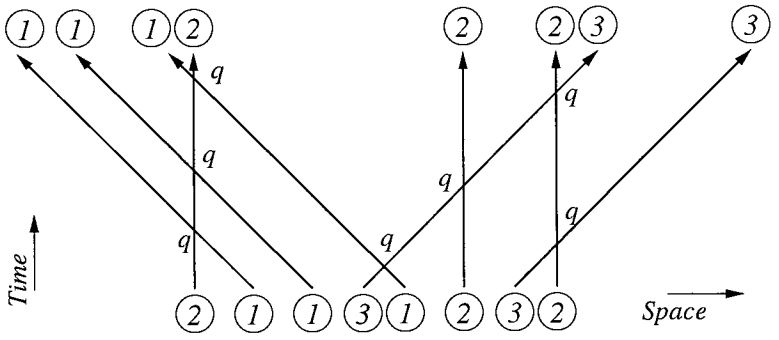

Figure 3.8: Eight particles with three different velocities. Note that because particles with the same velocity never meet, they are not reordered amongst themselves.

one might guess that $S_{N,k}(q)$ is a q-deformation of the binomial coefficient (see e.g. [56, app. I]), that is

$$S_{N,k}(q) = \begin{bmatrix} N \\ k \end{bmatrix}_q = \frac{N!_q}{k!_q (N-k)!_q} . \tag{3.2.13}$$

I show that this is the case by using an operator method that will be developed to describe particle reaction systems.

Operator approach to the problem

Following [57, ch. 10] one finds a closed form for the generating function $S_{N,k}(q)$ by introducing a pair of operators A_1 and A_2 that represent objects of type 1 and 2 respectively. A sequence of objects, e.g. 2121, is then expressed as an ordered operator product, in this case $A_2A_1A_2A_1$. The correctly ordered sequence of these objects is represented by the product $A_1A_1A_2A_2$ and three nearest neighbour exchanges are required to obtain this form.

If one now specifies the relationship between A_1 and A_2 using the q-commutation relation [57]

$$[A_2, A_1]_q \equiv A_2A_1 - qA_1A_2 = 0 \tag{3.2.14}$$

one finds that $A_2A_1A_2A_1 = q^3 A_1A_1A_2A_2$. The factor q^3 that has appeared is precisely that which should appear in generating function $S_{4,2}(q)$. Recall that this encodes the number of configurations requiring specific numbers of swaps to be correctly ordered.

Of particular importance is the combination of operators $(A_1 + A_2)^N$. When multiplied out, this expression gives rise to a sum over all possible initial sequences of N objects of two distinct types. Then, if one uses (3.2.14) repeatedly on each term in the summation, one obtains

$$(A_1 + A_2)^N = \sum_{k=0}^{N} S_{N,k}(q) A_1^k A_2^{N-k} . \tag{3.2.15}$$

That is, the prefactor of each ordered configuration comprising k objects of type 1 followed by $N-k$ objects of type 2 is precisely the generating function $S_{N,k}(q)$ that is of interest here.

To determine a recursion relation for $S_{N,k}(q)$ one multiplies both sides of the previous equation from the left by $(A_1 + A_2)$ and uses (3.2.14) to manipulate the right-hand side. This leads to the expression

$$(A_1 + A_2)^{N+1} = \sum_{k=0}^{N} \left(S_{N,k}(q) A_1^{k+1} A_2^{N-k} + S_{N,k}(q) q^k A_1^k A_2^{N+1-k} \right) . \tag{3.2.16}$$

By comparing particular powers of A_1 and A_2 in this equation with those of (3.2.15) with $N \to N+1$ one finds

$$S_{N+1,k}(q) = S_{N,k-1}(q) + q^k S_{N,k}(q) . \tag{3.2.17}$$

It is easy to verify that with $S_{N,k}(q)$ equal to the q-binomial $\begin{bmatrix} N \\ k \end{bmatrix}_q$ as defined by equation (3.2.13) the previous recursion relation is satisfied. This completes the derivation of the generating function $S_{N,k}(q)$.

In this derivation, the operators A_1 and A_2 have been used simply as a device to keep track of sequences and we have not been concerned with their existence or representation. One way in which these operators can be represented is through a basis of q-fermionic states (see e.g. [53]) and one can then verify explicitly that the expression

$$(A_1 + A_2)^N = \sum_{k=0}^{N} \begin{bmatrix} N \\ k \end{bmatrix}_q A_1^{N-k} A_2^k \tag{3.2.18}$$

comprises an identity for the pair of operators A_1 and A_2.

Generalisation to more types of object

It is quite easy to generalise the result of the previous section for two types (or species) of object to M species with the initial sequence comprising an arbitrary sequence of k_i objects of type i and with $\sum_i k_i = N$.

Again, reordering of the entire initial sequence can be performed iteratively. First one exchanges the k_1 objects of type 1 with their left neighbours until all type 1 objects are at the start of the sequence; then one sorts the remaining sequence of $N - k_1$ objects in the same way but exchanging the places of type 2 objects. Repeating this process for each type of object, the sequence eventually becomes correctly ordered—see figure 3.9. At each stage i, one is moving only objects of type i amongst those of types $i + 1, \ldots, M$ which are considered equivalent. Hence this process is equivalent to a sequence of separate two-species reorderings with progressively shorter initial sequences.

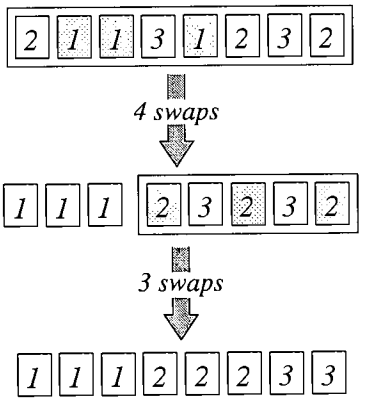

Figure 3.9: Reordering eight objects of three different types by moving those of type 1 first followed by those of type 2. At each step, the unsorted portion of the list is enclosed, and the shaded particles are those that should be exchanged with their left neighbours.

In terms of the generating function, defined as

$$S^{(M)}_{N,k_1,\ldots,k_M}(q) = \sum_n c_n\, q^n \tag{3.2.19}$$

where c_n is the number of sequences for which n nearest-neighbour swaps are required, the decomposition into separate two-species sorting processes implies

$$S^{(M)}_{N,k_1,\ldots,k_M}(q) = S^{(2)}_{N,k_1}(q)\, S^{(M-1)}_{N-k_1,k_2,\ldots,k_M}(q)\ . \tag{3.2.20}$$

If this is iterated, one can express the M-species generating function entirely in terms of the two-species generating function $S^{(2)}_n(k;q)$. Given that this latter function is equal to the q-binomial, it follows that

$$S^{(M)}_{N,k_1,\ldots,k_M}(q) = \left[\begin{array}{c} N \\ k_1\ k_2\ \ldots\ k_{M-1} \end{array}\right]_q \equiv \frac{N!_q}{k_1!_q k_2!_q \ldots k_M!_q}\ . \tag{3.2.21}$$

This is recognisable as a q-deformation of the multinomial, which in turn is the number of initial sequences with k_i objects of type i.

Again, operators could be used to obtain this result in a manner analogous to that described above. In this case one obtains the identity

$$(A_1 + A_2 + \ldots + A_M)^N = \sum_{k_1,k_2,\ldots,k_{M-1}} \left[\begin{array}{c} N \\ k_1\ k_2\ \ldots\ k_{M-1} \end{array}\right]_q A_1^{k_1} A_2^{k_2} \cdots A_M^{k_M} \tag{3.2.22}$$

with $k_M = 1 - \sum_{i=1}^{M-1} k_i$ for the operator algebra defined by the set of q-commutation relations

$$[A_i, A_j]_q = 0 \quad \text{for} \quad i > j\ . \tag{3.2.23}$$

Note these commutators implicitly imply that $[A_i, A_j]_{1/q} = 0$ for $i < j$ and express the fact that each time an object of type j is moved to the left of an object of type i (with $i > j$), a factor of q appears in the generating function. Again $A_1, \ldots, A_M$ can be represented using a basis of q-deformed fermionic states [53].

3.3 Combinatorics of ballistic reaction systems

In the previous section I detailed the connection between trajectories of noninteracting particles undergoing ballistic motion and the sorting of sequences and showed that q-deformations of mathematical objects (such as the factorial and binomial coefficient) arise naturally in the study of sorting problems. Furthermore I illustrated the utility of employing operator products to keep track of sequences of particles. I now extend this approach to a model of reacting particle systems and use it solve exactly a model of stochastic ballistic annihilation and coalescence.

3.3.1 Definition of the stochastic ballistic annihilation and coalescence (SBAC) model

The initial configuration of the stochastic ballistic annihilation and coalescence (SBAC) model comprises N particles each of which is independently assigned a velocity $+\frac{1}{2}$ with probability f_R or a velocity $-\frac{1}{2}$ with probability $f_L = 1 - f_R$. For simplicity, these particles will be referred to as 'left-moving' (and denoted L) and 'right-moving' (R) respectively. At this stage, the inter-particle spacings are arbitrary and these will only be specified in the generalisation to infinite systems in section 3.5.

The particles begin moving at time $t = 0$ with constant velocity until two meet. Then one of three reactions may occur, or the particles may elastically scatter: the probability with which each reaction occurs is given in the table below. In the reaction probabilities three parameters have been introduced: $p = 1 - q$ is the probability that two particles react upon contact and η_L, η_R are proportional to the probability that two particles coalesce into a single left-moving or right-moving particle respectively. Note one must have $\eta_L + \eta_R \leq 1$ and when the equality is satisfied there can be no annihilation reactions.

Reaction		**Probability**
Elastic scattering	$RL \to LR$	q
Rightward coalescence	$RL \to R$	$p\eta_R$
Leftward coalescence	$RL \to L$	$p\eta_L$
Annihilation	$RL \to \emptyset$	$p(1 - \eta_L - \eta_R)$

A realisation of the initial condition and subsequent reaction dynamics is shown in figure 3.10. Note that each vertex corresponds to a reaction or scattering event, and these have been labelled according the corresponding probability.

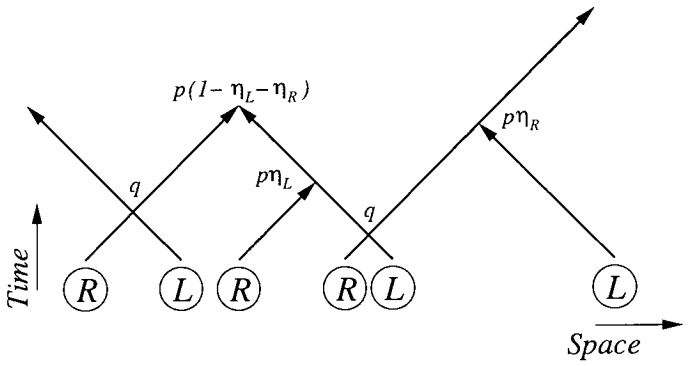

Figure 3.10: A realisation of the dynamics in the SBAC model. Note that the elastic scattering of two approaching particles can be equivalently thought of as two particles passing through each other. The vertex labels indicate the probability a particular type of reaction occurs.

3.3.2 Operator approach to reaction dynamics

As in section 3.2.2 a sequence of particles in the SBAC model is represented using a product of two operators: L for left-moving particles and R for right-moving particles. The initial distribution of sequences of N particles is then written as

$$(f_L L + f_R R)^N \tag{3.3.1}$$

so that each possible initial sequence is weighted according to the probability that it is constructed.

The distribution of final sequences (i.e. when no more reactions are possible) comprises a string of left-moving particles followed by a string of right-moving particles. In terms of operator products, the final state will take the form

$$\sum_{k_L=0}^{N} \sum_{k_R=0}^{N} p_N(k_L, k_R) L^{k_L} R^{k_R} \tag{3.3.2}$$

in which $p_N(k_L, k_R)$ should be the probability that a final sequence comprises exactly k_L left-moving particles and k_R right-moving particles given the initial distribution (3.3.1).

The probabilities $p_N(k_L, k_R)$ can be determined by drawing all possible space-time diagrams like figure 3.10 with k_L (respectively k_R) outgoing left-moving (right-moving) particles and summing the products of all the vertex labels on each diagram. Wherever two particles meet, four new diagrams result, each corresponding to the four possible outcomes of a reaction detailed above. In terms of operator products, this implies the introduction of new terms in a sum over sequences by making the replacement

$$RL \to qLR + p(\eta_L L + \eta_R R + [1 - \eta_L - \eta_R]\mathbb{I}) \,. \tag{3.3.3}$$

Here $\mathbb{I}$ is an identity operator and is used when both particles are removed from the system.

Note that this procedure ensures that each new term is correctly weighted according to the probability that a particular sequence of particles is generated. Note that if the operators R and L to satisfy the q-commutation relation

$$[RL, LR]_q \equiv RL - qLR = p(\eta_L L + \eta_R R + [1 - \eta_L - \eta_R]\mathbb{I}) \,. \tag{3.3.4}$$

that the operator products representing the initial (3.3.1) and final (3.3.2) distribution of particles are mathematically equal. That is

$$(f_L L + f_R R)^N = \sum_{k_L=0}^{N} \sum_{k_R=0}^{N} p_N(k_L, k_R) L^{k_L} R^{k_R} \,, \tag{3.3.5}$$

It is by using these two last relationships that I solve the SBAC model defined above. Note that the operator reordering procedure just described is essentially a convenient mechanism for enumerating all possible paths from an initial to a final sequence of particles, i.e. the method for determining probability distributions described in section 2.1.2.

3.3.3 Generalisation to models with more particle species

Before proceeding with the solution of the SBAC model, I remark on the generalisation to models with more particle species. For the operator method described above to be successful, one must ensure that the operators are associative under multiplication, i.e. $(AB)C = A(BC)$. Then there is only one way to reorder a particular product, and the identity (3.3.5) holds. It is easily verified from (3.3.4) that the pair of operators R and L are associative under multiplication.

If, however, one were to introduce stationary particles that interact nontrivially with both right- and left-moving particles, the above operator reordering method no longer works. To understand why this is so, consider the special case of pure stochastic annihilation (i.e. $\eta_L = \eta_R = 0$ in the SBAC model) in the presence of stationary particles S which interact with the left- and right-moving particles under the reactions $R+S \xrightarrow{p} S$ and $S+L \xrightarrow{p} S$ where $p = 1-q$. Then, in an operator formulation one has

$$RS = qSR + pS \quad \text{and} \quad SL = qLS + pS \,. \tag{3.3.6}$$

Using these relations and (3.3.4) with $\eta_L = \eta_R = 0$ one finds

$$(RS)L - R(SL) = p^2 q(LS - SR) \,. \tag{3.3.7}$$

If the triad of operators L, S and R are to be associative under multiplication, one must have $LS = SR$. However, one cannot in general simultaneously satisfy this relationship and (3.3.6).

The nonassociativity of the operators has its physical origin in the dependence of the set of final-state probabilities on the initial inter-particle separations—see figure 3.11. For the two-velocity system, defined above, the final-state probabilities are invariant under inter-particle spacings and so the operator algebra is associative.

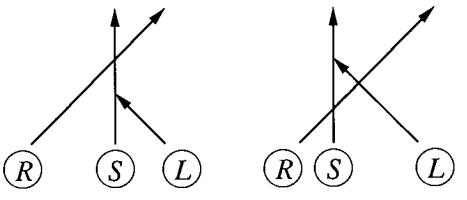

Figure 3.11: A particle reaction system that cannot be solved using the operator methods of this chapter. In the left-hand sequence of reactions, the left-moving particle always meets the stationary particle whereas in the right-hand sequence, they meet only with probability q^2. Hence the distribution of final configurations arising from the combination RSL varies according to the inter-particle spacing.

3.4 Survival probabilities in the SBAC model

I now use the operator formalism developed in the previous section to calculate particle survival probabilities in the SBAC model.

3.4.1 Operator form of survival probabilities

Consider again the N-particle SBAC model, figure 3.10, and add to it two further *test* particles. One of these test particles is a right-moving particle initially positioned to the left of the system and the other a left-moving particle positioned to the right of the system. The resulting system is shown in figure 3.12.

Let us define $P_{LR}(N)$ as the probability that the two test particles meet. In order for this to happen, neither test particle may be removed by any particles emerging from the original system of N particles (i.e. through the dotted lines in figure 3.12).

Given that the right-moving test particle survives after meeting each approaching left-moving particle with probability $q + p\eta_R$ and the left-moving test particle survives each approaching right-moving particle with probability $q + p\eta_L$, the probability that the two meet is given by

$$P_{LR}(N) = \sum_{k_L,k_R} p(k_L,k_R)(q+p\eta_R)^{k_L}(q+p\eta_L)^{k_R} = F_N(q+p\eta_R, q+p\eta_L) \tag{3.4.1}$$

in which the generating function

$$F_N(x,y) = \sum_{k_L,k_R} p(k_L,k_R)x^{k_L}y^{k_R} \tag{3.4.2}$$

has been introduced.

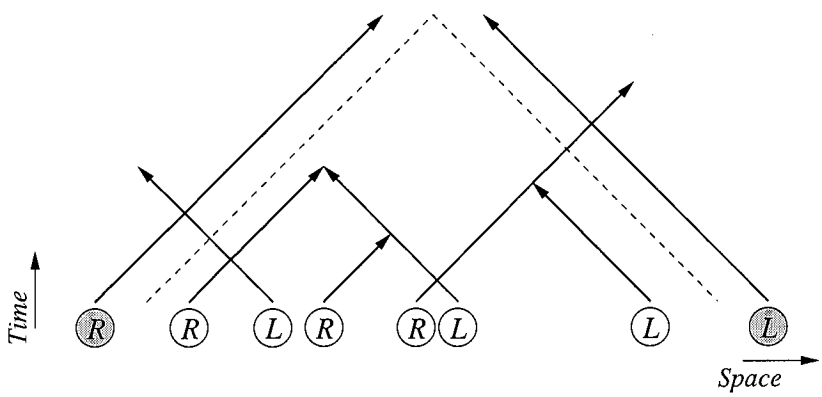

Figure 3.12: The system of figure 3.10 but with the addition of two test particles shown shaded.

Note that the form of the generating function (3.4.2) is very similar to (3.3.2). Indeed, one can obtain (3.4.2) from the operator product (3.3.2) by employing two vectors $|V\rangle$ and $\langle W|$ which obey

$$R|V\rangle = y|V\rangle \tag{3.4.3}$$

$$\langle W|L = \langle W|x \tag{3.4.4}$$

and have scalar product $\langle W|V\rangle = \mathcal{N}$. Then

$$F_N(x,y) = \frac{1}{\mathcal{N}} \sum_{k_L,k_R} p(k_L,k_R) \langle W|L^{k_L} R^{k_R}|V\rangle = \frac{1}{\mathcal{N}} \langle W|(f_L L + f_R R)^N|V\rangle \tag{3.4.5}$$

where the second equality follows from (3.3.5).

If one is interested only in the right-moving test particle, a similar logic can be applied and one finds that its survival probability $P_R(N)$ (regardless of what happens to the other test particle) is $P_R(N) = F_N(q + p\eta_R, 1)$. Similarly, the survival probability of the left-moving test particle is $P_L(N) = F_N(1, q + p\eta_L)$. Hence evaluation of the operator product on the right-hand side of (3.4.5) is of primary importance in the solution of the SBAC model.

Note that the survival probability of the left-moving test particle could also be obtained from that of the right-moving test particle under the exchange $f_L \leftrightarrow f_R$ and $\eta_L \leftrightarrow \eta_R$. This particular property of the model will be referred to as left-right symmetry.

3.4.2 Evaluation of survival probabilities

Having established that the generating function

$$F_N(x,y) = \frac{1}{\mathcal{N}} \langle W|(f_L L + f_R R)^N|V\rangle \tag{3.4.6}$$

is closely related to survival probabilities in the SBAC model, I now explain how it is evaluated. This is achieved by diagonalising the operator $f_L L + f_R R$, i.e. finding the complete set of vectors $\{|\theta\rangle\}$ for which

$$(f_L L + f_R R)|\theta\rangle = \lambda(\theta)|\theta\rangle \; . \tag{3.4.7}$$

Then, one can construct a representation of the identity operator $\mathbb{I}$ from the set $\{|\theta\rangle\}$ which takes the form

$$\mathbb{I} = \int \mathrm{d}\theta \, \nu(\theta) |\theta\rangle\langle\theta| \tag{3.4.8}$$

and where $\nu(\theta)$ is a weight function. Upon insertion of this identity operator into (3.4.6) one obtains the integral expression

$$F_N(x,y) = \frac{1}{\mathcal{N}} \int \mathrm{d}\theta \, \nu(\theta) W(\theta) [\lambda(\theta)]^N V(\theta) \tag{3.4.9}$$

where $W(\theta)$ and $V(\theta)$ arise from the decomposition of the vectors $\langle W|$, $|V\rangle$ onto the basis $\{|\theta\rangle\}$. This integral can then be obtained in the asymptotic (large N) limit using standard techniques, namely the saddle-point method and analytic continuation.

We seek therefore the scalar product $\mathcal{N} = \langle W|V\rangle$ and the four functions $\nu(\theta)$, $W(\theta)$, $V(\theta)$ and $\lambda(\theta)$. These are found by first expressing the R, L operators representing right- and left-moving particles in a sequence of reactants in terms of new pair of operators which satisfy a known operator algebra: that of the q-harmonic oscillator. This is a continuous deformation of the quantum mechanical harmonic oscillator algebra in much the same way that the q-factorial (3.2.8) and q-binomial (3.2.13) are deformations of the ordinary factorial and binomial respectively. The generating function (3.4.6) is then evaluated by using known properties of the q-harmonic oscillator algebra that I review below.

Relationship to the q-harmonic oscillator algebra

Two new operators, a and $a^\dagger$, are initially introduced through the relationships

$$R = \frac{\sqrt{f_L^* f_R^*}}{f_R} a + \eta_L \quad \text{and} \quad L = \frac{\sqrt{f_L^* f_R^*}}{f_L} a^\dagger + \eta_R \,. \tag{3.4.10}$$

in which the two parameters f_L^* and f_R^* that appear are given by

$$f_L^* = f_L(1-\eta_R) \quad \text{and} \quad f_R^* = f_R(1-\eta_L) \,. \tag{3.4.11}$$

These parameters play an important role in the SBAC model as will be seen below.

In terms of a and $a^\dagger$, the generating function (3.4.6) becomes

$$F_N(x,y) = \frac{1}{\mathcal{N}} \langle W| \left[\mathcal{K}(a+a^\dagger) + \mathcal{L}\right]^N |V\rangle \tag{3.4.12}$$

in which, to lighten the notation, the shorthand

$$\mathcal{K} = \sqrt{f_R^* f_L^*} \quad \text{and} \quad \mathcal{L} = 1 - f_R^* - f_L^* \tag{3.4.13}$$

has been introduced. This reveals that the operator to diagonalise is a linear combination of the operator $a + a^\dagger$ and the identity $\mathbb{I}$. Thus the eigenvalue $\lambda(\theta)$ defined in (3.4.7) can be written as

$$\lambda(\theta) = \mathcal{K}\,\tilde{\lambda}(\theta) + \mathcal{L} \tag{3.4.14}$$

in which $\tilde{\lambda}(\theta)$ obeys

$$(a + a^\dagger)|\theta\rangle = \tilde{\lambda}(\theta)|\theta\rangle \,. \tag{3.4.15}$$

The relationship between the operators $a, a^\dagger$ and the vectors $|V\rangle, \langle W|$ can be found by using equations (3.4.3), (3.4.4) and (3.4.10). We find that $|V\rangle$ and $\langle W|$ are eigenvectors of $a^\dagger$ and a respectively:

$$a|V\rangle = \sqrt{\frac{f_R^*}{f_L^*}} \frac{y-\eta_L}{1-\eta_L} |V\rangle \quad \text{and} \quad \langle W|a^\dagger = \langle W| \sqrt{\frac{f_L^*}{f_R^*}} \frac{x-\eta_R}{1-\eta_R} \,. \tag{3.4.16}$$

Additionally, substitution of (3.4.10) into (3.3.4) gives a reordering relation which takes the form of the q-commutator

$$[a, a^\dagger]_q \equiv aa^\dagger - qa^\dagger a = 1 - q \tag{3.4.17}$$

which is now adopted as the definition of $a, a^\dagger$.

This last relationship is reminiscent of the commutation relation for the boson creation $b^\dagger$ and annihilation b operators

$$[b, b^\dagger] \equiv bb^\dagger - b^\dagger b = 1 \tag{3.4.18}$$

and one recovers these from $a, a^\dagger$ in the limit $q \to 1$ if one writes

$$b = \lim_{q\to 1} \frac{a}{\sqrt{1-q}} \quad \text{and} \quad b^\dagger = \lim_{q\to 1} \frac{a^\dagger}{\sqrt{1-q}} \,. \tag{3.4.19}$$

As a result of this, $a, a^\dagger$ are q-boson ladder operators and the algebra (3.4.17) defines the q-harmonic oscillator algebra introduced in [59]. We recall that the quantum mechanical harmonic oscillator has a countably infinite set of boson-number eigenstates. These states have a representation in the continuous coordinate basis (the elements of which are eigenstates of the coordinate operator $b + b^\dagger$) known as Hermite polynomials. In the present work, both the q-boson-number eigenstates and their q-coordinate representation (q-Hermite polynomials) must be used so that the generating function (3.4.6) can be analysed.

The boson-number representation

The representation of the operators a and $a^\dagger$ in the boson-number basis $\{|n\rangle\}, n = 0, 1, 2, \ldots$ is well-known and I review this here. First I assume that the basis vectors $|n\rangle$ and $\langle m|$ satisfy

$$\langle m|n\rangle = \delta_{n,m} \quad \leftrightarrow \quad \mathbb{I} = \sum_{n=0}^{\infty} |n\rangle\langle n| \tag{3.4.20}$$

where $\mathbb{I}$ is the identity operator.

Then, using this basis set, it is easily verified that the definitions

$$a^\dagger|n\rangle = (1 - q^{n+1})^{\frac{1}{2}}|n+1\rangle \tag{3.4.21}$$

$$a|n\rangle = (1 - q^n)^{\frac{1}{2}}|n-1\rangle \tag{3.4.22}$$

$$a|0\rangle = 0 \tag{3.4.23}$$

are consistent with the q-commutator (3.4.17). These definitions allow one to determine the boson-number representation of the vectors $\langle W|$ and $|V\rangle$ as

$$\langle W|n\rangle = \frac{w^n}{\sqrt{(q;q)_n}} \quad \text{and} \quad \langle n|V\rangle = \frac{v^n}{\sqrt{(q;q)_n}} \tag{3.4.24}$$

in which

$$w = \sqrt{\frac{f_L^*}{f_R^*}}\,\frac{x-\eta_R}{1-\eta_R} \quad \text{and} \quad v = \sqrt{\frac{f_R^*}{f_L^*}}\,\frac{y-\eta_L}{1-\eta_L} \tag{3.4.25}$$

and $(q;q)_n$ is the notation for the q-factorial that was introduced through equation (3.2.8).

At this point, one should note an important consequence of the use of the boson-number representation. The vector elements given by (3.4.24) are unbounded as one takes $n \to \infty$ if $v > 1$ or $w > 1$. This means that not all products of the form $\langle W|X|V\rangle$ converge in this representation: for example, the scalar product $\mathcal{N} = \langle W|V\rangle$ has the series expression

$$\mathcal{N} = \sum_{n=0}^{\infty} \frac{(wv)^n}{(q;q)_n} \tag{3.4.26}$$

which is recognisable as the q-exponential (3.2.10). Unlike the classical exponential, this series diverges for $|wv| > 1$. Within its circle of convergence, however, the sum can be written as [56, §1.3]

$$\sum_{n=0}^{\infty} \frac{x^n}{(q;q)_n} = \frac{1}{(x;q)_\infty} \quad |x| < 1\,. \tag{3.4.27}$$

in which a the q-factorial notation has been extended to include the infinite product

$$(x;q)_\infty = \lim_{n\to\infty} \prod_{k=0}^{n} (1 - xq^k)\,. \tag{3.4.28}$$

Hence, when $|vw| < 1$ the factor $\mathcal{N}$ can also be written as

$$\mathcal{N} = \frac{1}{(vw;q)_\infty}\,. \tag{3.4.29}$$

In the analysis below, it will be necessary (at least initially) for *both* $|v| < 1$ and $|w| < 1$ to be satisfied, and so the condition $|vw| < 1$ will automatically hold true.

The coordinate representation

Recall that to obtain the integral representation of the generating function (3.4.9) one must solve the eigenvalue equation (3.4.7) or equivalently (3.4.15) which read

$$(a + a^\dagger)|\theta\rangle = \tilde{\lambda}(\theta)|\theta\rangle\,. \tag{3.4.30}$$

The operator $a + a^\dagger$ is analogous to the coordinate operator of the ordinary harmonic oscillator and so this basis $\{|\theta\rangle\}$ is called the coordinate basis with the coordinate parametrised through the continuous variable θ. The properties of this basis are also well known, and I now describe those that are pertinent to the solution of the SBAC model.

A convenient form of the eigenvalue $\tilde{\lambda}(\theta)$ is

$$\tilde{\lambda}(\theta) = 2\cos(\theta) = \mathrm{e}^{i\theta} + \mathrm{e}^{-i\theta} \tag{3.4.31}$$

and hence $\lambda(\theta)$ that appears in (3.4.9) is given by $\lambda(\theta) = 2\mathcal{K}\cos(\theta) + \mathcal{L}$.

Using the basis vectors $\{|\theta\rangle\}$ a representation of the identity can be constructed as

$$\mathbb{I} = \int \mathrm{d}\theta \nu(\theta) |\theta\rangle\langle\theta| \,. \tag{3.4.32}$$

Note that

$$\langle n|\mathbb{I}|m\rangle = \int \mathrm{d}\theta \nu(\theta) \langle n|\theta\rangle\langle\theta|m\rangle = \delta_{n,m} \tag{3.4.33}$$

which implies that $\nu(\theta)$ is a weight function that orthogonalises the functions $h_n(\theta) = \langle n|\theta\rangle = \langle\theta|n\rangle$.

The set of functions $h_n(\theta)$ is found by evaluating the product $\langle n|a + a^\dagger|\theta\rangle$. By using (3.4.30) and (3.4.21) one obtains

$$\langle n|(a + a^\dagger)|\theta\rangle = 2\cos(\theta)h_n(\theta) = \sqrt{1 - q^{n+1}}h_{n+1}(\theta) + \sqrt{1 - q^n}h_{n-1}(\theta) \,. \tag{3.4.34}$$

We additionally set $h_{-1}(\theta) = 0$ and $h_0(\theta) = 1$ so that $h_n(\theta)$ for $n > 0$ is completely specified by the three-term recursion (3.4.34).

By applying Favard's theorem (see e.g. [57, §5.2]), one can determine that a positive weight function $\nu(\theta)$ that orthogonalises the functions $h_n(\theta)$ exists as long as $1 - q^{n+1} > 0$ for $n = 0, 1, 2, \ldots$. In the SBAC model, $q < 1$ for all nontrivial sets of reaction parameters (q is the probability that a reaction takes place, so $q = 1$ corresponds to a trivial system in which the number of particles remains constant) and so this condition is satisfied which implies $\nu(\theta)$ exists.

There appears to be no systematic method for determining $\nu(\theta)$ from a three-term recursion like (3.4.34), although one can easily determine the form of $h_n(\theta)$ from (3.4.34). This we do by introducing the generating function $G(\theta, x)$ of the functions $h_n(\theta)$ defined as

$$G(\theta, x) = \sum_{n=0}^{\infty} \frac{h_n(\theta)}{\sqrt{(q;q)_n}} x^n \,. \tag{3.4.35}$$

Multiplying both sides of (3.4.34) by $x^n/\sqrt{(q;q)_n}$ and summing over n, one obtains (after some rearrangement) the functional relation

$$G(\theta, x) = \frac{G(\theta, qx)}{(1 - x\mathrm{e}^{i\theta})(1 - x\mathrm{e}^{-i\theta})} \tag{3.4.36}$$

in which the choice $\tilde{\lambda}(\theta) = \cos(\theta)$ admitted the symmetric form in θ of the denominator on the right-hand side. More significantly, this equation relates $G(\theta, x)$ to the same function but with $x \to qx$. Repeated application of this equation gives

$$G(\theta, x) = \frac{G(\theta, 0)}{(x\mathrm{e}^{i\theta}, x\mathrm{e}^{-i\theta}; q)_\infty} = \frac{1}{(x\mathrm{e}^{i\theta}, x\mathrm{e}^{-i\theta}; q)_\infty} \tag{3.4.37}$$

in which the fact that $G(\theta,0) = h_0(\theta) = 1$ has been used and a standard extension [56, ch. 1] to the q-factorial notation

$$(a_1, a_2, \ldots, a_m; q)_n = \prod_{k=0}^{n-1}(1 - a_1 q^k)(1 - a_2 q^k)\cdots(1 - a_m q^k) \tag{3.4.38}$$

has been introduced.

To obtain an explicit form of the functions $h_n(\theta)$, one now needs to expand (3.4.37) in powers of x and compare coefficients with (3.4.35). To do this we employ the identity (3.4.27) for the series representation of an infinite product. This then implies that where $G(\theta,x)$ converges $(x < 1)$,

$$h_n(\theta) = \frac{1}{\sqrt{(q;q)_n}} \sum_{k=0}^{n} \begin{bmatrix} n \\ k \end{bmatrix}_q \mathrm{e}^{i(n-2k)} \tag{3.4.39}$$

in which the q-binomial that arose from the sorting process of section 3.2.2 has made a reappearance.

Having found the functions $h_n(\theta)$, we recognise them as q-Hermite polynomials introduced by Rogers in 1894 [60]. Although Rogers did not compute the weight function for these polynomials, subsequent studies, e.g. [61], have revealed that if one has

$$\nu(\theta) = \frac{(q, e^{2i\theta}, e^{-2i\theta}; q)_\infty}{2\pi} \quad \theta \in [0,\pi] \tag{3.4.40}$$

equation (3.4.33) is satisfied. Additionally, the domain of the weight function indicates the proper limits to use for the integration, i.e. θ runs from 0 to π.

As a byproduct of the foregoing analysis we have, in fact, determined the functions $W(\theta) = \langle W|\theta\rangle$ and $V(\theta) = \langle\theta|V\rangle$: they are both equal to the generating function of q-Hermite polynomials. To be explicit

$$W(\theta) = \langle W|\theta\rangle = \sum_{n=0}^{\infty}\langle W|n\rangle\langle n|\theta\rangle \tag{3.4.41}$$

$$= \sum_{n=0}^{\infty} \frac{w^n}{\sqrt{(q;q)_n}} h_n(\theta) = G(\theta, w)\ . \tag{3.4.42}$$

Similarly, $V(\theta) = G(\theta, v)$.

3.4.3 Integral representation and asymptotic forms of the generating function

Using the connection to known properties of the q-deformed harmonic oscillator algebra, we have determined that the generating function $F_N(v,w)$

$$F_N(v,w) = \frac{1}{\mathcal{N}}\langle W|(f_L L + f_R R)^N|V\rangle = \frac{1}{\mathcal{N}}\langle W|[\mathcal{K}(a + a^\dagger) + \mathcal{L}]^N|V\rangle \tag{3.4.43}$$

supplemented by the relations $a|V\rangle = v|V\rangle$, $\langle W|a^\dagger = \langle W|w$ and $aa^\dagger - qa^\dagger a = 1 - q$ can be expressed in the form

$$F_N(v,w) = \frac{1}{\mathcal{N}} \int_0^\pi \mathrm{d}\theta\, \nu(\theta) G(\theta,w)[2\mathcal{K}\cos(\theta) + \mathcal{L}]^N G(\theta,v)\ . \tag{3.4.44}$$

In this equation

$$\nu(\theta) = \frac{(q, e^{2i\theta}, e^{-2i\theta}; q)_\infty}{2\pi} \quad \text{and} \quad G(\theta,x) = \frac{1}{(x\mathrm{e}^{i\theta}, x\mathrm{e}^{-i\theta}; q)_\infty} \tag{3.4.45}$$

and w and v are given by (3.4.25).

Noting that the integrand of (3.4.44) is even in θ, equation (3.4.44) can be recast as the closed contour integral over the complex variable $z = \mathrm{e}^{i\theta}$

$$F_N(v,w) = \frac{(q;q)_\infty}{4\pi i \mathcal{N}} \oint_K \frac{\mathrm{d}z}{z}[\mathcal{K}(z+z^{-1}) + \mathcal{L}]^N \frac{(z^2, z^{-2}; q)_\infty}{(wz, w/z, vz, v/z; q)_\infty} \tag{3.4.46}$$

where K is the unit circle ($|z| = 1$) and is directed anti-clockwise.

As has already been stressed, a direct consequence of the elements of the vectors $\langle W|, |V\rangle$ (3.4.24) remaining finite only for $w < 1$, respectively $v < 1$, is that this integral holds as it stands only for $v < 1, w < 1$. However, the integrand of (3.4.46) defines an analytic structure which allows the asymptotic (large-N) form of $F_N(v,w)$ to be determined not just for $v < 1, w < 1$ but also for general v and w as I now discuss.

First of all, it is convenient to split the integrand of (3.4.46) into two parts: a function

$$\lambda(z) = \mathcal{K}(z + z^{-1}) + \mathcal{L} \tag{3.4.47}$$

which is raised to the power of N and the remaining part

$$g(z) = \frac{1}{4\pi i \mathcal{N} z} \frac{(q, z^2, z^{-2}; q)_\infty}{(vz, v/z, wz, w/z; q)_\infty} \tag{3.4.48}$$

which is independent of N. As N is taken ever larger, it will be the value of $\lambda(z)^N$ evaluated at specific points that constitute the dominant contribution to the integral (3.4.46). One such point is where $\lambda(z)$ has a maximum in its real part: this is at $z = z_0 = 1$ and along a path parallel to the imaginary axis. In the transverse direction, $\Re(\lambda(z))$ has a minimum and so z_0 is in fact a saddle point. A second saddle point exists at $z = -1$; however its contribution to the integral (3.4.46) is found always to diminish in the limit $N \to \infty$ (compared to other terms) and hence we do not consider it any further.

The contour K in (3.4.46) passes through z_0 along the path of steepest descent, and so by expanding both $\lambda(z)^N$ and $g(z)$ around this point one obtains

$$F_N(v,w) = F_N^*(v,w) \sim \lambda(z_0)^N \left(\frac{a_0}{N^{1/2}} + \frac{a_1}{N^{3/2}} + \ldots\right) \tag{3.4.49}$$

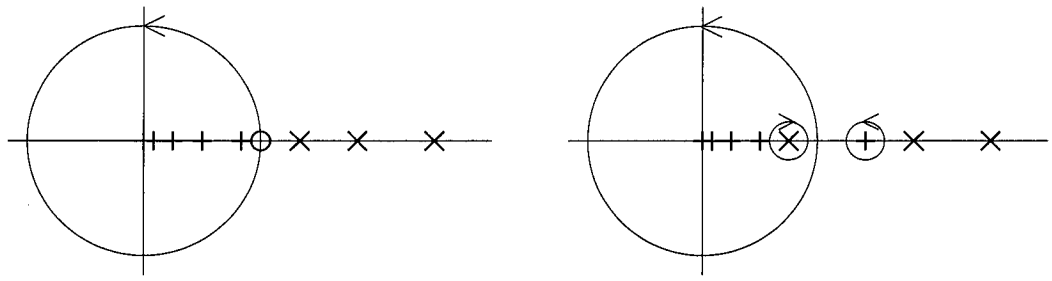

Figure 3.13: The left-hand figure shows the structure of the integrand of equation (3.4.46) when $|v| < 1$, $|w| < 1$. The contour K passes through the saddle point (indicated by the thick open circle), and the poles enclosed by K are schematically indicated by + symbols and those outside it by × symbols. The right-hand figure illustrates the effect of increasing, say, $v > 1$. A pole which was enclosed by K has moved outside the unit circle (and conversely one which was outside is now inside) and so two additional contours are drawn which must be integrated over in order to obtain the correct expression for the integral (3.4.46).

where the coefficients a_n will be quantified in the more detailed analysis below. This resulting asymptotic expansion of $F_N(v,w)$ will hold for the case $|v| < 1, |w| < 1$.

As v and/or w is increased to a value greater than one, some of the poles of $g(z)$ move from inside the unit circle to outside and vice-versa. In order to retain the analytic structure defined by the integral (3.4.46), one must distort the contour so that it still encloses the same set of poles (specifically those at $z = q^k v, q^k w$ with $k = 0, 1, 2, \ldots$) no matter what the values of v and w. This idea (known as analytic continuation) is illustrated in figure 3.13 in which the distorted contour has been split into several distinct closed paths.

To evaluate the integral, we add to the result from the unit circle (3.4.49) the contributions from the poles enclosed by the distorted contour using, e.g. the residue theorem. Noting that $\lambda(1/z) = \lambda(z)$, contours centred around, say, $z = q^k v, 1/q^k v$, will contribute terms proportional to $\lambda(q^k v)^N$. Given that $\lambda(z)$ increases monotonically along the real axis from the saddle point $z = 1$, the pair of poles furthest from the saddle point (in this case those at $z = v, 1/v$) will always dominate over the saddle-point contribution (3.4.49) in the large-N limit.

I now present the results one obtains from this analysis.

Case A $(|v| < 1, |w| < 1)$

Here the saddle-point expression (3.4.49) is valid and the expansion of $\lambda(z)$, $g(z)$ gives

$$F_N(v,w) \sim \sqrt{2\pi i}[\lambda(z_0)]^N \left[\frac{g(z_0)}{|h''(z)|^{\frac{1}{2}}} \frac{1}{N^{\frac{1}{2}}} + \left(\frac{g'(z_0)h'''(z_0)}{2|h''(z_0)|^{\frac{5}{2}}} - \frac{g''(z_0)}{2|h''(z_0)|^{\frac{3}{2}}} \right) \frac{1}{N^{\frac{3}{2}}} + \cdots \right] \tag{3.4.50}$$

in which $h(z) = \ln \lambda(z)$. One can show that $g(z_0) = g'(z_0) = 0$ and so

$$F_N(v,w) \sim \frac{1}{\sqrt{4\pi}\mathcal{N}} \frac{(q;q)_\infty^3}{(v,w;q)_\infty^2} \frac{(2\mathcal{K}+\mathcal{L})^{N+\frac{3}{2}}}{(\mathcal{K}N)^{\frac{3}{2}}} . \tag{3.4.51}$$

Case B1 $(v = 1, |w| < 1)$

Here things are a little subtle as the apparent poles in the function $g(z)$ at $z = 1$ are cancelled by zeros in the numerator. Indeed,

$$\lim_{v\to 1} g(z) = \frac{1}{\pi i \mathcal{N}} \frac{(q;q)_\infty}{(w;q)_\infty^2} \tag{3.4.52}$$

and the saddle-point formula given above implies

$$F_N(1,w) \sim \frac{1}{\sqrt{\pi}\mathcal{N}} \frac{(q;q)_\infty}{(w;q)_\infty^2} \frac{(2\mathcal{K}+\mathcal{L})^{N+\frac{1}{2}}}{(\mathcal{K}N)^{\frac{1}{2}}} . \tag{3.4.53}$$

Case B2 $(v < 1, w = 1)$

As (3.4.46) is symmetric in v and w, one can simply replace w in the previous expression with v to obtain

$$F_N(v,1) \sim \frac{1}{\sqrt{\pi}\mathcal{N}} \frac{(q;q)_\infty}{(v;q)_\infty^2} \frac{(2\mathcal{K}+\mathcal{L})^{N+\frac{1}{2}}}{(\mathcal{K}N)^{\frac{1}{2}}} . \tag{3.4.54}$$

Case C1 $(v > 1, w \leq 1)$

From the above discussion we know that the integrals over contours enclosing the poles at $z = v, 1/v$ will form the dominant contribution to the integral (3.4.46). If it is assumed that there is only a single pole at each of these points, one finds using the residue theorem that

$$F_N(v,w) = F_N^*(v,w) + 2\pi i \left(\operatorname*{Res}_{z=v} [\lambda(z)]^N g(z) - \operatorname*{Res}_{z=1/v} [\lambda(z)]^N g(z) \right) + \cdots . \tag{3.4.55}$$

The residues are easily computed and one finds

$$F_N(v,w) \sim \frac{1}{\mathcal{N}} \frac{(v^{-2};q)_\infty}{(vw, w/v; q)_\infty} \left[\mathcal{K}(v+v^{-1}) + \mathcal{L}\right]^N \tag{3.4.56}$$

in which the saddle-point contribution $F_N^*(v,w)$ has been discarded due to its being sub-dominant in the large-N limit as discussed above.

Case C2 $(v \leq 1, w > 1)$

Again, through the symmetry of (3.4.46) in v and w, we can treat the case $v \leq 1, w > 1$ by simply exchanging v and w in the previous expression:

$$F_N(v,w) = \frac{1}{\mathcal{N}} \frac{(w^{-2};q)_\infty}{(vw, v/w; q)_\infty} \left[\mathcal{K}(w+w^{-1}) + \mathcal{L}\right]^N . \tag{3.4.57}$$

Other cases The combinations of v and w not included in the above cases do not arise in the analysis of the SBAC model because they correspond to unphysical scenarios such as reaction probabilities larger than one.

3.4.4 Forms of survival probabilities in the finite system

Using the above results, we can now (finally) write down asymptotic forms of the survival probabilities introduced in section 3.4.1. These then constitute new results for the SBAC model introduced in section 3.3.1.

Meeting probability of two approaching test particles

Recall that the probability that two initially oppositely directed test particles separated by N other particles meet with probability $P_{LR}(N) = F_N(q + p\eta_R, q + p\eta_L)$. From (3.4.25) we find

$$w = q\sqrt{\frac{f_L^*}{f_R^*}} \quad \text{and} \quad v = q\sqrt{\frac{f_R^*}{f_L^*}} \,. \tag{3.4.58}$$

It is convenient to introduce the ratio

$$\chi = \frac{f_R^*}{f_L^*} \tag{3.4.59}$$

so that $w = q/\sqrt{\chi}$ and $v = q\sqrt{\chi}$. This makes it explicit that $vw = q^2 < 1$ for all combinations of the SBAC model parameters and that the cases listed above are sufficient to construct all the physically accessible forms of the meeting probability.

For each of the cases that arise, the two-particle meeting probability decays to zero with N as

$$P_{LR}(N) \sim A\frac{(1-B)^{N+\gamma}}{N^\gamma} \,. \tag{3.4.60}$$

The parameters A, B and γ and the values of χ for which they apply in are given in table 3.1 in which the constant K is defined as

$$K = \frac{(q;q)_\infty^4}{(q\chi, q/\chi; q)_\infty^2} \,. \tag{3.4.61}$$

	A	B	γ
$\chi > 1/q^2$	$(1-q\chi)(1-q^2\chi)$	$(1-q)f_R^* - (q^{-1}-1)f_L^*$	0
$\chi = 1/q^2$	$\frac{1-q}{\sqrt{\pi}(f_R^* f_L^*)^{\frac{1}{4}}}$	$(f_R^* - f_L^*)^2$	$\frac{1}{2}$
$q^2 < \chi < 1/q^2$	$\frac{1}{\sqrt{4\pi}(f_R^* f_L^*)^{3/4}} \frac{K}{1-q}$	$(f_R^* - f_L^*)^2$	$\frac{3}{2}$
$\chi = q^2$	$\frac{1-q}{\sqrt{\pi}(f_R^* f_L^*)^{\frac{1}{4}}}$	$(f_R^* - f_L^*)^2$	$\frac{1}{2}$
$\chi < q^2$	$(1-q/\chi)(1-q^2/\chi)$	$(1-q)f_L^* - (q^{-1}-1)f_R^*$	0

Table 3.1: Parameters that arise for different values of the ratio χ in the decay of the two-particle meeting probability $P_{LR}(N) \sim A(1-B)^{N+\gamma}N^{-\gamma}$.

	F_∞	A	B	γ
$\chi > 1$	$1 - f_L^*/f_R^*$	$\frac{1}{\sqrt{4\pi}(f_R^* f_L^*)^{3/4}} \frac{K}{(1-\sqrt{\chi})^2}$	$(f_R^* - f_L^*)^2$	$\frac{3}{2}$
$\chi = 1$	0	$\frac{1}{\sqrt{\pi}(f_R^* f_L^*)^{\frac{1}{4}}}$	0	$\frac{1}{2}$
$q^2 < \chi < 1$	0	$\frac{1}{\sqrt{4\pi}(f_R^* f_L^*)^{3/4}} \frac{K}{(1-\sqrt{\chi})^2}$	$(f_R^* - f_L^*)^2$	$\frac{3}{2}$
$\chi = q^2$	0	$\frac{1}{\sqrt{\pi}(f_R^* f_L^*)^{\frac{1}{4}}}$	$(f_R^* - f_L^*)^2$	$\frac{1}{2}$
$\chi < q^2$	0	$(1 - q^2/\chi)$	$(1-q)(f_L^* - f_R^*/q)$	0

Table 3.2: Parameters that arise for different values of the ratio χ in the decay of the right-moving particle survival probability $P_R(N) \sim F_\infty + A(1-B)^{N+\gamma} N^{-\gamma}$.

Note that the special case $\chi = 1$ implies $f_L^* = f_R^*$ and the decay of the meeting probability is pure power law:

$$P_{LR}(N) \sim \frac{1}{\sqrt{4\pi}(f_R^* f_L^*)^{3/4}(1-q)N^{\frac{3}{2}}} . \tag{3.4.62}$$

Furthermore, for sufficiently small χ ($\chi < q^2$) the decay is purely exponential ($\gamma = 0$).

Survival probability of a single particle

The survival probability of a single particle is obtained in manner similar to that just described. Considering first the right-moving test-particle at the left of the system, we recall that its survival probability is given by $P_R(N) = F_N(p+q\eta_R, 1)$ which translates to $w = q/\sqrt{\chi}, v = \sqrt{\chi}$ in (3.4.46). In the following only the right-moving test particle is considered because the left-moving test particle's survival probability $P_L(N)$ can be obtained by exchanging $L \leftrightarrow R$ and hence $\chi \to 1/\chi$.

The asymptotic form of the survival probability takes the generic form

$$P_R(N) \sim F_\infty + A\frac{(1-B)^{N+\gamma}}{N^\gamma} . \tag{3.4.63}$$

The values of the parameters that apply for different values of χ appear in table 3.2 for all $\chi > 0$. Note in this expression we have for $\chi > 1$ a residual constant term F_∞: i.e. there is some probability that a right-moving test particle manages to survive extinction even in a system containing an infinite number of other particles. Note also that the condition for this state of affairs, $\chi > 1$ can be written $f_R^* > f_L^*$: that is the 'dominant' particle species is that with the larger initial effective density, as the quantities f_R^* and f_L^* will now be called.

The pure power-law decay of the survival probability

$$P_R(N) \sim \frac{1}{\sqrt{\pi}(f_L^* f_R^*)^{3/4}} \frac{1}{N^{1/2}} \tag{3.4.64}$$

arises when the effective densities are exactly equal. Furthermore, this is the only situation in which both the left- and right-moving test-particles are guaranteed to become extinct if one takes the number of other particles in the system to infinity.

3.5 Properties of the infinite system

Until now only the final ordered state of systems comprising a finite number N of particles has been considered. I turn now to the study of the time evolution of physical quantities in an infinite system. The transition from one to the other is easily made because if one follows a chosen particle in an infinite system until a time t, it can only be influenced by those other particles that were positioned within a distance ct of it at time $t = 0$. Here, c is the difference in velocity between a right- and left-moving particle which was earlier set to unity.

To be more specific, I now define the way in which the infinite system is constructed. Firstly, particle positions are chosen such that the nearest-neighbour particle distances are independent random variables drawn from the distribution $g_1(x)$ which has mean $\int_0^\infty \mathrm{d}x\, x g_1(x) = 1$. From this distribution one can calculate $g_n(x)$, the distribution of n^{th}-nearest neighbour particle separations, through the convolution

$$g_{n+1}(x) = \int_0^x \mathrm{d}x'\, g_n(x') g_1(x - x') \,. \tag{3.5.1}$$

For simplicity I consider only a Poisson initial condition, i.e. $g_1(x) = \exp(-x)$.

As in the finite system, each particle is independently assigned a velocity $+\frac{1}{2}$ with probability f_R and $-\frac{1}{2}$ with probability f_L. The set of allowed reactions and their probabilities is also as given in section 3.3.1.

The results of the previous section can new be used to calculate the density of reactions in the (x, t) plane and the overall particle density as a function of time.

3.5.1 Density of reaction events

We define $\rho_R(x,t)\mathrm{d}x\mathrm{d}t$ as the probability that a reaction occurs between position x and $x + \mathrm{d}x$ and between time t and $t + \mathrm{d}t$, averaged over all possible initial particle orderings and spacings as well as reaction sequences. This quantity takes the form

$$\rho_R(x,t) = [1 - q] \sum_{N=0}^{\infty} [f_R f_L g_{N+1}(t)][P_{LR}(N)] \tag{3.5.2}$$

which arises as follows. The factor $(1-q)$ is simply the probability that, given two particles meet, they react. The sum over N reflects the fact that, after a time t, two approaching particles could originally have been—in principle—separated by any number N of other particles. This sum is weighted by $f_R f_L g_{N+1}(t)$ which is the probability that two particles were initially separated by a distance t, are approaching and had N particles in between at time $t=0$. The final factor that appears is the probability that these two approaching particles actually meet. Note that this expression contains no dependence on x because the initial distribution of particles is homogeneous (and so I drop the argument x of ρ_R).

For the exponential nearest-neighbour distribution $g_1(x) = \exp(-x)$ the n^{th}-nearest-neighbour distribution $g_n(x)$ is found from (3.5.1) to be

$$g_n(x) = \frac{x^{n-1}}{(n-1)!}\exp(-x)\ . \tag{3.5.3}$$

At large fixed x, this function is peaked around values of $n \approx x$. This means that for sufficiently large t, one can use the asymptotic forms of $P_{LR}(N)$ in (3.5.2) to find

$$\rho_R(t) \sim (1-q) f_R f_L A (1-B)^{\gamma} \mathrm{e}^{-t} \sum_{N=0}^{\infty} \frac{[t(1-B)]^N}{N!}\frac{1}{N^{\gamma}}\ . \tag{3.5.4}$$

This sum is dominated by the terms around the maximum of $[t(1-B)]^N/N!$ which occurs at $N_0 = (1-B)t$ and so we expand the power-law contribution to first order in $N - N_0$. Performing the resulting summation yields

$$\rho_R(t) \sim (1-q) f_R f_L A \frac{\exp(-Bt)}{t^{\gamma}}\ . \tag{3.5.5}$$

It is now a simple matter of substituting in the values of A, B and γ from table 3.1 to obtain the explicit expressions for the reaction density.

3.5.2 Density of particles

I now calculate $\rho(x,t)$, the density of particles remaining at position x after time t. Once again, the homogeneous initial conditions imply that this quantity will not depend on x. Thus one can write

$$\rho(t) = f_L \widetilde{P}_L(t) + f_R \widetilde{P}_R(t) \tag{3.5.6}$$

in which $\widetilde{P}_L(t)$ and $\widetilde{P}_R(t)$ are the probabilities that a left-moving, respectively right-moving, particle survives up to a time t. Again one need consider only the survival probability of the right-moving particle, as the equivalent quantity for the left-moving particle can be obtained from the left-right symmetry of the problem.

The right-moving test particle that is traced to determine $\widetilde{P}_R(t)$ can, after a time t, only have interacted with those particles that were initially positioned within a distance t of the test particle. This implies

$$\widetilde{P}_R(t) = \sum_{N=0}^{\infty} P_R(N) G_N(t) \tag{3.5.7}$$

where $P_R(N) = F_N(q+p\eta_R, 1)$ is the probability that the test particle survives after interacting with N other particles and $G_N(t)$ is the probability that there were initially *exactly* N particles in a region of size t.

This latter quantity can be expressed in terms of the neighbour-distribution functions $g_n(x)$ via

$$G_N(t) = \int_0^t \mathrm{d}x\, g_N(x) \int_{t-x}^{\infty} \mathrm{d}x'\, g_1(x') \,. \tag{3.5.8}$$

That is, the probability that there are exactly N particles in an interval t is the probability that the N^{th} particle is at a position $0 \leq x \leq t$ and the $(N+1)^{\text{th}}$ particle is at a position $x' > t$. For the Poisson initial conditions $g_1(x) = \exp(-x)$ equation (3.5.8) implies

$$G_N(t) = \frac{t^N}{N!} \exp(-t) = g_{N+1}(t) \tag{3.5.9}$$

and the sum (3.5.7) is of the same general form as that for the density of reactions (3.5.4).

Following the same procedure as that in the previous section to find both the left- and right-moving test-particle survival probabilities $\widetilde{P}_L(t)$ and $\widetilde{P}_R(t)$ we find from (3.5.6) that when $\chi \leq 1$ the density decays as

$$\rho(t) \sim f_L(1-\chi) + \Delta(t) \tag{3.5.10}$$

where the approach $\Delta(t)$ to the constant density state (which comprises a set of particles all moving in the same direction) is given in full in table 3.3. Results for the case $\chi > 1$

	$\Delta(t)$
$\chi = 1$	$\dfrac{1}{\sqrt{2\pi}(f_L^* f_R^*)^{3/4}} \dfrac{1}{t^{1/2}}$
$q^2 < \chi < 1$	$\dfrac{K}{4\sqrt{2\pi}(f_L^* f_R^*)^{3/4}} \dfrac{f_L f_R^* + f_R f_L^*}{(\sqrt{f_L^*} - \sqrt{f_R^*})^2} \dfrac{\exp[-(\sqrt{f_L^*} - \sqrt{f_R^*})^2 t]}{t^{3/2}}$
$\chi = q^2$	$\dfrac{f_R}{\sqrt{2\pi}(f_L^* f_R^*)^{3/4}} \dfrac{\exp[-(\sqrt{f_L^*} - \sqrt{f_R^*})^2 t]}{t^{1/2}}$
$\chi < q^2$	$f_R \left(1 - \dfrac{f_R^*}{q^2 f_L^*}\right) \exp[-(1-q)(f_L^* - f_R^*/q)\, t]$

Table 3.3: The approach to the asymptotic constant density state in the SBAC model.

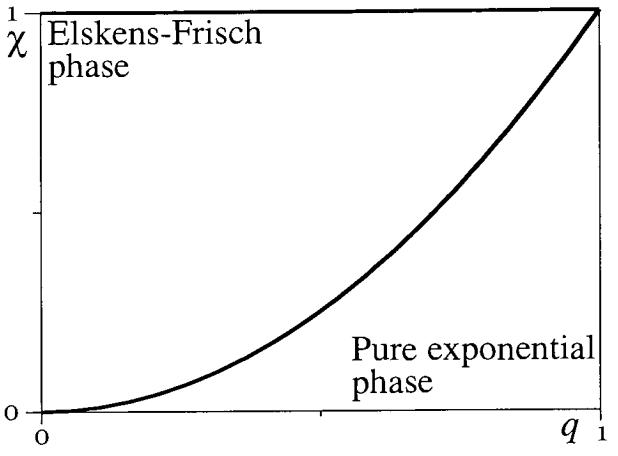

Figure 3.14: Phase diagram for all combinations of stochastic ballistic annihilation and coalescence with a bimodal velocity distribution of particles. The line $\chi = 1$ corresponds to a power-law decay form; the Elskens-Frisch phase corresponds to that found in [49] in which the density decay is exponential times a $t^{-3/2}$ power-law; the phase induced by stochasticity in the reactions (increased q) exhibits a pure exponential decay.

can be obtained by making the left-right exchange as discussed previously. We see then that the particle species with the smaller effective density (f_L^* vs. f_R^*) is always completely wiped out by the other species, except in the special case of equal effective densities in which every right-moving particle is removed by a partnering left-moving particle and vice versa.

These results are conveniently presented in the form of a phase diagram for the SBAC model—see figure 3.14. This emphasises the striking result that stochastic ballistic annihilation and coalescence reaction dynamics are very closely related, at least for the case of a bimodal velocity distribution. Recall that for the diffusive counterparts, both reactions give rise to the same density decay form and for all initial conditions. Here the situation is different, in that all information concerning the reactions of a particular model, along with the initial densities, are encoded into two parameters χ and q which in turn govern the density decay form. That is, for a generic combination of annihilation and coalescence (expressed through the parameters η_L and η_R), the same four decay regimes are found by varying the initial densities or stochasticity parameter q. This behaviour can be likened to a law of corresponding states [5]. One should note, however, that although a particular combination of χ and q gives rise to the same *general* density decay form (e.g. exponential), the values of the decay amplitudes and characteristic timescales may vary if f_R^* and f_L^* are varied with χ held fixed—see table 3.3.

Note that the line $q = 0$ (particles always react) corresponds to the model studied by Elskens and Frisch [49]. The two phases found in that case persist for nonzero q. The pure-power law phase (which is suggestive of the system being invariant under rescaling of space and time) applies where the two effective densities f_L^* and f_R^* are equal, no matter what the value of q. The value of the decay exponent can be understood through the arguments of Elskens and

Frisch [49]. The idea is that as the mean reaction rates of the two species are matched, the density decay is governed by the excess number of particles of one species in a zone of size ℓ. From the central-limit theorem, this excess is $\sim \ell^{1/2}$ and hence after a time t, a number $\sim t^{1/2}$ of particles will have reacted giving rise to the $\rho \sim t^{-1/2}$ decay form. Furthermore, we note that the amplitude of this power-law decay is unaffected by the stochasticity parameter q.

The Elskens-Frisch phase, in which the two particle species share the same decay form (exponential times $t^{-3/2}$) ceases to exist once q is increased to a value $\sqrt{\chi}$. That is, once the probability $1-q$ of a reaction occuring is sufficiently small, new decay forms are realised in which the two species decay at *unequal* rates. An extreme, and easily understood, case of this phenomenon is the case $\eta_R = 0$, $\eta_L = 1$ which implies that left-moving particles do not decay. Rather, they absorb the onslaught of exponentially distributed right-moving particles with probability $1-q$. This implies $\rho(t) = f_R \mathrm{e}^{-2(1-q)f_L t}$, which is indeed the result obtained through the full analysis.

3.6 Summary of results and open questions

In this chapter I have studied ballistic annihilation and coalescence processes in the framework of sorting processes, through which a connection with the commutation properties of operators was established. I then used this relationship to solve the two-velocity model exactly for all possible reaction probabilities and through this I found that stochastic ballistic annihilation and coalescence reactions are closely related. Mathematically, the correspondence of the two processes is expressed in the mapping to the q-deformed harmonic oscillator algebra (3.4.10) and physically we likened it to a law of corresponding states.

It should be emphasised that the correspondence between the two reactions is not the same as that previously established for the case of diffusive reactants. In the latter case, the same density decay form is found for all initial conditions and combinations of reaction probabilities. Here, the decay amplitudes, exponents and timescales all depend on the initial conditions and combination of reaction probabilities. Nevertheless, a single phase diagram can be drawn for all combinations of stochastic ballistic annihilation and coalescence and two particle velocities (figure 3.14).

One particular feature that emerged in the phase diagram was the special line of equal effective particle densities. Along this line, a power-law decay form is exhibited, which in turn is suggestive of a dynamical scaling regime (i.e. invariance of the system under simultaneous rescaling of both space and time). That this arises only for certain combinations of the model parameters is also in contrast to the case of the diffusive reaction system. Secondly, I noted

the phenomenon of the two particle species decaying at unequal rates which comes about when the probability of reaction is sufficiently small and the effective densities are unequal.

The natural question which arises from this work is how much the correspondence of the two reactions is retained in more general ballistic reaction systems. As was noted in section 3.3.3, the operators corresponding to reacting particles with more than two velocities are not associative under multiplication. It appears that none of the remaining known (associative) quadratic algebras can be interpreted as a description of reacting particle dynamics (e.g. one finds that probability is not conserved). Nonetheless, I feel that other discrete velocity distributions could be treated using similar methods if one could identify a suitable invariance property of the system: in the case of the bimodal velocity distribution, it was the invariance of survival probabilities under the change of particle spacings that allowed the system to be treated using operators that directly related to the particle species.

Within the two-species model discussed here, there remain a number of quantities that could, in principle, be calculated using the methods that have described in this chapter. One is the nature of correlation functions in the model; it is possible to define an operator that 'projects out' particle sequences in which two specified co-moving particles survive, although there remains a technical problem in converting this to a *bona fide* survival probability. Additionally, one can consider a disordered system in which each right-moving particle carries its own reaction probability q_i. It is easily shown that the operator algebra for such a system is associative; however, it turns out that one cannot find a pair of vectors that allows the conversion of operator products to survival probabilities. It may, however, be possible to resolve this technical problem by using, e.g., tensor products of operators and vectors.

Finally there remains the question of models in greater than one dimension. In the continuing absence of a generic approach to ballistic reaction systems, exact solutions of such models remain an immense challenge.

Chapter 4

Exclusion processes

Exclusion processes are stochastic microscopic models of driven diffusive systems, i.e. systems that are driven by their environment rather than being at equilibrium with it. In this chapter I concentrate on the steady-state properties of driven diffusive systems, beginning with an overview of the "standard model" of a driven lattice gas. In particular I examine the role that the boundary conditions play in causing a model to exhibit a nonequilibrium steady state. In section 4.2 I define the one-dimensional partially asymmetric exclusion process as a limit of the more general driven lattice gas and in subsequent sections I use a matrix product method to obtain exact new results for the process. Interestingly, the operator algebra of the previous chapter (the q-deformed harmonic oscillator algebra) is central to the solution of the model. Finally, in section 4.8 I review other results from the literature for exclusion processes.

4.1 Review of the driven lattice gas

Nearly twenty years ago, Katz, Lebowitz and Spohn [62, 63] introduced a model system intended to capture the physics of fast ionic conductors [64]. Since then, this model has attracted a great deal of attention due to the diversity of phenomena it exhibits: so much so that it now frequently referred to as the "standard model" of a driven diffusive system. Detailed accounts of the known results for the model can be found in the reviews by Schmittmann and Zia [65] and by Marro and Dickman [66]: I select only a few features of interest for discussion here.

It is customary to define the Katz-Lebowitz-Spohn (KLS) model using the nearest-neighbour Ising model as a starting point. A configuration ω in an Ising model is a set of spins $\omega = \{s_i\}$ where the spin s_i on lattice site i is either pointing up ($s_i = 1$) or down ($s_i = -1$). In the absence of a field, the Hamiltonian reads

$$\mathcal{H}(\omega) = -J \sum_{\langle i,j \rangle} s_i s_j \tag{4.1.1}$$

where $\langle i,j \rangle$ denotes pairs of nearest-neighbour lattice sites. In driven diffusive systems it is more common to think in terms of particles than spins and so one maps down-spins onto particles and up-spins onto holes through the relation $s_i = 2\tau_i - 1$ in which τ_i is the occupation number of site i. Then, if the number of particles is conserved, the Hamiltonian becomes

$$\mathcal{H}(\omega) = -4J \sum_{\langle i,j \rangle} \tau_i \tau_j - \text{const} \tag{4.1.2}$$

which, of course, is equivalent to (4.1.1). Note that positive J corresponds to an attractive interaction between particles, $J < 0$ a repulsive interaction and $J = 0$ a special case where the particle-particle interaction is simply a hard-core repulsion (i.e. the interaction energy between particles on neighbouring sites is zero, rising sharply to infinity for particles on the same site).

If one wishes the steady-state distribution to be a Boltzmann distribution, one constructs dynamics in terms of transition rates $\mathcal{W}(\omega \to \omega')$ that satisfy the detailed balance condition given in chapter 2 through equation (2.3.2), namely

$$\frac{\mathcal{W}(\omega \to \omega')}{\mathcal{W}(\omega' \to \omega)} = \exp(-\beta[\mathcal{H}(\omega') - \mathcal{H}(\omega)]) \, . \tag{4.1.3}$$

In the KLS model, the transition rates are constructed by first setting $\mathcal{W}(\omega \to \omega')$ to zero if ω' and ω differ by more than a nearest-neighbour particle-hole exchange. For the remaining configurations, one introduces the unit vector $\boldsymbol{k}$ that points in the direction that a particle must move to obtain the final configuration ω' from the initial configuration ω. Then the transition rates (4.1.3) are defined through

$$\frac{\mathcal{W}(\omega \to \omega')}{\mathcal{W}(\omega' \to \omega)} = \exp(-\beta[\mathcal{H}(\omega') - \mathcal{H}(\omega) - \boldsymbol{k} \cdot \boldsymbol{E}]) \tag{4.1.4}$$

in which $\boldsymbol{E}$ is some uniform 'electric' field. In words this equation implies that transitions for which the particle displacement direction $\boldsymbol{k}$ is parallel to the field $\boldsymbol{E}$ are more likely to occur than those antiparallel to the field. Particle hops transverse to the field are unaffected.

This alteration of the transition rates is not in itself sufficient to give rise to a nonequilibrium steady state (in the sense of section 2.3.6). For example, if the system is closed, one can simply introduce an electric field potential and incorporate it into the Hamiltonian $\mathcal{H}$ so that (4.1.3) and (4.1.4) are then equivalent. With other boundary conditions, such as periodic boundaries or open boundaries (where particles may enter or leave the system) it may be possible for a particle current to flow. This would be indicative of a nonequilibrium steady state.

One can understand this situation more clearly in terms of the detailed balance condition (2.3.3) and discussion of section 2.3. This stated that for any sequence of configurations $\omega_1, \omega_2, \ldots, \omega_n$, detailed balance is satisfied if and only if

$$\mathcal{W}(\omega_1 \to \omega_2)\mathcal{W}(\omega_2 \to \omega_3) \ldots \mathcal{W}(\omega_n \to \omega_1) = \mathcal{W}(\omega_1 \to \omega_n)\mathcal{W}(\omega_n \to \omega_{n-1}) \ldots \mathcal{W}(\omega_2 \to \omega_1) \, . \tag{4.1.5}$$

For rates defined by (4.1.4) and closed boundaries, any loop in configuration space must comprise an equal number of hops parallel and antiparallel to the field. Then, the ratio of the left-hand side to the right-hand side of (4.1.5) is easily shown from (4.1.4) to be one, implying that detailed balance is satisfied in the case of a closed system as claimed above.

If, however, periodic boundary conditions are imposed, one can construct a loop in configuration space using only hops parallel to the applied field. Then, the ratio of the left-hand to the right-hand side of (4.1.5) is some nonzero power of $e^{\beta E}$ and thus detailed balance is destroyed in a nonzero field. A similar situation emerges with open boundaries where particles are inserted at one edge of the lattice and removed from the other. In the following I briefly explore some of the consequences of periodic and open boundaries in the KLS model. For simplicity, only hard-core ($J = 0$) and attractive ($J > 0$) particle-particle interactions will be considered.

4.1.1 KLS model with periodic boundaries

In the case $J = 0$, the Hamiltonian (4.1.1) has no dependence on the spins s_i, and so each configuration is equally likely in equilibrium. The same is also true for the steady state of a driven system with periodic boundary conditions. This can be seen by noting that for every configuration ω' that can be reached from configuration ω, there is a configuration ω'' from which ω is accessed at the same rate—see figure 4.1. Then the rate of flow of probability into configuration ω is exactly balanced by the outflow from ω as long as the probability for the system to be in a particular configuration is the same as every other, and so the probability distribution $P^*(\omega) = \text{const}$ is a stationary distribution. This balance of transition rates is a generalisation of detailed balance (recall section 2.3) sometimes referred to as dynamic reversibility [15] or pairwise balance [67].

For positive J, the undriven Ising system exhibits (in dimensions greater than one) a phase transition at a critical temperature T_c above which the mean magnetisation is zero and below which the magnetisation is spontaneously either positive or negative. For example, on the

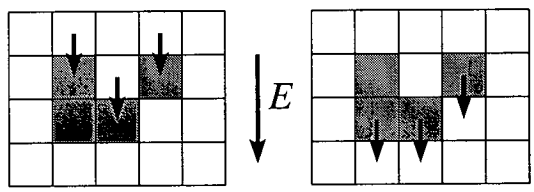

Figure 4.1: The left figure shows the number of ways into a particular configuration using particle hops parallel to the field; the right figure shows the corresponding number of ways out of the same configuration, again using particle hops parallel to the field. The fact that the two are equal for all configurations when periodic boundaries are imposed implies that every configuration is equally likely in the steady state.

two-dimensional square lattice, and with the Hamiltonian (4.1.1), $T_c = 0.567J/k_B$ where k_B is Boltzmann's constant.

In the driven case, this transition survives [62, 63], although T_c rises with E. Also, there are differences associated with the fact that the dynamics conserve particle number; in the spin picture this implies that the total magnetisation must remain constant and so it is not possible to have a transition from a phase with zero magnetisation to one with a nonzero magnetisation as the temperature is varied. Instead it is found that the KLS model assumes a disordered homogeneous phase above T_c and phase separates into regions of high and low density below T_c. One should note that the dynamics of phase separation, along with the related phenomenon of coarsening (see e.g. [68]), is a generic nonequilibrium phenomenon of great interest in its own right. Also of interest are the interfaces between the regions of high and low density: some aspects of interfacial dynamics are studied later, in chapter 6, where references to the literature are also provided.

A particular benefit of having the Ising model as a special equilibrium case of the nonequilibrium KLS model is that, because the two-dimensional Ising model has been exactly solved (see e.g. [69]), one can immediately recognise new phenomena as being a consequence of the driving force alone. Two noteworthy nonequilibrium effects in the KLS model are long-range (power-law) correlations above T_c and that at criticality there are two lengthscales that diverge with different exponents due to the anisotropy induced by the driving force. The generic power-law correlations should be compared with power-laws in equilibrium systems that generally arise only at isolated critical points. For a full discussion of these (and other) effects, one should consult the comprehensive reviews [65, 66] and references therein.

4.1.2 KLS model with open boundaries

In the case of open boundaries, particles are inserted at one edge of the lattice and removed at the opposite edge. Even in the absence of a bulk drive, the open boundaries enforce a current of particles across the lattice and this, combined with attractive interactions between the particles, is found to give rise to long-range (power-law) density correlations in the steady state [70].

With the drive E reactivated, the nonequilibrium effects of both the bulk drive and the boundary conditions combine to give unusual 'finger' density configurations [65], i.e. triangular high-density regions pointing in the direction of the drive. More significantly, it was pointed out by Krug [71] that nonequilibrium bulk dynamics combined with open boundaries can, in fact, lead to phase transitions induced by a change in the parameters controlling the boundary interactions. This particular feature will be demonstrated explicitly for the partially asymmetric

exclusion process below. The case of finite systems with open boundaries is of particular interest since these are more realistic than models with periodic boundary conditions and hence potentially more accessible in experimental situations.

4.2 Partially asymmetric exclusion process (PASEP)

The partially asymmetric exclusion process (PASEP) is a one-dimensional limit of the KLS model with open boundaries. In this section I define this model, indicate some of its many applications and establish some easily identified properties.

4.2.1 Definition and applications

Using as a starting point the KLS model defined on a one-dimensional lattice comprising N sites, the partially asymmetric exclusion process is obtained by first setting the parameter J in (4.1.2) to zero. That is, the particles interact through hard-core exclusion alone. Then, a parameter q is introduced through $q = \mathrm{e}^{-\beta E}$ such that the rate at which particles hop against the field (taken to be to the left) is a factor of q times the hop-rate in the same direction as the field (to the right). Two additional parameters that are related to the boundaries are also introduced: particles are inserted at the left boundary with rate α and extracted from the right boundary with rate β. Figure 4.2 illustrates these dynamics. Note that the rightward (or forward) hop-rate has been set to unity with no loss of generality, and hence that the leftward (reverse) hop-rate is equal to q.

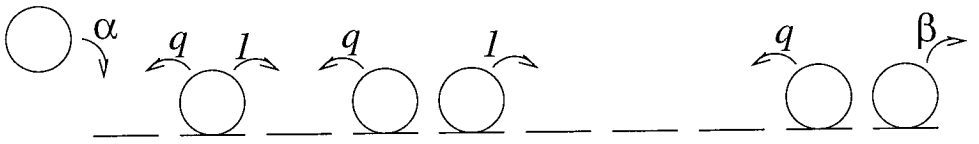

Figure 4.2: Typical particle configuration and dynamics of the partially asymmetric exclusion process.

This definition of the model's dynamics should be supplemented by an *updating scheme*—i.e. whether the elementary rules are performed in parallel, in series or at random. Here a random sequential update is assumed: that is, in each infinitesimal time interval $\mathrm{d}t$ at most one particle move occurs and the particle selected for update is chosen entirely at random. In terms of transition rates, this implies that $\mathcal{W}(\omega \to \omega') = 0$ if the configurations ω and ω' differ by more than one particle move.

Although the PASEP has been introduced here in the context of the driven lattice gas, it should be noted that the study of exclusion processes predates the work of Katz, Lebowitz

and Spohn. The interest in mathematical circles since the 1970s stems from its being a simple realisation of interacting Markov processes [72, 34]. Even before that, an exclusion process had been constructed to model the kinetics of biopolymerisation on nucleic acid templates [73]. This experimental situation is perhaps best described in the authors' own words:

> The synthesis of DNA and RNA on DNA templates, being catalysed by exoenzymes both have the property of progressing sequentially along their templates by the addition of monomers only. ... In the case where only one polymer chain is permitted to grow on a given template at a given time, but in which allowance is made for a possible back reaction (depolymerisation), such a polymerisation may be viewed as the diffusion of a single point along a one-dimensional lattice. ... In order to incorporate some additional features known about certain biopolymerisations on nucleic acid templates, we now extend the model to the diffusion of several nonoverlapping segments on the lattice.

This system maps onto the PASEP directly: the particles represent the large ribosomes (enzymes) catalysing the polymerisation: each time monomers are added to the polymer, the ribosome hops forwards; the depolymerisation corresponds to the reverse hop; and the current of particles measures the overall rate of polymerisation. The only difference between the model studied in [73] and the PASEP is that the biophysicists were interested in particles that are larger than a single lattice site.

A further application of exclusion processes is to transport through microscopic channels such as those found in biological membranes or other porous media. In this case, the width of a channel is of the same order of magnitude as that of the particles being transported though it. Hence a one-dimensional exclusion process is appropriate for its description. In the system described in [74] there are two species in baths at opposite ends of a channel and each species may enter and exit at both ends of the channel. Additionally, the forward and reverse hopping-rates of each species are independent—see figure 4.3. This model is clearly a generalisation of the partially asymmetric exclusion process and the physical interest in the system is the result

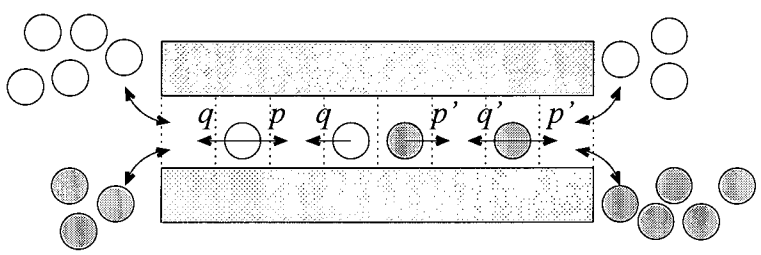

Figure 4.3: The model of transport through a narrow channel, recreated from [74]. The model comprises two particle species (represented by open and filled circles) which have independent left- and right-hopping rates. Each of the eight insertion and extraction rates (not shown) are also independent. Compare this model with the partially asymmetric exclusion process, figure 4.2.

that it is possible for direction of flow of a particle species to be in the opposite direction to that expected from simple arguments based on osmotic pressures alone [74].

A further application of exclusion processes is to traffic flow on a single-lane road. Many models are based on variants of exclusion processes (see e.g. [75]), particularly those where the particles (vehicles) hop simultaneously, i.e. the system is defined in terms of parallel update dynamics rather than the random sequential dynamics that are considered here. Some additional discussion of the extension of the methods used in this chapter to the case of parallel dynamics can be found in section 4.8.

From a fundamental viewpoint, the appeal of exclusion processes is not just in the fact that many exact results are known (including the new results for the PASEP derived below) but also because the boundary-induced phase transitions of the type described by Krug [71] are present even in one dimension. As we will see in sections 4.5 and 4.6 these phase transitions are manifested as nonanalyticies in the current and density profiles as one varies the boundary parameters α and β.

4.2.2 Master equation

A configuration of particles and holes (empty sites) on the lattice can be completely specified by the set of occupation numbers $\{\tau_j\}$, $j = 1, 2, \ldots, N$ where $\tau_j = 1$ if site j is occupied and $\tau_j = 0$ otherwise. Using the vector-space notation of chapter 2 we introduce the probability vector

$$\boldsymbol{P_t} = \sum_{\{\tau_j\}} P_t(\tau_1, \ldots, \tau_N) \boldsymbol{e}(\tau_1, \ldots, \tau_N) \tag{4.2.1}$$

where $\boldsymbol{e}(\tau_1, \ldots, \tau_N)$ is a unit vector that represents the specified configuration and $P_t(\tau_1, \ldots, \tau_N)$ is the probability of observing that configuration at time t (given some suitable initial condition).

From the definition of the PASEP, we see that the only interactions are between pairs of neighbouring sites in the bulk, and between the boundary sites and the outside world. Thus the transition rate matrix W in the master equation

$$\frac{\mathrm{d}}{\mathrm{d}t} \boldsymbol{P}_t = W \boldsymbol{P}_t \tag{4.2.2}$$

can be decomposed into a left-boundary matrix W_L, a set of matrices $W_{j,j+1}$ that act on pairs of neighbouring sites $(j, j+1)$ and a right-boundary matrix W_R. Explicitly

$$W = W_L + \sum_{j=1}^{N-1} W_{j,j+1} + W_R \,. \tag{4.2.3}$$

As a consequence of the random sequential updating scheme, in which only one particle may be moved per infinitesimal time step, the number of nonzero elements of these matrices is quite small. The nonzero off-diagonal elements are given through the transition rates

$$\begin{aligned} \mathcal{W}_L\left(\tau_1 = 0 \to \tau_1 = 1\right) &= \alpha \\ \mathcal{W}_{j,j+1}\left(\tau_j = 1, \tau_{j+1} = 0 \to \tau_j = 0, \tau_{j+1} = 1\right) &= 1 \\ \mathcal{W}_{j,j+1}\left(\tau_j = 0, \tau_{j+1} = 1 \to \tau_j = 1, \tau_{j+1} = 0\right) &= q \\ \mathcal{W}_R\left(\tau_N = 1 \to \tau_N = 0\right) &= \beta\,. \end{aligned} \tag{4.2.4}$$

Note that the occupation numbers not explicitly specified in the above definitions are not changed in the transition. Additionally recall that the diagonal elements of W are set through the condition that the column sums of a transition rate matrix must be zero, as discussed in chapter 2.

The main quantity of interest is the steady-state probability distribution $\boldsymbol{P}^*$ that satisfies

$$\frac{\mathrm{d}}{\mathrm{d}t}\boldsymbol{P}^* = W\boldsymbol{P}^* = 0\,. \tag{4.2.5}$$

One can verify that the matrix defined by (4.2.3) and (4.2.4) is irreducible and so from the discussion in chapter 2 we know that there is only one stationary distribution $\boldsymbol{P}^*$. As was argued in section 4.1, the open boundaries imply that detailed balance is not satisfied and thus from the discussion of section 2.3.6 we do not exact $\boldsymbol{P}^*$ to be expressible in terms of a simple Hamiltonian.

We also wish to calculate steady-state ensemble averages of observables as defined in section 2.1.2. In particular the steady-state density ρ_j and the current (flux) of particles $J_{j,j+1}$ between sites j and $j+1$ are defined as

$$\rho_j = \lim_{t\to\infty}\langle\tau_{j+1}\rangle \tag{4.2.6}$$

$$J_{j,j+1} = \lim_{t\to\infty}\langle(\tau_j(1-\tau_{j+1}) - q(1-\tau_j)\tau_{j+1}\rangle \tag{4.2.7}$$

in which the meaning of the angle brackets is as defined by equation (2.1.1) and t is time. In the steady-state, the current must be the same at all points on the lattice and so the subscripts of J will often be omitted.

4.2.3 Basic properties

Before I present the exact solution for the steady state of the PASEP, I illustrate some of the properties of the model that can be appreciated from its definition alone. In particular I identify a symmetry (invariance) under particle-hole exchange and an important notion called *bias* just by studying figure 4.2.

Symmetry under particle-hole exchange

Consider the lattice configuration indicated in figure 4.2. By exchanging particles and holes, one obtains the system illustrated in figure 4.4 in which particles are inserted at the right boundary at rate β, hop with rate 1 towards the left, rate q towards the right and exit at the left at rate α.

If one now reflects figure 4.4 about a vertical axis, the same system as that illustrated in figure 4.2 is recovered, except that the roles of the insertion rate α and the extraction rate β have been exchanged. Mathematically, this invariance under particle-hole exchange and reflection implies that the steady-state density obeys

$$\rho_j(\alpha, \beta, q) = 1 - \rho_{N+1-j}(\beta, \alpha, q) \tag{4.2.8}$$

and the steady-state particle current

$$J(\alpha, \beta, q) = J(\beta, \alpha, q) \ . \tag{4.2.9}$$

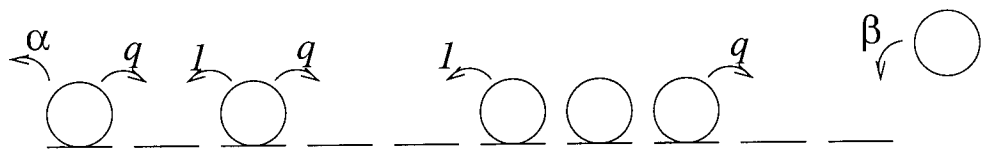

Figure 4.4: The partially asymmetric exclusion process of figure 4.2 after a particle-hole exchange. Note the model maps onto a reflection of itself with the roles of α and β exchanged.

The role of bias

Under the particle-hole exchange just described, the role of the parameter q is unaffected: it always expresses the rate at which particles hop towards the boundary where particles are inserted but not removed. If this rate q is less than one, the particles perform random walks (subject to the exclusion constraint) that are biased towards the boundary where particles are removed. Thus in this case we expect an uninhibited flow (current) of particles across the entire system. This situation will be called the *forward-bias* regime.

Conversely, when $q > 1$, particles prefer to hop towards the boundary where they cannot escape. Typically, therefore, there will be a build-up of particles at that boundary, and in order for any new particles to be inserted, all those on the lattice must hop at least one site away from the boundary. The probability that this collective action occurs will be extremely small and therefore we expect a much smaller current in this *reverse-bias* regime as compared with the case of forward bias.

The physicist's instinct is to home in on any transition points: here the case of no bias ($q = 1$). In fact this choice of q is particularly simple to treat, and so I discuss this first.

4.2.4 Exact solution in the absence of bias

In the absence of bias ($q = 1$), the steady-state current between two sites (4.2.7) takes a very simple exact form

$$J_{j,j+1} = \langle \tau_j \rangle - \langle \tau_{j+1} \rangle = \rho_j - \rho_{j+1} \,. \tag{4.2.10}$$

As has already been noted, the steady-state current has no dependence on the position j, and in particular the current between any two neighbouring sites on the lattice must equal the rate at which particles leave the system at the right boundary. That is,

$$\rho_j - \rho_{j+1} = \beta \rho_N \quad \Rightarrow \quad \rho_j = (1 + (N - j)\beta)\rho_N \,. \tag{4.2.11}$$

The rate at which particles enter the system, $\alpha(1 - \rho_1)$, must also equal the rate at which particles leave and so

$$\alpha(1 - \rho_1) = \alpha[1 - (1 + (N - 1)\beta)\rho_N] = \beta \rho_N \tag{4.2.12}$$

whose solution for ρ_N implies that

$$\rho_j = \frac{\alpha(1 + (N - j)\beta)}{\alpha + \beta + \alpha\beta(N - 1)} \quad \text{and} \quad J = \frac{\alpha\beta}{\alpha + \beta + \alpha\beta(N - 1)} \,. \tag{4.2.13}$$

Thus for $q = 1$, the density decreases linearly from the left boundary to the right and, in the limit $N \to \infty$ the density at the left boundary approaches unity whilst that at the right vanishes for all α, β larger than zero.

The ease with which these results were obtained is due entirely to the cancellation of the two-point density functions $\langle \tau_j \tau_{j+1} \rangle$ in expression (4.2.7) for the current when $q = 1$. This does not occur in general and some understanding of the case $q \neq 1$ can be gained by applying the mean-field approximation $\langle \tau_j \tau_k \rangle \to \langle \tau_j \rangle \langle \tau_k \rangle$. The results obtained [73, 76, 77, 16] are in qualitative (although not entirely quantitive) agreement with those I now to proceed to derive exactly for general q.

4.3 Exact solution of the PASEP

4.3.1 The matrix product method

In this section I describe the matrix product method for the construction of a steady-state probability distribution for an exclusion process. Before applying the method to obtain new results for the partially asymmetric exclusion process, I discuss its development since the original application to the *totally* asymmetric exclusion process (TASEP), i.e. the present model with the reverse hop-rate q set to zero.

Before the advent of the matrix product method, the TASEP was treated by obtaining expressions for the steady-state weights (unnormalised configurational probabilities) for a system of size N in terms of those for a system of size $N-1$ [78]. By iterating these one can, in principle, relate any steady-state weight to the easily solved $N = 1$ system. In [78] this procedure was followed for the case $q = 0, \alpha = \beta = 1$ to obtain the density profile ρ_j and current J. Subsequently [79] these recursion relations were used to calculate correlation functions for the same set of parameters.

Shortly afterwards, in the seminal paper of Derrida *et al.* [80], it was shown that the recursion relations could be expressed more conveniently and powerfully in terms of reduction relations for matrix products. This allowed immediate treatment of general α and β with $q = 0$. Since then, the matrix method has been applied to other models, including the partially asymmetric exclusion process of this chapter and further generalisations which will be discussed in section 4.8.

I now describe the way in which the matrix method is employed in the specific context of the PASEP. For a given set of occupation variables τ_i one writes an ordered product of matrices $A_{\tau_1} A_{\tau_2} \dots A_{\tau_N}$ in which $A_1 = D$ (site occupied) and $A_0 = E$ (site empty). To obtain a probability (a scalar value) from this matrix product, two vectors $\langle W|$ and $|V\rangle$ are employed in the following way:

$$P^*(\tau_1, \dots, \tau_N) = \frac{\langle W| A_{\tau_1} A_{\tau_2} \dots A_{\tau_N} |V\rangle}{Z_N} \,. \tag{4.3.1}$$

The factor Z_N is included to ensure that the probability distribution is properly normalised. This normalisation (analogous to a partition function—see section 2.3.2) has the following simple matrix expression through which a new matrix C is defined:

$$Z_N = \langle W|(D+E)^N|V\rangle = \langle W|C^N|V\rangle \,. \tag{4.3.2}$$

Note that in the matrix product expression (4.3.1) the same matrix D is used each time a particle is present in a configuration. Nevertheless, dependence of a probability on particles' positions arises from commutation relations imposed on D and E and their relationship to the vectors $\langle W|$ and $|V\rangle$. As I now explain, the origin of these commutation relations is the set of allowed interactions expressed through the master equation (4.2.2).

First one introduces the vector of steady-state weights

$$\boldsymbol{f} = \sum_{\{\tau\}} \langle W| A_{\tau_1} A_{\tau_2} \dots A_{\tau_N} |V\rangle \, \boldsymbol{e}(\tau_1 \tau_2 \dots \tau_N) \tag{4.3.3}$$

and it is emphasised that the vector space of the A matrices is *not* the same as that of the configuration space spanned by the vector $\boldsymbol{f}$. It is now assumed that the action of the matrices

W_L, $W_{j,j+1}$, W_N introduced in (4.2.3) on $\boldsymbol{f}$ can be expressed as

$$-W_L\boldsymbol{f} = \sum_{\{\tau\}}\langle W|X_{\tau_1}A_{\tau_2}\ldots A_{\tau_N}|V\rangle\,\boldsymbol{e}(\tau_1\tau_2\ldots\tau_N) \tag{4.3.4}$$

$$W_{j,j+1}\boldsymbol{f} = \sum_{\{\tau\}}\langle W|\ldots A_{\tau_{j-1}}(X_{\tau_j}A_{\tau_{j+1}} - A_{\tau_j}X_{\tau_{j+1}})A_{\tau_{j+2}}\ldots|V\rangle\,\boldsymbol{e}(\tau_1\tau_2\ldots\tau_N) \tag{4.3.5}$$

$$W_R\boldsymbol{f} = \sum_{\{\tau\}}\langle W|A_{\tau_1}\ldots A_{\tau_{N-1}}X_{\tau_N}|V\rangle\,\boldsymbol{e}(\tau_1\tau_2\ldots\tau_N) \tag{4.3.6}$$

where $X_1 = \widetilde{D}$ and $X_0 = \widetilde{E}$ are matrices complementary to the particle matrix D and the hole matrix E respectively. Putting $\boldsymbol{P} = \boldsymbol{f}/Z_N$ into the master equation (4.2.2) one finds the terms in the summation cancel and thus that $\boldsymbol{P}$ is the stationary distribution $\boldsymbol{P}^*$.

Contracting each of the above three expressions from the left with an appropriate unit (row) vector and using the definitions of W_L, $W_{j,j+1}$, W_R through (4.2.4) one finds that

$$\widetilde{D}E - D\widetilde{E} = E\widetilde{D} - \widetilde{E}D = qED - DE \tag{4.3.7}$$

$$\widetilde{D}D - D\widetilde{D} = \widetilde{E}E - E\widetilde{E} = 0 \tag{4.3.8}$$

$$-\langle W|\widetilde{D} = \langle W|\widetilde{E} = \alpha\langle W|E \tag{4.3.9}$$

$$\widetilde{E}|V\rangle = -\widetilde{D}|V\rangle = \beta D|V\rangle\,. \tag{4.3.10}$$

This set of relations can be satisfied if one sets $\widetilde{D} = -1$ and $\widetilde{E} = 1$. Then one finds that if the following relations are satisfied, equation (4.3.1) correctly describes the steady-state probability distribution for the PASEP:

$$DE - qED = D + E \tag{4.3.11}$$

$$\langle W|E = \langle W|\frac{1}{\alpha} \tag{4.3.12}$$

$$D|V\rangle = \frac{1}{\beta}|V\rangle\,. \tag{4.3.13}$$

Although only two lattice-site states (occupied and empty) were considered here, the method generalises to a larger number of states, e.g. where there is more than one type of particle on the lattice. The matrix product method has been shown rigorously to be equivalent to the problem of solving for the steady-state probability distribution in the presence of nearest-neighbour interactions [21]. However when there are more than two lattice-site states, the auxiliary matrices X are generally not simple scalars, nor is there a recipe for finding them.

A particular benefit of the matrix product method is that the expectation values of observables have simple matrix expressions. For example, the density

$$\rho_j = \sum_{\{\tau\}}\tau_j P^*(\tau_1,\ldots,\tau_N) = \frac{\langle W|C^{j-1}DC^{N-j}|V\rangle}{Z_N} \tag{4.3.14}$$

and the current

$$J_{j,j+1} = \frac{\langle W|C^{j-1}(DE - qED)C^{N-j-1}|V\rangle}{Z_N} = \frac{Z_{N-1}}{Z_N} \tag{4.3.15}$$

where the last equality follows from relation (4.3.11). We see that, as expected in the steady state, the current is independent of the bond chosen. The determination of new results for the PASEP now amounts to the evaluation of the matrix products in the above expressions for Z_N, J and ρ_i. This is achieved (essentially) by applying the rules (4.3.11)–(4.3.13).

4.3.2 Matrix reordering and the q-harmonic oscillator

The matrix expressions and algebraic relations derived in the previous section define a reordering problem of the type discussed in chapter 3 of this thesis. Consider the normalisation Z_N (4.3.2) and its implicit representation in terms of α, β and q through (4.3.11)–(4.3.13). If (4.3.11) is used repeatedly (4.3.2) can be developed as

$$Z_N = \langle W|(D+E)^N|V\rangle = \sum_{n,m} a_{N,n,m}\langle W|E^n D^m|V\rangle = \langle W|V\rangle \sum_{n,m} a_{N,n,m}\alpha^{-n}\beta^{-m} \tag{4.3.16}$$

where the intermediate step contains a sum of products (weighted by the coefficients $a_{N,n,m}$), each with any E matrices appearing to the left of any D matrices. This is precisely the situation we had in the stochastic ballistic annihilation and coalescence (SBAC) model of chapter 3 and indeed this expression for the normalisation has a very similar form to that of the matrix product (3.4.6) which was evaluated for the SBAC model. Furthermore, the reduction relation (4.3.11) is a special case of that satisfied by the matrices L and R that appeared in the previous chapter—cf. equation (3.3.4). Hence we can use the methods of the previous chapter to find the normalisation and therewith the current using (4.3.15).

To be more specific about the connection, recall the q-harmonic oscillator ladder operators a and $a^\dagger$ which satisfy

$$aa^\dagger - qa^\dagger a = 1 - q\,. \tag{4.3.17}$$

One can verify that if

$$D = \frac{1+a}{1-q} \quad \text{and} \quad E = \frac{1+a^\dagger}{1-q} \tag{4.3.18}$$

equation (4.3.11) is equivalent to (4.3.17). Again $\langle W|$ and $|V\rangle$ are eigenvectors of $a^\dagger$ and a respectively, and once again v and w denote the eigenvalues:

$$\langle W|a^\dagger = \langle W|w \quad \text{and} \quad a|V\rangle = v|V\rangle \tag{4.3.19}$$

with

$$w = \frac{1-q}{\alpha} - 1 \quad \text{and} \quad v = \frac{1-q}{\beta} - 1 \tag{4.3.20}$$

which is easily found from (4.3.12), (4.3.13) and (4.3.18). In order to calculate the normalisation, one must evaluate

$$Z_N = \langle W|(D+E)^N|V\rangle = \frac{\langle W|[(a+a^\dagger)+2]^N|V\rangle}{(1-q)^N} \ . \tag{4.3.21}$$

Recall now equation (3.4.43) from the previous chapter:

$$F_N(v,w) = \frac{1}{\mathcal{N}}\langle W|(f_L L + f_R R)^N|V\rangle = \frac{1}{\mathcal{N}}\langle W|[\mathcal{K}(a+a^\dagger)+\mathcal{L}]^N|V\rangle \ . \tag{4.3.22}$$

Comparison of this with (4.3.21) above reveals that the normalisation in the PASEP model has a near-identical form to the generating function that was of great importance in understanding the SBAC model of the previous chapter. The differences lie only in the values of the parameters v, w, $\mathcal{K}$, $\mathcal{L}$ and $\mathcal{N}$.

In section 3.4.2 I explained a procedure for recasting (4.3.22) as an integral which could subsequently be analysed in the large-N limit. This involved a number of steps which are summarised as follows.

Step 1: Definition of boson-number and coordinate bases The boson-number basis $\{|n\rangle\}$ was introduced through the operator relationships $a|n\rangle = \sqrt{1-q^n}|n-1\rangle$, $a^\dagger|n\rangle = \sqrt{1-q^{n+1}}|n+1\rangle$. The coordinate basis $\{|\theta\rangle\}$ arose from the eigenvalue equation $(a+a^\dagger)|\theta\rangle = 2\cos(\theta)|\theta\rangle$.

Step 2: Coordinate representation of the boson-number state vectors By evaluating the product $\langle n|(a+a^\dagger)|\theta\rangle$, and putting $h_n(\theta) = \langle n|\theta\rangle = \langle\theta|n\rangle$ we found the three-term recurrence relation

$$2\cos(\theta)h_n(\theta) = \sqrt{1-q^{n+1}}\,h_{n+1}(\theta) + \sqrt{1-q^n}\,h_{n-1}(\theta) \ . \tag{4.3.23}$$

From this it was a simple matter to obtain the generating function of the q-Hermite polynomials

$$G(\theta,x) = \sum_{n=0}^{\infty} \frac{x^n}{\sqrt{(q;q)_n}} h_n(\theta) = \frac{1}{(xe^{i\theta}, xe^{-i\theta};q)_\infty} \tag{4.3.24}$$

and thence an explicit form for $h_n(\theta)$ through a series expansion of the right-hand side. (Recall the q-factorial notation defined in section 3.2.1).

Step 3: Orthogonality of the q-Hermite polynomials By using the solution of

$$\int \mathrm{d}\theta\, \nu(\theta)\, h_n(\theta)h_m(\theta) = \delta_{n,m} \tag{4.3.25}$$

for $\nu(\theta)$ we constructed the identity $\mathbb{I} = \int \mathrm{d}\theta \nu(\theta)|\theta\rangle\langle\theta|$ which, when inserted into (4.3.22) gave

$$F_N(v,w) = \frac{1}{\mathcal{N}}\int \mathrm{d}\theta \langle W|\theta\rangle[2\mathcal{K}\cos(\theta)+\mathcal{L}]^N\langle\theta|V\rangle \ . \tag{4.3.26}$$

Step 4: Evaluation of the remaining inner products It transpired that the inner products $\langle W|\theta\rangle$ and $\langle\theta|V\rangle$ were both equal to the generating function of the q-Hermite polynomials (4.3.24) required to complete step 2 above.

To treat (4.3.21) one follows exactly the same series of steps. For the forward-bias regime ($q < 1$), the procedure is identical to that just described and one can simply take the results of section 3.4.2 and substitute in the values of the parameters $\mathcal{K}, \mathcal{L}$ and $\mathcal{N}$ that are appropriate here. See section 4.4.1 below and also independent contemporaneous results [81] for the forward-bias regime.

In the present work, I also treat the reverse-bias regime ($q > 1$). There is, however, a technical difficulty one must first overcome, in that step 3 of the above procedure cannot be completed with the three-term recurrence relation appearing as it does in equation (4.3.23). The reason for this is that if the weight function $\nu(\theta)$ is to exist, the coefficients that appear in a three-term recurrence relation must be real. Therefore, to treat the case $q > 1$, some manipulation of (4.3.23) must be performed first (see section 4.4.2 below).

Before I present exact expressions for the normalisation Z_N for general α, β and q I remark on some special cases. Firstly, the algebra defined by (4.3.11)–(4.3.13) can be satisfied by the scalars

$$D = \frac{1}{\beta} \quad \text{and} \quad E = \frac{1}{\alpha} \tag{4.3.27}$$

as long as $\alpha + \beta = 1 - q$. In other words, the matrices D and E have a one-dimensional representation for this range of parameters and every configuration with the same number of particles on the lattice has the same steady-state probability: i.e. the model exhibits no correlations. One can also find finite-dimensional representations ($n \times n$ matrices) of the algebra [82] which again are restricted to particular combinations of α, β and q.

Furthermore, in the limit $q \to 1$, the representation (4.3.18) diverges which implies that the use of the q-harmonic oscillator algebra is not beneficial here. Indeed, as shown in section 4.2.4 above, the case $q = 1$ can be solved without recourse to matrices. Nevertheless, I later (in section 4.4.3) derive an exact formula for the normalisation from the integral that applies for $q < 1$. This formula is easily extended to *all* values of the model parameters α, β and q and will be of particular benefit when it comes to analysing the reverse-bias ($q > 1$) regime.

4.4 Exact expressions for the normalisation

I now use the connection to the q-deformed oscillator algebra to calculate exact expressions for the steady-state normalisation in the partially asymmetric exclusion process. First I derive integral representations of the normalisation that apply in the forward- and reverse-bias regimes

(resulting in equations (4.4.4) and (4.4.15) respectively). Then, in section 4.4.3, I obtain an exact expression for the normalisation in the form of a sum with a finite number of terms (as long as the system size N too is finite). This expression, equation (4.4.44), is valid for all combinations of the model parameters.

4.4.1 Integral representation in the forward-bias regime

As stated above, the results of section 3.4.2 can be transferred to the present case by comparing equations (4.3.21) and (4.3.22). One finds

$$Z_N = F_N(v,w) \quad \text{if} \quad \mathcal{N} = (1-q)^N, \quad \mathcal{K} = 1 \quad \text{and} \quad \mathcal{L} = 2 \tag{4.4.1}$$

where $F_N(v,w)$ is the generating function that was considered in the SBAC model.

Thus from section 3.4.2 we know that one form of Z_N is the integral over θ

$$Z_N = \left(\frac{2}{1-q}\right)^N \int_0^\pi \mathrm{d}\theta\, \nu(\theta) G(\theta,w)[\cos(\theta)+1]^N G(\theta,v) \tag{4.4.2}$$

in which

$$\nu(\theta) = \frac{(q, e^{2i\theta}, e^{-2i\theta}; q)_\infty}{2\pi} \quad \text{and} \quad G(\theta,x) = \frac{1}{(x\mathrm{e}^{i\theta}, x\mathrm{e}^{-i\theta}; q)_\infty}\,. \tag{4.4.3}$$

Alternatively, one can convert this to the contour integral

$$Z_N = \frac{(q;q)_\infty}{4\pi i}\left(\frac{1}{1-q}\right)^N \oint_K \frac{\mathrm{d}z}{z}[z+z^{-1}+2]^N \frac{(z^2, z^{-2}; q)_\infty}{(wz, w/z, vz, v/z; q)_\infty} \tag{4.4.4}$$

where the contour K is the unit circle $|z| = 1$ and is directed anti-clockwise.

Recall that in the derivation of these integrals, it was assumed that $|v| < 1$ and $|w| < 1$ and that the expressions for Z_N must be analytically continued to cater for other values of v, w when they arise.

4.4.2 Integral representation for $q > 1$

It was noted earlier that there is a problem with the three-term recurrence relation (4.3.23) in that the coefficients of the functions $h_n(\theta)$ are imaginary if $q > 1$. Favard's theorem [57, §5.2] then no longer holds, which means that a positive weight function $\nu(\theta)$ that orthogonalises the functions $h_n(\theta)$ does not exist with (4.3.23) in its present form.

To address this problem one makes the change of variable due to Askey [83], namely $\theta \to \pi/2 - iu$ (or equivalently $\cos(\theta) \to i\sinh(u)$). Then (4.3.23) becomes

$$2\sinh(u)h_n(u) = \sqrt{q^{n+1}-1}\,h_{n+1}(u) + \sqrt{q^n-1}\,h_{n-1}(u) \tag{4.4.5}$$

in which all the prefactors are positive. We are now assured from Favard's theorem that for the new set of q-Hermite polynomials

$$\langle n|u\rangle = \langle u|n\rangle = h_n(u) = \frac{i^n}{\sqrt{(q;q)_n}} \sum_{k=0}^{n} (-)^k \begin{bmatrix} n \\ k \end{bmatrix}_q \mathrm{e}^{(n-2k)u} \tag{4.4.6}$$

there exists a positive weight function such that

$$\int \mathrm{d}u\, \nu(u) h_n(u) h_m(u) = \delta_{n,m} \,. \tag{4.4.7}$$

In [83] Askey showed that

$$\nu(u) = \frac{1}{\ln q} \frac{1}{(q^{-1}, -q^{-1}e^{2u}, -q^{-1}e^{-2u}; q^{-1})_\infty} \quad \text{for} \quad u \in (-\infty, \infty) \tag{4.4.8}$$

is the correct choice for $\nu(u)$. Note now that the range of integration (i.e. domain of ν) is infinite.

Finally, one must determine the inner products $\langle W|u\rangle$ and $\langle u|V\rangle$. The boson-number representation of $\langle W|$ implies that

$$\langle W|u\rangle = \sum_{n=0}^{\infty} \langle W|n\rangle\langle n|u\rangle = \sum_{n=0}^{\infty} \frac{w^n}{\sqrt{(q;q)_n}} h_n(u) = G(u,w) \tag{4.4.9}$$

where $G(u,w)$ is the generating function of q-Hermite polynomials for $q > 1$. By multiplying (4.4.5) on both sides by $x^n/\sqrt{(q;q)_n}$ and summing over n it is found that

$$G(u,x) = (1 - iq^{-1}x\mathrm{e}^{u})(1 + iq^{-1}x\mathrm{e}^{-u}) G(u, q^{-1}x) \,. \tag{4.4.10}$$

Note that this expression relates the value of $G(u,x)$ to itself with $x \to q^{-1}x$ (i.e. a smaller value of x) and so it can be used iteratively to find

$$G(u,x) = (iq^{-1}xe^{u}, -iq^{-1}xe^{-u}; q^{-1})_\infty G(u,0) = (iq^{-1}xe^{u}, -iq^{-1}xe^{-u}; q^{-1})_\infty \tag{4.4.11}$$

since $G(u,0) = 1$. The same procedure can be used to find $\langle u|V\rangle$ and to summarise

$$\langle W|u\rangle = G(u,w) = (iq^{-1}we^{u}, -iq^{-1}we^{-u}; q^{-1})_\infty \tag{4.4.12}$$

$$\langle u|V\rangle = G(u,v) = (iq^{-1}ve^{u}, -iq^{-1}ve^{-u}; q^{-1})_\infty \,. \tag{4.4.13}$$

In contrast to the case of forward bias, these inner products are equal to their respective infinite products for all v and w.

Inserting the identity $\mathbb{I} = \int_{-\infty}^{\infty} \mathrm{d}u\, \nu(u) |u\rangle\langle u|$ into (4.3.21) we find

$$Z_N = \left(\frac{1}{1-q}\right)^N \int_{-\infty}^{\infty} \mathrm{d}u\, \nu(u) G(u,w) [2(1 + i\sinh u)]^N G(u,v) \tag{4.4.14}$$

which, when written out in full becomes

$$Z_N = \frac{1}{\ln q}\frac{1}{(q^{-1};q^{-1})_\infty}\left(\frac{2}{1-q}\right)^N \int_{-\infty}^{\infty} \mathrm{d}u\,(1+i\sinh u)^N \frac{(iq^{-1}ve^{u}, -iq^{-1}ve^{-u}, iq^{-1}we^{u}, -iq^{-1}we^{-u}; q^{-1})_\infty}{(-q^{-1}e^{2u}, -q^{-1}e^{-2u}; q^{-1})_\infty} . \quad (4.4.15)$$

Note that in contrast to the case of $q < 1$, this expression cannot be conveniently rewritten as an integral over a closed contour.

4.4.3 Explicit formula valid in all regimes

In this section I derive (using a number of standard results from the study of q-series [56]) an alternative expression for Z_N which takes the form of a finite sum rather than an integral and is valid for all values of the model parameters. Such an expression is useful for two main reasons: firstly it allows us to extract the asymptotic form of the normalisation when $q > 1$; secondly, as the sum contains a finite number of terms, it can be evaluated numerically if one wishes to study finite-sized systems.

I work from the integral for $q < 1$, $|v| < 1$ and $|w| < 1$ (4.4.2) which can be written as the generic functional

$$\mathcal{I}[f] = \int_0^\pi \mathrm{d}\theta\,\nu(\theta)G(\theta,w)G(\theta,v)f(\theta) . \quad (4.4.16)$$

Here f is an arbitrary real function of $\theta \in [0,\pi]$ and the functions $\nu(\theta)$ and $G(\theta,t)$ are as given by (4.4.3). As is well known, $f(\theta)$ can be expressed as a sum of orthogonal functions, as long as the latter form a complete set. The q-Hermite polynomials $h_n(\theta)$ form such a set, and so one can write

$$\mathcal{I}[f] = \sum_n a_n \int_0^\pi \mathrm{d}\theta\,\nu(\theta)G(\theta,w)G(\theta,v)h_n(\theta) = \sum_n a_n I_n \quad (4.4.17)$$

in which the expansion coefficients a_n are found by evaluating

$$a_n = \int_0^\pi \mathrm{d}\theta\,\nu(\theta)h_n(\theta)f(\theta) . \quad (4.4.18)$$

The benefit of making this transformation is that the integral

$$I_n = \int_0^\pi \mathrm{d}\theta\,\nu(\theta)G(\theta,w)G(\theta,v)h_n(\theta) \quad (4.4.19)$$

has a closed-form solution, as will be shown below. Therefore, in order to find an explicit formula for $\mathcal{I}[f]$, one first needs to know this closed-form solution and then to compute the expansion coefficients a_n. In the case of the integral (4.4.2) the function $f(\theta) = (2[\cos\theta + 1]/[1-q])^N$ which I now consider explicitly in the following[1].

[1] The reader not interested in the technical details of this calculation is advised to skip ahead to the result, equation (4.4.44).

Solution of the integral I_n

The solution of the integral I_n arises from an important identity satisfied by the generating function of Hermite polynomials $G(\theta,t)$ (4.3.24). This identity, proved by Askey and Wilson [84], reads

$$\int_0^\pi \mathrm{d}\theta\,\nu(\theta)G(\theta,a)G(\theta,b)G(\theta,c)G(\theta,d) = \frac{(abcd;q)_\infty}{(ab,ac,ad,bc,bd,cd;q)_\infty} \tag{4.4.20}$$

and holds as long as the denominator of the right-hand side is nonzero (e.g. $ab \neq q^{-j}, j = 0,1,2,\ldots$). By setting $a = w$, $b = v$, $c = x$ and $d = 0$ one finds

$$\int_0^\pi \mathrm{d}\theta\,\nu(\theta)G(\theta,w)G(\theta,v)G(\theta,x) = \frac{1}{(vw,vx,wx;q)_\infty} \tag{4.4.21}$$

and that for the case being considered here ($v < 1$, $w < 1$) one can always find a value of x for which the right-hand side is finite.

By writing $G(\theta,x)$ out explicitly in terms of Hermite polynomials (4.3.24) one finds from the previous expression that

$$\sum_{n=0}^{\infty} \frac{x^n}{\sqrt{(q;q)_n}} \int_0^\pi \mathrm{d}\theta\,\nu(\theta)G(\theta,w)G(\theta,v)h_n(\theta) = \frac{1}{(vw,vx,wx;q)_\infty}\,. \tag{4.4.22}$$

Then, using the identity (3.4.27) for the infinite products on the right-hand side of the previous expression, and comparing powers of x we find

$$I_n = \frac{1}{(vw;q)_\infty}\frac{1}{\sqrt{(q;q)_n}} \sum_{k=0}^{n} \begin{bmatrix} n \\ k \end{bmatrix}_q v^{n-k}w^k\,. \tag{4.4.23}$$

The sum on the right-hand side of this expression can be thought of as a q-deformation of the binomial expansion of $(v+w)^n$ and will henceforth be denoted as $B_n(v,w;q)$. That is, the equation

$$I_n = \frac{1}{(vw;q)_\infty}\frac{B_n(q;q)_n}{\sqrt{(q;q)_n}} \tag{4.4.24}$$

supplemented with the definition

$$B_n(v,w;q) = \sum_{k=0}^{n} \begin{bmatrix} n \\ k \end{bmatrix}_q v^{n-k}w^k \tag{4.4.25}$$

are the main results of this calculation.

As an aside, we note that these polynomials obey

$$B_n(v,w;q) = v^n \sum_{k=0}^{n} \begin{bmatrix} n \\ k \end{bmatrix}_q (w/v)^k = v^n \widetilde{B}_n(w/v;q) \tag{4.4.26}$$

in which $\widetilde{B}_n(x;q)$ are Szegö polynomials which have a number of properties similar to those of the q-Hermite polynomials [85].

Expansion of an arbitrary function $f(\theta)$ in q-Hermite polynomials

I now turn to the problem of determining the coefficients a_n in the expansion of an arbitrary function $f(\theta), \theta \in [0, \pi]$ in q-Hermite polynomials

$$f(\theta) = \sum_{n=0}^{\infty} a_n h_n(\theta) \ . \tag{4.4.27}$$

This problem was first tackled by Rogers in 1894 [60] in his study of the q-Hermite polynomials. His method was to note that the q-Hermite polynomial can be expressed explicitly as a Fourier cosine series

$$h_n(\theta) = \sum_{k=0}^{n} \begin{bmatrix} n \\ k \end{bmatrix}_q c_{n-2k}(\theta) \tag{4.4.28}$$

in which

$$c_m(\theta) = \begin{cases} 0 & m < 0 \\ 1 & m = 0 \\ 2\cos(m\theta) & m > 0 \end{cases} \ . \tag{4.4.29}$$

Rogers then represented $\cos(m\theta)$ in terms of q-Hermite polynomials by means of a longhand solution of the set of linear equations defined by (4.4.28).

Over a hundred years later, we have at our disposal an orthogonality relation for the q-Hermite polynomials, which means that one can evaluate the coefficients a_n more efficiently using (4.4.18). We set

$$f(\theta) = [2(1 + \cos\theta)]^N = \sum_n a_n h_n(\theta) \tag{4.4.30}$$

in accordance with the integral (4.4.2), but leaving aside for the moment the prefactor $(1-q)^{-N}$, and note that

$$[2(1 + \cos\theta)]^N = \sum_{n=0}^{N} \binom{2N}{N-n} c_n(\theta) \tag{4.4.31}$$

Now we expand $c_n(\theta)$ in q-Hermite polynomials

$$c_n(\theta) = \sum_{k=0}^{\lfloor \frac{n}{2} \rfloor} b_{n,k} h_{n-2k}(\theta) \ . \tag{4.4.32}$$

This form arises by noting from (4.4.28) that only those q-Hermite polynomials of the same parity as the cosine function c_n appear in the expansion and there can be no contributions from $h_k(\theta)$ with $k > n$.

The coefficients $b_{n,k}$ are now found by applying (4.4.18) after writing the weight function $\nu(\theta)$ as a series rather than a product (this is a special case of Jacobi's triple product, see e.g. [56, App. II]):

$$\nu(\theta) = \frac{1}{2\pi} \sum_{s=-\infty}^{\infty} (-)^s q^{\binom{s}{2}} (1 + q^s)\, e^{2is\theta} \ . \tag{4.4.33}$$

One finds, after some manipulation of the resulting integral that

$$b_{n,k} = \frac{1}{\sqrt{(q;q)_{n-2k}}} \sum_{s=-\infty}^{\infty} (-)^s q^{\binom{s}{2}} (1+q^s) \sum_{r=0}^{n-2k} \begin{bmatrix} n-2k \\ r \end{bmatrix}_q \frac{1}{2\pi} \int_0^{2\pi} \mathrm{d}\theta \, \cos((k+r-s)\theta) \; . \quad (4.4.34)$$

The solution of the integral that appears under the summation is just $\delta_{s,k+r}$, which allows one to eliminate the sum over s. This gives

$$\begin{aligned} b_{n,k} &= \frac{(-1)^k}{\sqrt{(q;q)_{n-2k}}} \sum_{r=0}^{n-2k} \begin{bmatrix} n-2k \\ r \end{bmatrix}_q (-1)^r q^{\binom{k+r}{2}} (1+q^{k+r}) \\ &= \frac{(-1)^k q^{\binom{k}{2}}}{\sqrt{(q;q)_{n-2k}}} \sum_{r=0}^{n-2k} \begin{bmatrix} n-2k \\ r \end{bmatrix}_q (-1)^r q^{\binom{r}{2}+kr} \left(1+q^{k+r}\right) \\ &= \frac{(-1)^k q^{\binom{k}{2}}}{\sqrt{(q;q)_{n-2k}}} \left((q^k;q)_{n-2k} + q^k (q^{k+1};q)_{n-2k}\right) \end{aligned} \quad (4.4.35)$$

in which the series expansion

$$(x;q)_n = \sum_{k=0}^{n} \begin{bmatrix} n \\ k \end{bmatrix}_q (-1)^k q^{\binom{k}{2}} x^k \quad (4.4.36)$$

was used [56, App. II].

This latest expression allows a little simplification by noting that

$$(aq^k;q)_{n-2k} = \frac{(a;q)_{n-k}}{(a;q)_k} \quad (4.4.37)$$

and so we find

$$b_{n,k} = (-1)^k \sqrt{(q;q)_{n-2k}}\, q^{\binom{k}{2}} \left(\begin{bmatrix} n-k-1 \\ k-1 \end{bmatrix}_q + q^k \begin{bmatrix} n-k \\ k \end{bmatrix}_q \right) \quad (4.4.38)$$

which is true for all n and k if one observes the convention that $\begin{bmatrix} n \\ k \end{bmatrix}_q = 0$ when $k < 0$ or $k > n$.

Equations (4.4.31), (4.4.32) and (4.4.38) can now be combined to find

$$\begin{aligned} [2(1+\cos\theta)]^N &= \sum_{n=0}^{N} \binom{2N}{N-n} \sum_{k=0}^{\lfloor \frac{n}{2} \rfloor} b_{n,k}(q) h_{n-2k}(\theta) \\ &= \sum_{n=0}^{N} \sum_{k=0}^{\lfloor \frac{N-n}{2} \rfloor} \binom{2N}{N-(n+2k)} b_{n+2k,k} h_n(\theta) \\ &= \sum_{n=0}^{N} R_{N,n}(q) \sqrt{(q;q)_n}\, h_n(\theta) \end{aligned} \quad (4.4.39)$$

in which

$$R_{N,n}(q) = \sum_{k=0}^{\lfloor \frac{N-n}{2} \rfloor} (-1)^k \binom{2N}{N-n-2k} q^{\binom{k}{2}} \left\{ \begin{bmatrix} n+k-1 \\ k-1 \end{bmatrix}_q + q^k \begin{bmatrix} n+k \\ k \end{bmatrix}_q \right\} \quad (4.4.40)$$

which may be alternatively written as

$$R_{N,n}(q) = \sum_{k=0}^{\lfloor \frac{N-n}{2} \rfloor} (-1)^k \left[\binom{2N}{N-n-2k} - \binom{2N}{N-n-2k-2} \right] q^{\binom{k+1}{2}} \begin{bmatrix} n+k \\ k \end{bmatrix}_q . \quad (4.4.41)$$

Explicit formula

Putting the results (4.4.17), (4.4.24) and (4.4.39) together, and reintroducing the prefactor $(1-q)^{-N}$, the normalisation as given by the integral (4.4.2) becomes

$$Z_N = \frac{1}{(vw;q)_\infty} \left(\frac{1}{1-q} \right)^N \sum_{n=0}^{N} R_{N,n}(q) B_n(v,w;q) \quad (4.4.42)$$

in which $R_{N,n}(q)$ is given by (4.4.40) or (4.4.41) and $B_n(v,w;q)$ by (4.4.25).

This exact formula, valid for $q<1$, $|v|<1$ and $|w|<1$ admits extension to general q, v and w. Note first that the infinite product in the prefactor can be replaced with $\langle W|V\rangle$, a fact which follows from (3.4.27):

$$\frac{1}{(vw;q)_\infty} = \sum_{n=0}^{\infty} \frac{(vw)^n}{(q;q)_n} = \langle W|V\rangle . \quad (4.4.43)$$

I now claim that the resulting representation-independent expression for the normalisation

$$Z_N = \langle W|V\rangle \left(\frac{1}{1-q} \right)^N \sum_{n=0}^{N} R_{N,n}(q) B_n(v,w;q) \quad (4.4.44)$$

holds for *all* choices of the model parameters.

This is justified by observing that once v and w are written in terms of α and β using (4.3.20) we obtain a power series in α and β of the same form as (4.3.16). As discussed in section 4.3.2, an equation with this structure arises when one re-orders a matrix product directly using the relation (4.3.11). To perform this direct manipulation, it is not necessary to employ a specific representation of the quadratic algebra. Therefore, although the representation used above to derive (4.4.44) breaks down for $q=1$, $q<1, |v|>1$ or $q<1, |w|>1$, had one chosen a representation that *does* converge in the region of interest, equation (4.4.44) would still have been obtained for the normalisation.

As a check of this formula, let us consider the case $q=0$. Then (4.4.41) and (4.4.25) become

$$R_{N,n}(0) = \binom{2N}{N-n} - \binom{2N}{N-n-2} \quad (4.4.45)$$

$$B_n(v,w;0) = \frac{v^{n+1} - w^{n+1}}{v-w} \quad (4.4.46)$$

where now $v = 1/\beta - 1$ and $w = 1/\alpha - 1$. It can be verified, using the identity

$$\sum_{n=r}^{X-Y} (-1)^n \binom{X}{Y-n} \binom{n}{r} = (-1)^r \binom{X-1-r}{Y-r} , \quad (4.4.47)$$

that (4.4.44) can be rewritten as

$$Z_N = \langle W|V\rangle \sum_{k=0}^{N} \left[\binom{2N-2-k}{N-k} - \binom{2N-2-k}{N-2-k} \right] \left[\frac{(1/\beta)^{k+1} - (1/\alpha)^{k+1}}{1/\beta - 1/\alpha} \right] \tag{4.4.48}$$

which is equivalent to the result obtained in [80].

Finally in this section I note that for $q \to 1$ the singularity in the denominator of (4.4.44) is cancelled by the sum over $R_{N,n}(q)B_n(v,w;q)$ and the expression is in fact well-behaved. Although I have checked for small system sizes that (4.4.44) agrees with the results in section 4.2.4 I have not been able to show this in a simple way; nevertheless it is known [77] that the matrix relation (4.3.11) with $q = 1$ gives the same results as those in section 4.2.4 and so the summation (4.4.44) should too.

4.5 Phase diagram for the current

The results that were obtained above for the normalisation Z_N are sufficient to determine an exact phase diagram for the partially asymmetric exclusion process. I use the methods described in section 3.4.3 to analyse the integral (4.4.2) and thence to determine the current in the forward-biased regime ($q < 1$). For the case of reverse bias ($q > 1$), the analysis is conducted by approximating the exact sum formula (4.4.44).

4.5.1 The forward bias regime ($q < 1$)

Recalling the contour integral for the normalisation (4.4.4) which reads

$$Z_N = \frac{(q;q)_\infty}{4\pi i} \left(\frac{1}{1-q} \right)^N \oint_K \frac{\mathrm{d}z}{z} [z + z^{-1} + 2]^N \frac{(z^2, z^{-2}; q)_\infty}{(wz, w/z, vz, v/z; q)_\infty} \tag{4.5.1}$$

I list the results that were obtained in section 3.4.3 for an integral of the same form, valid in the limit of large N (i.e. the leading terms in an asymptotic expansion). The only difference is that here the parameters take the values $\mathcal{N} = (1-q)^N$, $\mathcal{K} = 1$ and $\mathcal{L} = 2$.

Case A $(v < 1, w < 1)$

$$Z_N \sim \frac{4}{\sqrt{\pi}} \frac{(q;q)_\infty^3}{(v,w;q)_\infty^2} \left(\frac{4}{1-q} \right)^N \frac{1}{N^{\frac{3}{2}}} \tag{4.5.2}$$

Case B1 $(v = 1, w < 1)$

$$Z_N \sim \frac{2}{\sqrt{\pi}} \frac{(q;q)_\infty}{(w;q)_\infty^2} \left(\frac{4}{1-q} \right)^N \frac{1}{N^{\frac{1}{2}}} \tag{4.5.3}$$

Case B2 $(v < 1, w = 1)$

$$Z_N \sim \frac{2}{\sqrt{\pi}} \frac{(q;q)_\infty}{(v;q)_\infty^2} \left(\frac{4}{1-q}\right)^N \frac{1}{N^{\frac{1}{2}}} \tag{4.5.4}$$

Case C1 $(v > 1, w \leq 1)$

$$Z_N \sim \frac{(v^{-2};q)_\infty}{(vw, w/v;q)_\infty} \left(\frac{v + v^{-1} + 2}{1-q}\right)^N \tag{4.5.5}$$

Case C2 $(v \leq 1, w > 1)$

$$Z_N \sim \frac{(w^{-2};q)_\infty}{(vw, v/w;q)_\infty} \left(\frac{w + w^{-1} + 2}{1-q}\right)^N \tag{4.5.6}$$

In contrast to the SBAC model, v and w are independent quantities in the PASEP. In particular, both may be greater than one and/or equal. There are thus a further four cases to treat.

Case B3 $(v = 1, w = 1)$

Here there would appear to be a problem because the integrand of (4.5.1) is infinite at the saddle-point $z = 1$. This suggests that the saddle-point analysis is not the way to determine the normalisation Z_N in this case. In fact, the correct approach is much simpler.

In terms of the original model parameters, the condition $v = w = 1$ implies $\alpha = \beta = (1-q)/2$ and so $\alpha + \beta = (1-q)$. This is precisely the condition given in section 4.3.2 for the operators D and E to have a one-dimensional (scalar) representation $D = 1/\alpha$ and $E = 1/\beta$. Substituting this into the original expression for Z_N (equation (4.3.2)) one finds

$$Z_N = \langle W|[D+E]^N|V\rangle = \left(\frac{4}{1-q}\right)^N \tag{4.5.7}$$

in which the inner product $\langle W|V\rangle$ has been set to unity[2].

Case C3 $(v > w > 1)$

Although there are now poles at $q^k w$ which lie at $z > 1$, none of these lies to the right of the pole at $z = v$, and so that will still supply the dominant contribution to the integral (4.4.4). Thus the result of case C1 applies here: that is,

$$Z_N \sim \frac{(v^{-2};q)_\infty}{(vw, w/v;q)_\infty} \left(\frac{v + v^{-1} + 2}{1-q}\right)^N \tag{4.5.8}$$

[2]The fact that the inner-product has been chosen differently for $v = w = 1$ to elsewhere is not important as results for different values of v and w are never combined.

Case C4 $(w > v > 1)$

By the same argument of the previous paragraph, the result C2 applies here.

$$Z_N \sim \frac{(w^{-2};q)_\infty}{(vw, v/w;q)_\infty}\left(\frac{w+w^{-1}+2}{1-q}\right)^N \tag{4.5.9}$$

Case C5 $(w = v > 1)$

Here, two poles lie at $z = w = v$ and we must evaluate their contribution to Z_N by differentiating the analytic part of the integrand with respect to z and putting $z = w = v$ in the result. Although there is, in principle, the daunting prospect of differentiating an infinite product, one can verify that the dominant term for $N \gg 1$ takes the form

$$Z_N \sim \frac{(v^{-2};q)_\infty}{(q, v^2;q)_\infty}(v - v^{-1})N\left(\frac{v+v^{-1}+2}{1-q}\right)^{N-1}. \tag{4.5.10}$$

Taking these asymptotic results (4.5.2) to (4.5.10) for Z_N the current can be computed for large N by evaluating $J = Z_{N-1}/Z_N$ (4.3.15). The results are presented in table 4.1. The regions for which the different forms apply are indicated in the form of a phase diagram in figure 4.5. Note that the apparently large number of cases enumerated above reduces to just three current phases. In particular, at a phase boundary, e.g. $\alpha = \beta < (1-q)/2$, the current approaches the same value from either side of the boundary. We also remark that the results obtained here confirm predictions based on mean-field arguments [76].

In figure 4.5, the distinct current phases have been labelled L, H and M. These correspond to a low-density phase, high-density phase and maximal-current phase respectively. That the current takes the largest value realisable in the system everywhere in phase M is easily confirmed from table 4.1. It is not immediately evident (until I consider density profiles more

Region	Current J
$\alpha \geq \frac{1-q}{2}, \beta \geq \frac{1-q}{2}$	$\frac{1-q}{4}$
$\alpha \leq \frac{1-q}{2}, \beta \geq \alpha$	$\frac{\alpha(1-q-\alpha)}{1-q}$
$\beta \leq \frac{1-q}{2}, \alpha \geq \beta$	$\frac{\beta(1-q-\beta)}{1-q}$

Table 4.1: The $N \to \infty$ forms of the particle current in the forward biased phases.

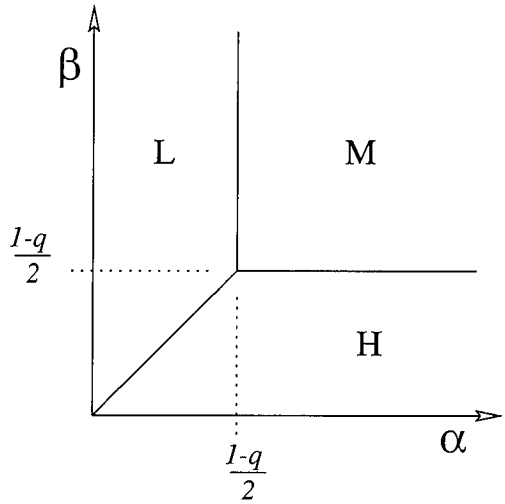

Figure 4.5: The phase diagram for the current in the forward-bias regime. The symbols L, H and M correspond to the low-density, high-density and maximal-current phases respectively.

closely in section 4.6 below) that phases L and H correspond to a low and high mean occupation of the lattice. However this seems likely because in the former the particle insertion rate α is smaller than the extraction rate β and vice-versa in the latter.

To investigate the nature of the currents a little more closely it is useful to plot their dependence on α for fixed β below and above the transition point $\beta = \frac{1}{2}(1-q)$—see figure 4.6. Of particular interest is the discontinuity in the first derivative of the current with respect to α at the transition from the low density to the high density phase. Recall from section 2.4 that a discontinuity in the current strictly exists only in the limit of infinite system size N. Due to the particle-hole symmetry of the model (see section 4.2.3) the same behaviour is exhibited by the current at fixed α and as one varies β.

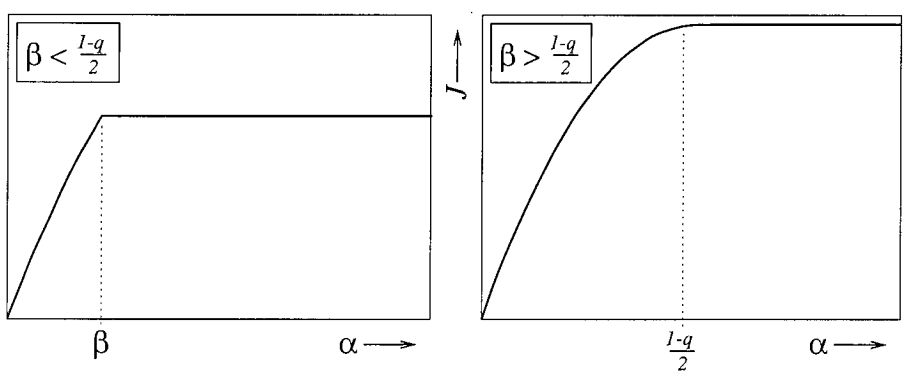

Figure 4.6: The dependence of the particle current in the PASEP on α for fixed β. On the left we see the sharp transition from the low-density to the high-density phase; on the right the transition to the maximal current phase is seen to be smooth.

There is another singularity at the transition line from either of the low- or high-density phases to the maximal current phase. In this case, it is the second derivative of the current that is discontinuous. Hence one can accordingly classify these two types of transition as first- and second-order; as will be discussed in section (4.6) below, the high- and low-density phases actually coexist along the line $\alpha = \beta < (1-q)/2$ and so this line is analogous to an equilibrium first-order transition.

4.5.2 The reverse bias regime ($q > 1$)

I turn now to the case $q > 1$. I will show shortly by examining the exact formula (4.4.44) that the normalisation behaves like $Z_N \sim q^{\frac{1}{4}N^2}$ for large N. This result indicates that a naive application of the saddle-point method on the integral (4.4.15) is not appropriate: by its nature, the saddle-point method gives expressions in which the exponent is linear in N rather than the desired quadratic.

In order to analyse the exact formula (4.4.44) I approximate the coefficients $R_{N,n}(q)$ and the polynomials $B_n(v, w; q)$ that appear in the summation. Both of these quantities contain two elementary objects, the q-factorial and the q-binomial coefficient, which I treat first.

The first step is to rewrite the q-factorial as

$$(q;q)_n = \prod_{j=1}^{n}(1-q^j) = (-1)^n q^{\binom{n+1}{2}} e^{M_n(q)} \tag{4.5.11}$$

where

$$M_n(q) = \sum_{j=1}^{n} \ln(1-q^{-j}) = -\sum_{k=1}^{\infty} \frac{1}{k}\frac{1}{q^k-1}(1-q^{-kn}) \,. \tag{4.5.12}$$

When n is large, $M_n(q)$ can be approximated by

$$M(q) \simeq -\sum_{k=1}^{\infty} \frac{1}{k}\frac{1}{q^k-1} \tag{4.5.13}$$

which is independent of n. This leads to an approximation of the q-binomial coefficient

$$\begin{bmatrix} n \\ k \end{bmatrix}_q \simeq q^{k(n-k)} e^{-M(q)} \tag{4.5.14}$$

which is valid when both n and k are large.

The function $B_n(v, w; q)$ is defined by the sum in equation (4.4.25). The dominant terms are those around $k = n/2$, and so the q-binomial is replaced with the approximation (4.5.14). Also the sum is recast as an integral over k which gives, for large n,

$$B_n(v, w; q) \sim (-1)^n A(v, w; q)\, q^{\frac{1}{4}n^2} |vw|^{\frac{1}{2}n} \tag{4.5.15}$$

where $A(v,w;q)$ is given by

$$A(v,w;q) = \sqrt{\frac{\pi}{\ln q}} \exp\left\{ M(q) + \frac{(\ln w/v)^2}{4\ln q} \right\} . \tag{4.5.16}$$

To deal with sum (4.4.40) for $R_{N,n}$ we keep only the term with largest k: the others are exponentially suppressed. Then

$$Z_N \simeq \frac{(q^{-1}vw;q^{-1})_\infty}{(1-q)^N} \left(S_N(v,w;q) + 2N S_{N-1}(v,w;q) \right) \tag{4.5.17}$$

where the product $\langle W|V\rangle$ has been expanded by using the identity

$$\sum_{r=0}^{\infty} \frac{(qx)^r}{(q;q)_r} = (x;q^{-1})_\infty \tag{4.5.18}$$

which holds when $q > 1$ and for all x and where

$$S_N(v,w;q) = (1-q^N) \sum_{r=0}^{\lfloor \frac{N}{2} \rfloor} (-1)^r q^{\binom{r}{2}} \frac{(q;q)_{N-r-1}}{(q;q)_r (q;q)_{N-2r}} B_{N-2r}(v,w;q) . \tag{4.5.19}$$

The main contribution to the sum $S_N(v,w;q)$ is where r is small. The approximation

$$(-1)^r q^{\binom{r}{2}} \frac{(q;q)_{N-r-1}}{(q;q)_r (q;q)_{N-2r}} \simeq -\frac{q^{-r^2+(N+1)r-N}}{(q;q)_r} , \tag{4.5.20}$$

which follows from (4.5.11) and (4.5.13), is valid in that region and when combined with the asymptotic expression for $B_n(v,w;q)$ yields

$$S_N(v,w;q) \sim (-1)^N A(v,w;q) (1-q^{-N}) |vw|^{\frac{1}{2}N} q^{\frac{1}{4}N^2} \sum_{r=0}^{\lfloor \frac{N}{2} \rfloor} \frac{1}{(q;q)_r} \left(\frac{q}{vw} \right)^r . \tag{4.5.21}$$

The single remaining summation may be estimated from the identity (4.5.18). Then

$$\sum_{r=0}^{\lfloor \frac{N}{2} \rfloor} \frac{1}{(q;q)_r} \left(\frac{q}{vw} \right)^r = (1/vw;q^{-1})_\infty + \mathrm{O}(q^{-\frac{1}{4}N^2}) \tag{4.5.22}$$

and so to leading order in q

$$S_N(v,w;q) \sim (-1)^N A(v,w;q) (1/vw;q^{-1})_\infty |vw|^{\frac{1}{2}N} q^{\frac{1}{4}N^2} \tag{4.5.23}$$

Noting that S_{N-1} is exponentially smaller than S_N the asymptotic form of Z_N for $q > 1$ is given by

$$Z_N \sim A(v,w;q) (q^{-1}vw, 1/vw; q^{-1})_\infty \left(\frac{\sqrt{vw}}{q-1} \right)^N q^{\frac{1}{4}N^2} \tag{4.5.24}$$

where the prefactor $A(v,w;q)$ is as defined in equation (4.5.16).

An expression for the reverse-bias current valid for large system sizes follows quickly from (4.3.15). This reads

$$J = \frac{Z_{N-1}}{Z_N} \sim \left(\frac{\alpha\beta(q-1)^2}{(q-1+\alpha)(q-1+\beta)} \right)^{\frac{1}{2}} q^{-\frac{1}{2}N+\frac{1}{4}} \tag{4.5.25}$$

which, in contrast with the currents in the forward bias regime, is a function of the number of lattice sites N and valid for all α and β. Hence we do not find any phase transitions in the reverse-bias regime.

4.6 Density profiles and correlations

Having solved the PASEP for the particle current, I now review further calculations performed by Sasamoto [86] which yield explicit expressions for the density profile in the forward-bias regime. I follow this up with some conjectures for density profiles in reverse-bias regime, for which no explicit calculations have yet been performed.

4.6.1 The forward-bias regime

Recall that the density (mean occupancy) is given by (4.3.14)

$$\rho_j = \frac{\langle W|C^{j-1}DC^{N-j}|V\rangle}{Z_N} \tag{4.6.1}$$

in which $C = D + E$. It is possible, by using methods similar to those presented above, to determine the current from this matrix product expression. The full calculations are presented step-by-step in [86]; here I wish only to indicate why the resulting phase diagram (figure 4.7) has a more complicated structure than that obtained for the current (figure 4.5).

To perform the calculations one works with the density difference between neighbouring sites $\Delta_j = \rho_j - \rho_{j+1}$ which has the matrix expression

$$\Delta_j = \frac{\langle W|C^{j-1}(DC-CD)C^{N-j-1}|V\rangle}{Z_N} = \frac{\langle W|C^{j-1}(DE-ED)C^{N-j-1}|V\rangle}{Z_N} \tag{4.6.2}$$

after using the fact that $DC - CD = DE - ED$ because $C = D + E$. The matrix $DE - ED$ is diagonal in the boson-number basis

$$DE - ED = \frac{aa^\dagger - a^\dagger a}{(1-q)^2} = \sum_n \frac{q^n}{1-q}|n\rangle\langle n| \tag{4.6.3}$$

as one can verify from (4.3.18) and (3.4.21). Insertion of two complete sets of coordinate basis vectors into (4.6.2) gives

$$\Delta_j = \frac{1}{Z_N}\frac{1}{(1-q)^{N-1}}\int_0^\pi \mathrm{d}\theta \int_0^\pi \mathrm{d}\phi\, G(\theta,w)[2(\cos\theta+1)]^{j-1} \left(\sum_{n=0}^{\infty} h_n(\theta)q^n h_n(\phi)\right)[2(\cos\phi+1)]^{N-j-1}G(\phi,v)\ . \quad (4.6.4)$$

The bracketed sum over n has an infinite product representation [57, §10.11]

$$\sum_{n=0}^{\infty} h_n(\theta)q^n h_n(\phi) = \left|\frac{(q;q)_\infty}{(qe^{i(\theta+\phi)}, qe^{-i(\theta-\phi)};q)_\infty}\right|^2 \quad (4.6.5)$$

whose insertion into (4.6.4) yields an integrand which comprises combinations of cosines and infinite products and can be analysed in a similar way to the integral for the normalisation in the forward-biased regime (4.4.2).

The fact that there are now two integrals to perform endows the results with some additional structure in that two of the phases found in section 4.5 subdivide into three—see figure 4.7 taken from [86]. The decay forms for $\beta > \alpha$ are given explicitly in table 4.2 and plotted schematically in figure 4.8. Those for $\beta < \alpha$ can be obtained by invoking the particle-hole symmetry as expressed through equation (4.2.8).

The maximal current phase (M) corresponds to a single density phase with the same boundaries $\alpha > (1-q)/2, \beta > (1-q)/2$. In this phase the bulk density is a half and the density increases as $\ell^{-1/2}$ towards the left boundary and decreases in the same way towards the right (in both cases ℓ is the distance from the boundary). This power-law density profile does not in itself imply long-range two-point density correlations; however long-range correlations have been confirmed explicitly [79] for the special case $\alpha = \beta = 1$, $q = 0$ and we expect similar results throughout the maximal current phase (M). This should be compared with the power-law correlations exhibited in the KLS model (section 4.1). There, the power-law correlations were observed above the critical temperature associated with the onset of spontaneous magnetisation in the Ising model whose origin lies in attractive interactions between particles. Here, the only interaction is hard-core exclusion, which corresponds to noninteracting Ising spins, and the temperature does not enter. In any case, the one-dimensional Ising model does not exhibit a phase transition and so we see that this behaviour lies entirely in the nonequilibrium boundary conditions.

A further point of interest in the maximal current phase is that the prefactor of the density decay is *independent* of q. This tells us something of the crossover to the symmetric case $q \to 1$. As one takes that limit, the maximal current phase is reached for smaller and smaller values of α and β: the maximal current phase takes over the whole of the phase diagram whilst the

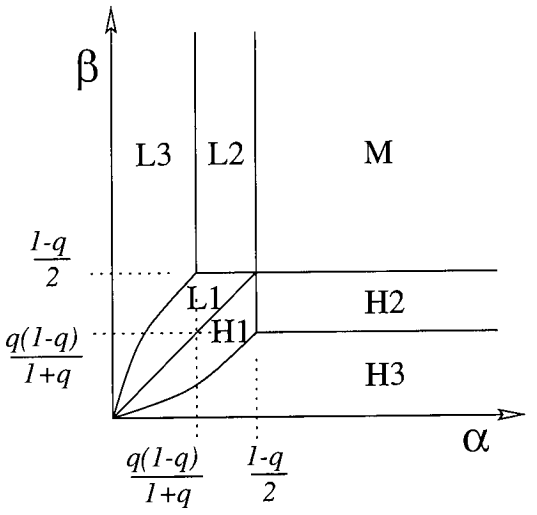

Figure 4.7: The phase diagram for the density in the forward-bias regime (taken from [86]). Note that the two phases H and L have subdivided into three subphases each.

Phase	Bulk $\rho_{N/2}$	Left boundary ρ_ℓ	Right boundary $\rho_{N+1-\ell}$
L1	$\frac{\alpha}{1-q}$	$\frac{\alpha}{1-q}$	$\frac{\alpha}{1-q} + a_1 \left(\frac{\alpha(1-q-\alpha)}{\beta(1-q-\beta)}\right)^\ell$
L2	$\frac{\alpha}{1-q}$	$\frac{\alpha}{1-q}$	$\frac{\alpha}{1-q} + a_2 \left(\frac{2}{1-q}\right)^{2\ell} \frac{(\alpha[1-q-\alpha])^\ell}{\ell^{3/2}}$
L3	$\frac{\alpha}{1-q}$	$\frac{\alpha}{1-q}$	$\frac{\alpha}{1-q} + a_3 \left(\frac{(q+\alpha)^2}{q}\right)^\ell$
M	$\frac{1}{2}$	$\frac{1}{2}\left(1+\frac{1}{\sqrt{\pi\,\ell}}\right)$	$\frac{1}{2}\left(1-\frac{1}{\sqrt{\pi\,\ell}}\right)$

Table 4.2: Density profiles in the low density (L) and maximal current (M) phases in the partially asymmetric exclusion process. The distance ℓ is relative to the specified boundary and $\{a_i\}$ is a set of prefactors given in [86] and which depend on α, β and q. The behaviour for the high density phases is obtained via the particle-hole symmetry $\rho_j(\alpha,\beta) = 1 - \rho_{N+1-j}(\beta,\alpha)$.

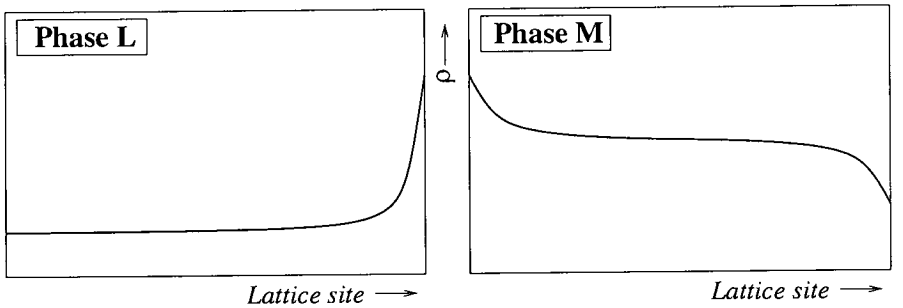

Figure 4.8: Schematic density plots in the one of the low-density phases L and the maximal current phase M. In the former, the density decay is exponential, or exponential times a power-law; in the latter case the density decays as a power-law from both boundaries.

asymptotic (large-N form) of the density profile remains unchanged. The system size required for this 'universal' density decay form to be seen increases as $q \to 1$. In fact, from explicit calculation [87] and dimensional analysis of the stochastic growth equations associated with exclusion processes (see section 6.2.1 and [88]), it can be shown that $u \propto (1-q)N^{1/2}$ is a *scaling variable* and that for $u \gg 1$ the asymptotic behaviour that has been calculated is observed. Hence, even with the slightest asymmetry in the particle dynamics, the nature of the density profile changes from linear across the lattice ($q = 1$) to power-law near the boundaries ($q < 1$) for sufficiently large N.

In the low density phases, the density is constant at the left boundary and decays exponentially (albeit with a power-law correction in phase L2). The boundary between phases L1 and L3 is given by the line $\alpha = q\beta/(1-\beta)$ and in the case of total asymmetry the phase L3 is not seen and with the boundary between L1 and L2 at $\beta = \frac{1}{2}$ [80].

On the line that separates phases L1 and H1 ($\alpha = \beta < (1-q)/2$) the ensemble-averaged density increases from the left boundary. It has further been determined [89, 90] that this average profile arises from the diffusion of a shock front (step profile) between a low-density region at the left of the lattice and a high-density region at the right. Hence, the high- and low-density phases coexist along the transition line $\alpha = \beta < (1-q)/2$ as remarked in section 4.5 above.

4.6.2 The reverse-bias regime

Although no explicit calculations have yet been performed for the density profile in the reverse-bias regime ($q > 1$), some conjectures can be made based on the information to hand. As was stated in section 4.2.3, we expect in the reverse bias-regime a build-up of particles at the left boundary which, because they cannot escape at the left boundary, are largely constrained from diffusing. I now argue that this implies the lattice is (on average) half full.

The build-up of particles at the left boundary implies that $\rho_j \simeq 1$ for $j \lesssim j_0$ and $\rho_j \simeq 0$ for $j \gtrsim j_0$ in which j_0 is the point at which the density decays rapidly to zero. Now, for a current to flow, we must have simultaneous stochastic diffusion of a particle from j_0 to the right boundary, and diffusion of a hole from j_0 to the left boundary. Both particles and holes diffuse at the same rate towards their respective boundary, and as the average time taken for the boundary to be approached must be the same in each case, we find that j_0 must be equal to $N/2$. In other words, the lattice is half full. That the current behaves as $J \sim q^{-N/2}$ is also suggestive of a half-full lattice. A similar argument for a half-full lattice was given in a study of transport through backbends ('uphill' segments) in a random network of bonds in a porous medium [91].

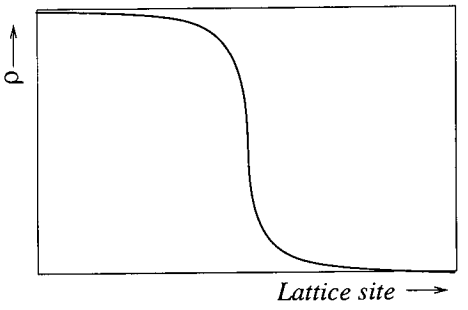

Figure 4.9: Proposed sigmoidal form of the density profile in the reverse bias regime

For $q < \infty$ we expect the density profile to decay rapidly, but continuously, around $j_0 = N/2$: i.e. in the style of a sigmoid—see figure 4.9. As the current decreases exponentially with system size, one can approximate the reverse-bias regime at large N with an asymmetric exclusion process with closed boundaries [92] in which the current is necessarily zero. In that case it is found that the density approaches the boundary values of 1 and 0 exponentially with a characteristic length that diverges as $q \to 1$ (where a flat profile results). In a study of a two-dimensional exclusion model [93] that coarsens into domains (which correspond to the high- and low-density regions of the present discussion) the interfacial profile was found to be well approximated by a Fermi function, i.e. $\rho_j \sim (1+\mathrm{e}^{\beta(j-j_0)})^{-1}$ in which β plays the role of inverse temperature.

4.7 Summary of results

So far in this chapter I have considered the partially asymmetric exclusion process (PASEP) in the context of a wider (and important) class of nonequilibrium systems, namely driven diffusive systems. In particular I obtained new results for the steady-state normalisation for the PASEP in the form of integral representations (equations (4.4.4) and (4.4.15)) and an exact summation formula, equation (4.4.44).

By using these expressions I was able to obtain the steady-state currents for large system sizes which in turn gave rise to the phase diagram shown in figure 4.5 and a proposal for the density profile in the reverse-bias regime (illustrated in figure 4.9). In the previous section I also reviewed a related work [86] in which the density phase diagram for the forward-bias regime was obtained exactly.

The results reported in this chapter were obtained using a matrix product method for the construction of the steady-state probability distribution function. In particular it became apparent that the commutation relations of the matrices required to solve the PASEP were nearly identical to those used in the previous chapter to determine the time evolution of the

model of stochastic ballistic annihilation and coalescence (SBAC). Hence the nonanalyticies that arose in the SBAC model and led to different density decay phases were also pertinent here and led to the phase diagram shown in figure 4.5.

The reason that similar methods could be applied to both models is that both reduce to sorting problems of the type described in section 3.2. In the case of the ballistic reaction system it was the particles themselves that were being sorted. In the PASEP the relationship to matrix reordering is more abstract and hence it is not obvious from a physical point of view that the two models should enjoy such a close mathematical relationship.

I now conclude this chapter with a brief review of results from the literature for the more general class of exclusion processes to illustrate the wide range of phenomena that arise.

4.8 Other studies of exclusion processes: a review

4.8.1 Dynamic properties

In this chapter, I have considered only static properties of the partially asymmetric exclusion process. However, there are a wealth of properties, such as correlation functions at unequal times, that give insight into the dynamics of exclusion processes.

One quantity that has received much attention is the distance travelled by a single (arbitrarily selected) particle in the asymmetric exclusion process with periodic boundary conditions. If, in a single realisation of the dynamics, the displacement of the chosen particle is $y(t)$ after a time t then one can define

$$v = \lim_{t\to\infty} \frac{\langle y(t)\rangle}{t} \quad \text{and} \quad \Delta = \lim_{t\to\infty} \frac{\langle y(t)^2\rangle - \langle y(t)\rangle^2}{t} \tag{4.8.1}$$

in which v is interpreted as the particle's mean velocity and Δ the diffusion constant of the process. These quantities were calculated using the matrix product method [94] and two particular limits ensue.

Firstly, if one takes the periodicity of the lattice to infinity, but keeps the number of occupied sites M constant, the diffusion constant $\Delta \to 4^{M+1}[(M+1)!]^2/[(2M+3)!]$. This number decreases monotonically with M which indicates that even in limit $\rho = M/N \to 0$, there are still sufficient collisions to impede the progress of a particle. In fact, the explanation for the decrease in Δ stems from the fact that collisions are correlated in time: if there are two particles on neighbouring sites at time t, they are quite likely still to be neighbours shortly afterwards. The other limit is when M/N is fixed as one takes both system size N and number of particles M to infinity. Then the diffusion constant decreases as a power-law $\Delta \sim N^{-1/2}$, a result that will be called upon in chapter 6 on interfacial growth.

One need not stop with the second moment of the displacement distribution, and indeed Derrida *et al.* [95, 96] show how one may obtain the scaling behaviour of the large-deviation function (which essentially measures the probability of observing rare events) by using a Bethe ansatz approach (see section 2.5.1) to calculate its generating function.

4.8.2 Multi-species exclusion processes

As hinted in section (4.3) the matrix method is applicable to systems with more than lattice states than just 'occupied' or 'empty'. In practice this involves a number of distinct particle species, each with its own dynamics. The most intensively studied case concerns particles, holes and stochastic defects on a periodic lattice.

In this case, the particles hop to the right at unit rate whereas a defect exchanges places with a hole to its right at rate α and with a particle to its left at rate β. Representing particles, holes and defects as the matrices D, E and A respectively, one finds the following algebraic relationships must be satisfied:

$$DE = D + E\,, \quad \beta DA = A \quad \text{and} \quad \alpha AE = A\,. \tag{4.8.2}$$

This can be achieved with D and E as used for the PASEP with $q = 0$, and with A given by the projector $A = |V\rangle\langle W|$. A steady state weight is then given by the trace of a matrix product (which ensures a translationally invariant distribution) and so, with a single defect one has a product of the form $\langle W|\prod_i X_i|V\rangle$ to evaluate, where X_i is either D or E. That is, the steady state weights for this model are exactly the same as those for the PASEP with $q = 0$. The only difference is that the normalisation is the sum over all configurations containing exactly M particles, and not over all M as is the case of open boundaries. In practice, one sidesteps this issue by allowing the number of particles to fluctuate by placing the system in equilibrium with a particle reservoir, and by then introducing fugacities to control the densities.

The model with a single defect has been used to study shock profiles [97] as the dynamics of the defect cause it to diffuse to a position where the density changes rapidly. Interestingly, if there are two defects on the ring they form an unusual bound state: the probability that one finds them a distance r apart decays as $r^{-3/2}$ so their mean separation is infinite. Additionally, the large deviation function of a single defect's displacement has been studied, again using a Bethe ansatz technique [98].

Another multi-species exclusion process [99, 100] comprises particles with opposite 'charges'. The positively-charged particles hop to the right with unit rate whilst the negatively-charged particles hop to the left with unit rate. The two exchange positions (so that $\oplus\ominus \to \ominus\oplus$) with rate q and the boundaries are open so that a current of both particle species can be

supported. The same current phases as those found for the PASEP (section 4.5) are present, in addition to new phases in which the currents of positively- and negatively-charged particles are different, even though the model is invariant under exchange of the two species: that is, the charge-symmetry is spontaneously broken.

A further variant of an exclusion process exhibits phase separation [101, 102]. This model comprises three particle species, A, B and C in which the dynamics are constructed to that A's prefer to be to the left of B's, B's to the left of C's and C's to the left of A's. With three species, the interfaces between domains are stable and, if one starts with an arbitrary configuration, the domain size coarsens logarithmically in time until it reaches the steady state in which there are just three domains, one for each species.

Finally I remark that a set of matrices for a model with four species (three types of particle and holes) has been found [103]. As in the PASEP, these take the form of semi-infinite matrices; however their elements are not scalars but the semi-infinite matrices D and E used in the PASEP. This structure implies a considerably more complicated set of matrix reduction relations (compared to $DE - qED = D + E$ of the PASEP) and performing any explicit calculations for the model with four or more species using the matrix method remains a challenge.

4.8.3 Simultaneous updating schemes

All models described so far have the common property of a random sequential updating scheme: that is, in an infinitesimal time interval, a single particle is chosen at random and made to hop with an appropriate probability. For certain applications, a simultaneous updating scheme is more suitable, e.g. where the particles are chosen for update in a prescribed order or in parallel. A case in point is that of traffic flow models (see e.g. [75] for a review). This is because cars move continuously and tend to accelerate sequentially from the start to the end of a queue rather than in some random order.

Perhaps the best-known model of this type is the Nagel-Schreckenberg model for traffic flow [104]. In this model, the roadway is divided into cells of approximately one car's length, and each cell is either occupied or empty. Cars are assigned an integer velocity which fluctuates according to the distance to the car in front with some noise that models differences between individual drivers. This velocity then specifies the number of lattice sites a car moves by in one time step, and is subject to an upper speed limit. In the special case of only one nonzero velocity, the model defines an asymmetric exclusion process in discrete time and with a simultaneous updating scheme.

Various types of simultaneous updating schemes exist. In a sublattice updating scheme, the lattice is divided up into disjoint pairs of neighbouring sites creating two sublattices (i.e. the

sublattices $(1,2)$, $(3,4)$, ... $(N-1,N)$ and $(2,3)$, $(4,5)$, ... $(N,1)$). In one time step the first sublattice is updated and in the following time step the other is updated. Alternatively, one can employ an ordered-sequential update, in which one starts at one end of the lattice, and considers each site in order, going either left-to-right or right-to-left. Finally, the dynamics may be fully parallel, in which all particles hop in concert. The various updating schemes for the asymmetric exclusion process are compared in [105] and the phase diagrams for the current and density are found to have the same basic form as that for the PASEP (sections 4.5 and 4.6).

A further variant, in which the hopping rates of particles are quenched random variables (i.e. each particle possess a hop rate which is drawn from some distribution and is kept by that particle for as long as it remains in the system) has been solved exactly for both random sequential [106] and simultaneous updating [107] schemes. In all cases an additional phase transition from a high-density homogeneous congested phase to a low-density inhomogeneous phase in which all the cars are in a queue behind the slowest moving vehicle. This can be understood as a direct consequence of allowing the particles to have different hopping rates if one considers the tailbacks that occur on country roads during the holiday season due to the presence of caravans.

Additionally, this scenario can be related to Bose condensation through the association of each vehicle at the head of a queue with a momentum state and of the size of the queue with the number of excitations of that state. Then one sees that the congested phase corresponds a finite occupancy of excited states whereas the inhomogeneous phase has total occupancy of the lowest momentum state. A similar 'jamming' transition occurs in a model in which the particles (buses) have the *same* hopping rates, but that long-range interactions are mediated by a third party (passengers) between them [108].

4.8.4 Connections to equilibrium statistical physics

Finally in this brief survey of exclusion processes I remark on some recent work in which connections with ideas from equilibrium statistical physics have been made, namely the Yang-Lee theory of phase transitions and the relationship between free energies and large deviation functions.

Recall from the discussion in section 2.4.1 that a phase transition should be accompanied by a convergence to the real axis of the zeros of the steady-state normalisation (partition function) expressed in terms of a suitable complex parameter. This was found by Arndt [109] to be the case for an asymmetric exclusion process in the presence of defect particles and periodic boundaries.

Specifically, the model in question contains a parameter q which controls the stability of the interface between a domain of ordinary and defect particles. At some critical value of $q = q_c$ something akin to a condensation transition occurs: for $q < q_c$, and suitable particle densities, there is a finite condensate whereas for $q > q_c$ the distribution of particles on the lattice is homogeneous. By introducing fugacities conjugate to the densities of ordinary and defect particles, and solving for the zeros of the partition function, Arndt found that for $q < q_c$ the zeros sat on ellipses that pinch the real-axis in the complex fugacity plane as one takes the system size to infinity. This corresponds to the first-order transition that heralds the appearance of the condensate for sufficiently large particle densities. For $q < q_c$, the partition function zeros lie on hyperbolæ and thus do not pinch the real axis: this indicates that no density induced transition takes place in this regime. Additionally, the analysis shows that the transition at $q = q_c$ is second order.

Finally I note a recent paper by Derrida *et al* [110] in which the probability of observing a particular coarse-grained density profile $\rho(x)$ in the symmetric exclusion process with open boundaries was determined. In an equilibrium system, this probability is given by $\exp(-N\mathcal{F}[\rho])$ in which $\mathcal{F}[\rho]$ is a local functional that expresses the free energy difference between the equilibrium profile and ρ. A form for a functional that plays a similar role in the symmetric exclusion process was found in [110] and was shown to be *nonlocal*. Through this formalism, Derrida *et al* were able to determine optimal density profiles given certain constraints (e.g. constant mean density of particles) and that fluctuations in particle number are suppressed compared to an equilibrium system. Additionally it is shown that when the dynamics satisfy detailed balance, $\mathcal{F}[\rho]$ reduces to the free energy functional known from equilibrium statistical physics and so the nonlocal functional would seem to generalise a well-known equilibrium concept.

4.9 Outlook

It is perhaps fair to say that exclusion processes are the Ising models of nonequilibrium steady states in the sense that they serve as excellent prototypes of a wide range of phenomena. The new results in this chapter for the partially asymmetric exclusion process (summarised in section 4.7) illustrate the existence of boundary-induced phase transitions whilst the work reviewed in the previous section reveal the presence of condensation transitions and spontaneous symmetry breaking in certain situations.

It is through exact solutions for one-dimensional systems that is has been possible to gain such a clear understanding into the origins of nonequilibrium macroscopic phenomena. Of particular relevance to the main work of this chapter are the long-range effects that can be induced by the boundaries in a nonequilibrium system. These should be contrasted to the case

of equilibrium systems where the effects of boundaries are usually finite in range (i.e. the bulk properties are unaffected).

Finally, I wish to draw particular attention to the papers of Arndt [109] and Derrida *et al* [110] which related properties of the nonequilibrium steady state of particular exclusion processes to well-established concepts in equilibrium physics, namely Yang-Lee theory and the relationship between free energy functionals and large deviations. It would be of interest to try and generalise these results to other exclusion processes to see what, if any, patterns emerge. This might allow one to establish a general characterisation of nonequilibrium steady states and then to make predictions about other processes that have so far eluded exact solution, such as exclusion processes in higher dimensions and thence driven diffusive systems in general.

Chapter 5

Models with an absorbing state

An absorbing state is a configuration which, once accessed by a system, cannot subsequently be left. In this chapter I investigate the implications of the presence of an absorbing state in a model of stochastic dynamics. I begin by reviewing some simple population models which display a transition from a regime in which the absorbing state is reached with certainty (irrespective of the initial condition) to one in which there is some probability of remaining active in perpetuity. This transition is believed to be related to that of a geometrical problem—directed percolation—and I discuss the connection in section 5.1.3. Then, in section 5.2 I introduce a one-dimensional wetting model which does not possess an absorbing state. Using a mean-field approximation and numerical simulation I argue that this model nevertheless exhibits characteristics of the absorbing-state phase transition seen in the simple population models. Finally, I will comment on the possibility of finding an exact solution for models with an absorbing-state phase transition.

5.1 The absorbing state phase transition

5.1.1 Continuous phase transition in a simple population model

A number of properties of models with an absorbing state can be understood from the simple example illustrated in figure 5.1 and which is discussed fully in chapter 1 of the book by Marro and Dickman [66]. The model comprises a population of n_t members at time t. An existing member of the population can reproduce and give birth to a new member at rate r, or it may die at unit rate. It is assumed that only one birth or death event occurs in an infinitesimal time interval $\mathrm{d}t$ in the limit $\mathrm{d}t \to 0$. It is easily seen that in this model the state $n = 0$ is an absorbing state, as the population cannot regenerate itself once it is extinct.

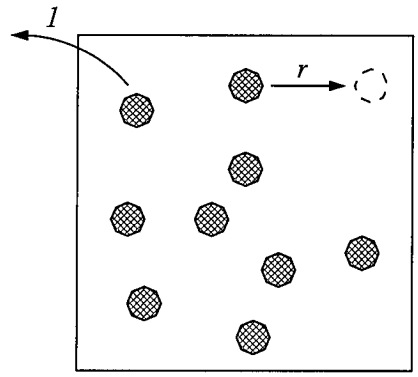

Figure 5.1: Simple model of population dynamics. Each of the n members of the population can reproduce at rate r or die at unit rate. Once all members have died, the population remains permanently extinct: this is an absorbing state.

This extinction is guaranteed to occur if there is a finite upper bound N on the population size, although it may require a rare fluctuation for the absorbing state to be entered. On the other hand, if the maximum population size is not bounded, the population can grow indefinitely. This can be seen by first writing down a master equation for the probability $P_t(n)$ that the population comprises n members at time t:

$$\frac{\mathrm{d}}{\mathrm{d}t}P_t(n) = (n-1)rP_t(n-1) + (n+1)P_t(n+1) - n(1+r)P_t(n)\,. \tag{5.1.1}$$

Then the mean population size $\langle n_t\rangle$ at time t can be shown (see section 2.5.2) to obey the exact differential equation

$$\frac{\mathrm{d}}{\mathrm{d}t}\langle n_t\rangle = (r-1)\langle n_t\rangle\,. \tag{5.1.2}$$

Thus $\langle n_t\rangle = n_0\exp[(r-1)t]$ where n_0 is the population size at $t=0$. Thus for $r>1$ the population grows exponentially on average which implies that the probability for the system to remain permanently active (i.e. $n>0\,\forall\, t$) is nonzero. Conversely for $r<1$, the population decays exponentially, suggesting that the probability the absorbing state is entered at some time t is unity. The transition point $r=1$ between these two regimes is called the *critical* birth rate r_c, and the analogy to an equilibrium critical point will be made clear below.

A little more is learned by solving (5.1.1) exactly for the probability $\phi_t = 1 - P_t(0)$ that the system is *active* ($n_t > 0$) at time t given an initial population $n_0 = 1$. Using a generating function approach (see [66] for the calculations) it can be shown that

$$\phi_t = \begin{cases} \dfrac{1}{1+t} & r = r_c = 1 \\ \dfrac{r-1}{r-\exp[(1-r)t]} & r \neq r_c \end{cases} \tag{5.1.3}$$

The limiting behaviour of ϕ_t as $t\to\infty$

$$\phi = \lim_{t\to\infty}\phi_t = \begin{cases} 0 & r \le r_c \\ 1-\frac{1}{r} & r > r_c \end{cases} \tag{5.1.4}$$

reveals that for $r > r_c$ the system supports two different types of infinite time behaviour. With a probability ϕ the process remains active, whereas with probability $1 - \phi$ the absorbing state is entered at some time. Below r_c there is a single steady state which is the absorbing state, as evidenced by the fact that the probability the system remains active is zero. Hence the transition at $r = r_c = 1$ is referred to as an *absorbing-state* phase transition[1].

The quantity ϕ used to discriminate between the two phases (it is zero below the critical point r_c and nonzero above it) is a suitable *order parameter* for the model. The variation of ϕ with r through the critical point r_c is continuous, and for r just above r_c, one can easily show from (5.1.4) that $\phi \sim |r - r_c|^\beta$ where the critical exponent $\beta = 1$.

As for an equilibrium system, the critical point r_c is accompanied by a divergence. In this case the probability ϕ_t decays as a power-law with time at $r = r_c$. Near criticality, one can introduce a timescale ξ_t through

$$\phi_t \sim \phi + A \exp(-t/\xi_t) \tag{5.1.5}$$

in which A is some amplitude. From equations (5.1.3) and (5.1.4) one finds that $\xi_t \sim |r - r_c|^{-\nu}$ with $\nu = 1$ for $r \simeq r_c$. The exponent ν is then a further critical exponent associated with the transition.

Even in this very simple model, we see evidence for a continuous phase transition characterised by a scalar order parameter and power-law behaviour at and near the critical point. This transition is generic in models with an absorbing state as I now discuss through a more refined population growth model called the contact process.

5.1.2 The contact process

The (basic) contact process was introduced in 1974 as a crude description of an epidemic [111] and here I review its properties.

The model comprises a set of 'organisms' (point particles) which may be either healthy or infected. An infected organism may pass the infection on to a neighbour, or it may overcome the infection and return to good health. The recovery rate is set to unity and the infection rate to r/z in which z is the number of neighbours belonging to each organism. Figure 5.2 illustrates the process on the one-dimensional lattice in which occupied sites denote infected organisms. As with the partially asymmetric exclusion process, we understand the updating scheme of this model to be random sequential (see section 4.2.1).

[1] In the mathematical literature, this transition is also referred to as an *ergodicity* transition, the reason being that if one defines ergodicity as the property of a model exhibiting only a single steady state (as was done in chapter 2 of this thesis) then below r_c the model is ergodic whereas above r_c it is not. However, physicists also associate ergodicity with the exploration of the entirety of configuration space in the steady state, which manifestly does not occur when the system is trapped in the absorbing state. Hence, I avoid the use of the term ergodicity in the rest of this chapter.

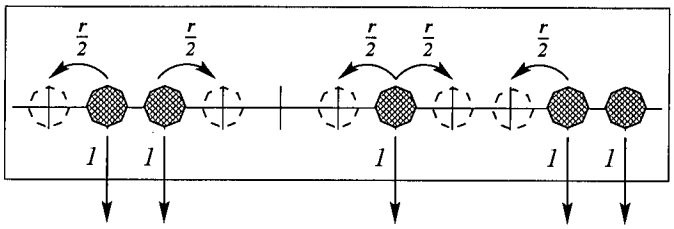

Figure 5.2: Contact process on the one-dimensional lattice. Infected (occupied) sites can infect their nearest neighbours at a rate $r/2$. The infection clears up at unit rate.

Of course, the contact process is a generic spreading process and one need not think in terms of an infection. In fact, to simplify terminology, I shall talk instead of a wetting process: the occupied sites of figure 5.2 are wet and the empty sites dry. The absorbing state of the process is then the completely dry lattice—wet patches cannot spontaneously appear.

It is easily intuited that the contact process could remain active (i.e. nonzero probability of wet sites) in the infinite time limit on the infinite lattice. To this end, consider figure 5.3 in which space-time plots for the one-dimensional contact process with two different values of r are presented. For the smaller wetting rate r the wet (dark) sites are localised at late time, and are likely to die out completely; for the larger wetting rate there are many strands of wet sites persisting and in the limit of an infinite system we guess that the activity could persist forever—i.e. for sufficiently large r we expect an epidemic.

To date the contact process has not been solved exactly, and this remains an outstanding problem—a point to which I return in section 5.4. In the literature, it has been rigorously shown that the absorbing-state transition exists [112] and through various approximation schemes, such as series expansions (see [113] for a review), the critical wetting rate has been estimated as $r_c = 3.329785(9)$.

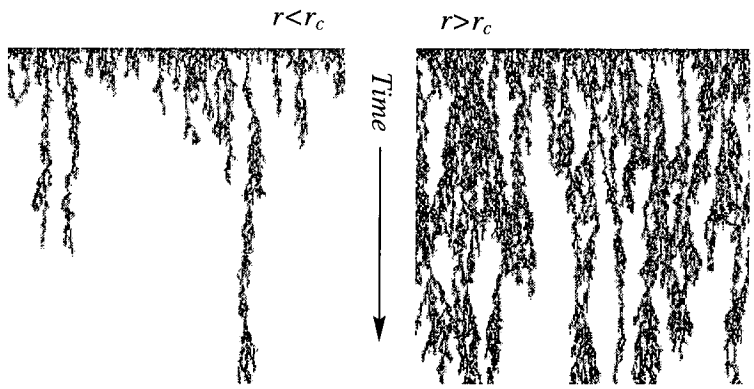

Figure 5.3: Space-time plots for the one-dimensional contact process with r just below r_c (left) and just above (right). The same initial condition (completely wet lattice) and sequence of random numbers sequence was used for both. This implies that pattern of wet sites on the left-hand plot is a subset of that on the right.

As with the simple population model, the critical point is characterised through an order parameter which is defined as the ensemble-averaged density of wet sites in the infinite system as time goes to infinity (given a suitable initial condition, such as a single wet site or a completely wet lattice). That is

$$\phi = \lim_{t\to\infty} \lim_{N\to\infty} \frac{1}{N} \sum_{j=1}^{N} \langle \tau_t(j) \rangle \tag{5.1.6}$$

where $\tau_t(j)$ is the occupation number of site j at time t and N the total number of lattice sites. Note that the order of the limits is significant in this expression: if they were exchanged $\phi \equiv 0$ as the process is guaranteed to become inactive on a *finite* lattice. Near criticality, the order parameter has been found to behave as

$$\phi \sim \begin{cases} 0 & r < r_c \\ |r - r_c|^{\beta} & r > r_c \end{cases} \tag{5.1.7}$$

with $\beta \approx 0.276$ for the one-dimensional contact process (as determined by, e.g., perturbation series expansions [114]).

As in the simple population model, there are divergences at the critical point. These can be seen from, e.g., the equal-time spatial correlation function

$$C_s(x) = \langle \tau_t(0)\tau_t(x) \rangle - \langle \tau_t(0) \rangle^2 \sim \exp(-x/\xi_s) \tag{5.1.8}$$

(in which translational invariance has been assumed) and the steady-state temporal correlation function

$$C_t(u) = \lim_{t_0\to\infty} \langle \tau_{t_0}(0)\tau_{t_0+u}(0) \rangle - \langle \tau_{t_0}(0) \rangle^2 \sim \exp(-u/\xi_t) \; . \tag{5.1.9}$$

The correlation length ξ_s and time ξ_t both exhibit distinct divergences near criticality. That is

$$\xi_s \sim |r - r_c|^{-\nu_s} \quad \text{and} \quad \xi_t \sim |r - r_c|^{-\nu_t} \tag{5.1.10}$$

in which the exponents are $\nu_s \approx 1.10$ and $\nu_t \approx 1.73$ [114].

A celebrated feature of equilibrium critical phenomena is the concept of universality, that is a set of critical exponents is shared by all systems that are characterised by the same type of order parameter and have the same dimensionality and symmetry properties. The values of the critical exponents quoted here for the contact process have also been observed in a number of systems which together form the nonequilibrium *directed percolation* universality class which I now review.

5.1.3 Directed percolation universality class

Consider figure 5.4(i) in which a random porous medium is represented as a two-dimensional square lattice with some of the bonds open (solid lines) and others closed (dotted lines).

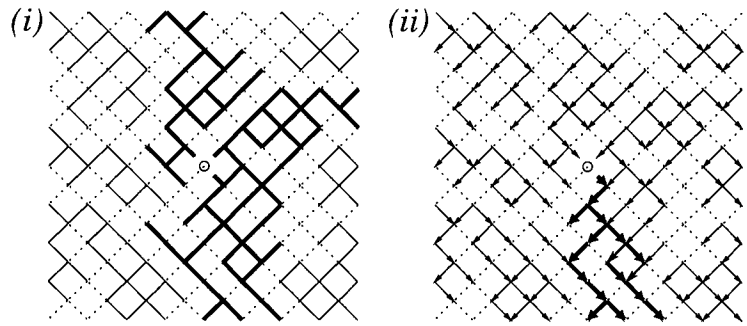

Figure 5.4: (i) Isotropic and (ii) directed percolation through a random medium. In both cases the solid lines are open bonds (each opened independently with some probability p) and the thick lines belong to the cluster that originates from the circled point.

Whether a particular bond is open or closed is statistically independent of the state of any other bond, and a given bond is open with a probability p. A contiguous set of open bonds (such as that indicated with the thick lines) forms a *cluster* and percolation is concerned with the statistics of clusters, such as their mean size, and in particular the situation where a cluster of infinite size exists.

This geometric model is known as isotropic percolation, to which a comprehensive introduction is presented in [115]. Directed percolation (DP) is a similar construction, introduced in 1957 by Broadbent and Hammersley [116]. This time a cluster is a set of bonds which together form a *directed* path with its origin at a specified point. Figure 5.4(ii) has the same set of open bonds as the isotropic model shown in figure 5.4(i), but the cluster originating from the indicated point is only a subset of the isotropic cluster containing that point. The case of directed percolation applies e.g. in the presence of a liquid flowing under gravity through a set of pores (note all bonds point 'downwards' either to the left or right).

One way to study the properties of a directed percolation cluster is to consider the bonds as transferring a fluid from one layer to another. A single layer is then a lattice comprising a sequence of wet and dry sites. The transfer of the fluid can then be specified using the

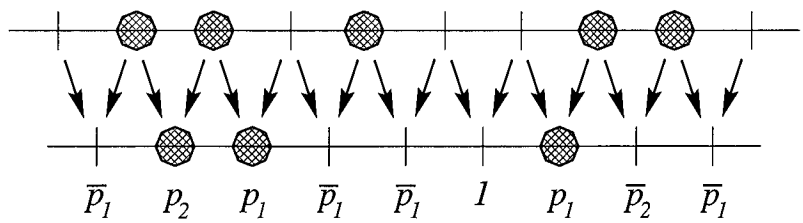

Figure 5.5: The Domany-Kinzel cellular automaton [117] in which directed percolation clusters are constructed layer-by-layer. The labels are probabilities that a site in the next layer is wet or dry given the state of the two sites immediately above (linked by the arrows). The quantities $\bar{p}_i = 1 - p_i$ and the rules are applied in parallel.

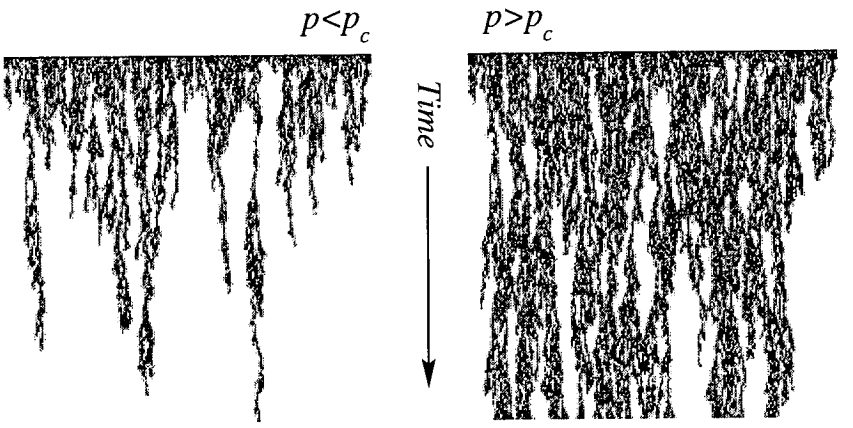

Figure 5.6: The cellular automaton of figure 5.5 with fluid transfer probabilities $p_1 = p$, $p_2 = p(2-p)$ which corresponds to the bond percolation of figure 5.4. The left figure is for a value of p just below the percolation threshold p_c; the right for a value just above. Note the similarities with the contact process, figure 5.3.

rules illustrated in figure 5.5 in which there are two parameters p_1 and p_2 which describe the probability that the fluid flows through a single open bond and two open bonds respectively. This model is the Domany-Kinzel cellular automaton [117] which for different values of p_1 and p_2 describes several variants of directed percolation models. The specific case of directed bond percolation introduced above is realised with $p_1 = p$ and $p_2 = 2p(1-p) + p^2$, where the latter is obtained by considering the probability that at least one of two neighbouring bonds is open in the directed bond percolation model.

Note that in the stochastic Domany-Kinzel cellular automaton of figure 5.5, one layer is constructed from another by applying the elementary rules simultaneously (cf. the random sequential update of the contact process). If one considers each layer as a different timestep, we see from figure 5.6 that the time-evolution of the cellular automaton is similar in character to that of the contact process. In particular there is evidence for a critical value of the bond probability $p = p_c$. Below this *percolation threshold* a layer is completely dry at some late time; above, part of a layer has some probability of being wet even in the limit $t \to \infty$. A completely dry layer is then the absorbing state of directed percolation.

Once again, one studies the percolation transition by defining a suitable order parameter ϕ. Here the density of wet sites in an infinite system as time $t \to \infty$ is appropriate. Also one can consider correlation lengths analogous to those of the contact process. Not only is the same generic behaviour observed in these quantities as for the contact process, but (to within the accuracy of the numerical estimates) the same exponents are obtained as well. I present the values of the 1+1-dimensional[2] DP exponents in table 5.1 as reported in [113].

At this point the standard notation of directed percolation has been adopted. That is, the

[2]The notation d+1 indicates that a dynamic model has d spatial dimensions and one temporal dimension; in the geometric picture of DP it implies a space of d+1 dimensions.

Quantity	Related exponent	Most precise estimate
ϕ	β	0.276486(8)
$\xi_{\parallel}$	$\nu_{\parallel}$	1.096854(4)
$\xi_{\perp}$	$\nu_{\perp}$	1.733847(6)

Table 5.1: Percolation exponents in the directed percolation universality class in 1+1 dimensions taken from [113]. The subscript $\parallel$ relates to the preferred direction of percolation in a DP model, or time in a dynamic model. The subscript $\perp$ relates to the remaining directions in a DP model, or space in a dynamic model. The high precision of the numerical values quoted reflects the amount of effort devoted to the study of directed percolation over the last few years.

symbol $\parallel$ denotes the preferred percolation direction (or time in a dynamic process) and $\perp$ the perpendicular directions (space in a dynamic process). One should note that there are other exponents associated with DP; however, all are related (through scaling and symmetry arguments) to the three given in table 5.1—see [113] for a detailed discussion.

Directed percolation exponents have also been observed in a number of models with an absorbing state, e.g. branching and annihilating random walks with an odd number of offspring[3] [118, 119]. The study of a wide range of systems has led to the suggestion of four sufficient criteria for a model to be in the DP universality class. Collectively known as the directed percolation conjecture [113] these state that if a model

- exhibits a continuous phase transition from a fluctuating active phase to a unique absorbing state;
- is characterised by a positive one-component order parameter;
- evolves under short-range (e.g. nearest-neighbour) interactions; and
- does not possess additional symmetry properties or disorder

then its behaviour near criticality should be the same as that of the directed percolation model with the appropriate dimensionality.

The fact that the character of the order parameter and symmetry properties enter into the criteria is reminiscent of the classification of universality classes in equilibrium statistical physics. It should be noted that this conjecture, whilst not discredited by any of the numerical or approximate analytical evidence, has not been proved. However I say no more of this here, leaving the reader to consult the review by Hinrichsen [113] for a considerably more detailed discussion of directed percolation.

[3] Branching and annihilating random walks with an even number of offspring belong to a different universality class.

5.2 Driven asymmetric contact process (DACP)

I now introduce a new wetting model derived from the basic contact process and in this and the next chapter I shall determine some of its properties.

The model, the driven asymmetric contact process (DACP), is defined as follows. A one-dimensional lattice configuration is specified through a set of occupation numbers $\{\tau(j)\}, 1 \leq j \leq N$ which have $\tau(j) = 1$ if site j is wet and $\tau(j) = 0$ if it is dry. Additionally there is a boundary site $j = 0$ which is permanently wet ($\tau(0) \equiv 1$) and so the process is *driven*. Wetting occurs from left to right *only* and with rate r. As with the basic contact process, wet sites (except the special boundary site at $j = 0$) can dry out independently with unit rate. See figure 5.7 for an illustration.

The motivation for introducing this model is threefold. Firstly, as a result of the boundary wetting process, the model does not possess an absorbing state. Nevertheless, as the wetting rate r is increased, we expect a transition from a phase in which the fluid can spread only a finite distance from the left boundary to one in which the fluid can spread to infinity (i.e. a wetting transition). In this chapter I will show from a mean-field approximation and numerical results that this transition can be observed in the steady-state properties of the model, and present evidence for the critical properties of this transition to belong to the directed percolation class.

The second motivation for studying the model is that, because the fluid travels only in one direction, statistical properties exhibit no finite-size effects. By this it is meant that if one has the solution for a system size N_1, one also has the solution for the first N_1 sites of a system with $N_2 > N_1$ sites because no information travels from right to left in the model. This observation applies also to numerical estimates of macroscopic quantities and can help to reduce simulation run times.

Finally, before the steady state is reached, a cluster of wet sites grows from the left boundary invading the empty lattice. Above the critical wetting rate, this growth process continues

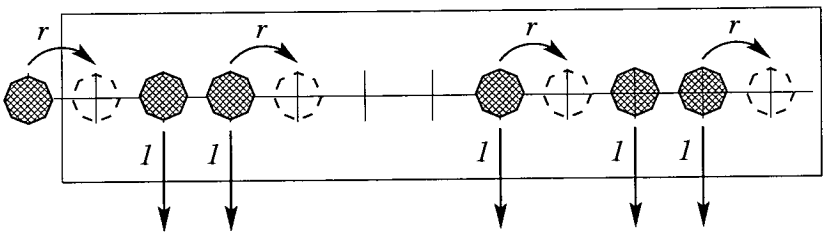

Figure 5.7: Evolution of the driven asymmetric contact process (DACP). Wet (occupied) sites can wet their right neighbours at rate r; sites dry out (become empty) with unit rate. The system is driven since the permanently wet boundary site (at $j = 0$) wets the leftmost lattice site ($j = 1$), also at rate r. This boundary process implies that the model does not have an absorbing state and hence a single unique steady state for all values of r.

indefinitely (at least on on the infinite lattice) and so the model is also a prototype of a propagating wavefront, or shock. Dynamic properties of the model will be discussed in the next chapter: here I concentrate on the steady state.

Before discussing the DACP in more detail, I note that studies of the asymmetric contact process *without* a boundary drive appear to be few in number. Two mathematical papers [120, 121] are concerned with the existence of a critical wetting rate as reflected by a suitably defined order parameter. In these works wetting is allowed to occur both to the right and left with different rates and the initial condition is a single wet site at the origin. Two distinct order parameters are considered: one is the probability that the process remains active indefinitely and the other is the probability that the site at the origin is wet as $t \to \infty$. Clearly for the case of total asymmetry (e.g. only rightward wetting is allowed) the second order parameter is zero for all wetting rates as once the initial seed has dried out, it cannot subsequently become wet. Hence, the second critical point does not exist when the asymmetry is total and it is believed that this is also the case for sufficiently large asymmetry. The first critical point exists for all degrees of asymmetry and when the wetting process is symmetric, the two critical points coincide [120]. Although a phase diagram is presented in [120], no estimates of the values of the critical wetting rates or exponents are given. A numerical and approximate analytical study of the totally asymmetric contact process without a drive [122] confirms the existence of a continuous transition (in terms of the probability that the process remains active) at a value of $r_c = 3.306(4)$ and with an exponent $\beta = 0.2760(1)$ which is close to the best estimate for DP. In the following I obtain similar results for the totally asymmetric contact process with a boundary drive.

5.3 Steady state of the driven asymmetric contact process

In this section I study the steady state of the DACP using a mean-field approximation and a numerical simulation.

5.3.1 Mean field approximation

By following the procedure described in section 2.5.2 one can derived from the master equation for the DACP an exact expression for the mean occupancy $\langle \tau_t(j) \rangle$ of site j at time t (see section 2.5.2). This derived equation reads

$$\frac{\mathrm{d}}{\mathrm{d}t}\langle \tau_t(j) \rangle = r\langle \tau_t(j-1)[1-\tau_t(j)] \rangle - \langle \tau_t(j) \rangle \quad \text{for} \quad j \geq 1 \tag{5.3.1}$$

with the boundary condition $\tau_t(0) = 1 \, \forall t$. Note how the time-dependence of the one-point density $\langle \tau_t(j) \rangle$ depends on the two-point density correlation function $\langle \tau_t(j-1)\tau_t(j) \rangle$. The

time-dependence of this two-point function depends in turn on three-point functions, and so on, and so the set of equations obtained in this way is not closed.

Instead an approximate (mean-field) solution to (5.3.1) can be found by assuming $\langle \tau_t(j-1)\tau_t(j)\rangle = \langle \tau_t(j-1)\rangle\langle \tau_t(j)\rangle$. Introducing the steady-state densities $\rho(j) = \lim_{t\to\infty}\langle \tau_t(j)\rangle$ one finds

$$\rho(j) = r\rho(j-1)[1-\rho(j)] \,. \tag{5.3.2}$$

With the boundary condition $\rho(0) = 1$, the mean-field solution for the density profile is

$$\frac{1}{\rho(j)} = \sum_{k=0}^{j} \frac{1}{r^k} \,. \tag{5.3.3}$$

Taking now the limit $j \to \infty$ we find

$$\phi \equiv \lim_{j\to\infty} \rho(j) = \begin{cases} 0 & r \le 1 \\ 1 - \frac{1}{r} & r > 1 \end{cases} \tag{5.3.4}$$

which is the same result as that obtained for the simple population model of section 5.1.1. As in that model, the DACP exhibits divergences at the critical point $r = r_c = 1$. For example, the steady-state density profile $\rho(j)$ exhibits a power-law decay from the left boundary

$$\rho(j) = \frac{1}{1+j} \tag{5.3.5}$$

which for large j behaves as $\rho(j) \sim j^{-\delta}$ with an exponent $\delta = 1$ whose interpretation will be left open until section 5.3.3 below. Furthermore, with $r \neq r_c$ we find

$$\rho(j) = \phi + |1-r|\exp(-j|\ln r|) \sim \exp(-j/\xi) \tag{5.3.6}$$

in which the lengthscale $\xi \approx |r - r_c|^{-\nu}$ near criticality with the mean-field critical exponent $\nu = 1$.

We see therefore—for the mean-field model at least—that the steady-state density profile exhibits a continuous transition from a phase in which the mean density at infinity ϕ is zero to one in which it is nonzero. The quantity ϕ is hence adopted as an order parameter for the model. However, one cannot tell from the mean-field solution whether this transition is a DP transition as although mean-field predictions are generally qualitatively correct, they usually fail quantitively. In the absence of an exact analytic approach to the model, I turn to numerical methods to estimate the critical exponents of the DACP.

5.3.2 Numerical study

The numerical method that I used to estimate the steady-state properties of the DACP was Monte Carlo simulation. This procedure amounts to little more than storing the set of occupation numbers $\{\tau(j)\}$ and updating them randomly according to the transition rates in the model definition.

Specifically, a Monte Carlo move comprises the random selection of a neighbouring pair of lattice sites $(j, j+1)$. Then, the occupation variables $\tau(j), \tau(j+1)$ are either updated with the probability shown in the table below or otherwise left unchanged.

$\tau(j)\ \tau(j+1)$		$\tau'(j)\ \tau'(j+1)$	Probability
1 0	$\to$	1 1	r/Z
1 1	$\to$	1 0	$1/Z$
0 1	$\to$	0 0	$1/Z$

The factor Z is present so that the rates (r and 1) are transformed to probabilities and is chosen so that the larger of the two rates corresponds to certain update if an appropriate pair of lattice sites is chosen, i.e. $Z = \max\{1, r\}$. This determinism in the updating scheme does not affect the statistics of the sequence of configurations generated as there is noise present in the choice of lattice-site pairs chosen for update.

Note also that, on average, NZ Monte Carlo moves should be performed per unit of 'real' time. That is $\Delta t = 1/(NZ)$ is considered to be the 'infinitesimal' time interval in which a single elementary event is performed—see equation (2.2.1) that relates a continuous-time to a discrete-time process. In fact, technically one should advance the real-time 'clock' by a random number drawn from an exponential distribution with mean Δt. For sufficiently large N, however, advancing the clock linearly is a good approximation to this. Furthermore, it reduces the number of random numbers (which can be time-consuming to compute) required during the course of a simulation. Of course, if one is only interested in the steady state, these considerations are rather academic.

In order for the critical point to be located, and the critical exponents determined, it is necessary to sample the density profile $\rho(j)$ in the steady state. This is done by initialising the array of occupancy variables so that only the left boundary site is wet. Then the above Monte Carlo updates (which are constructed so that the left boundary site is never changed) are iterated until the steady state is reached. In simulation terms, this is where lattice configurations are generated with a frequency proportional to their steady-state probability, at least in the infinite-time limit (see section 2.1.4). An estimate of the density profile $\rho(j)$ is then obtained by averaging $\tau(j)$ over a sufficiently large sequence of lattice configurations—although see below for what is deemed 'sufficiently large' from a practical point of view.

Once the density profile is obtained, the critical wetting rate r_c is estimated as that for which $\rho(j)$ is closest to a power-law. Now, near criticality we expect the scaling form

$$\rho(j) \sim j^{-\delta} f(j/\xi) \tag{5.3.7}$$

in which δ is a critical exponent, $f(u)$ is a scaling function and ξ is the correlation length. At criticality, the correlation length should diverge as $\xi \sim |r - r_c|^{-\nu}$, and so $f(u) \to f(0) = \text{const.}$

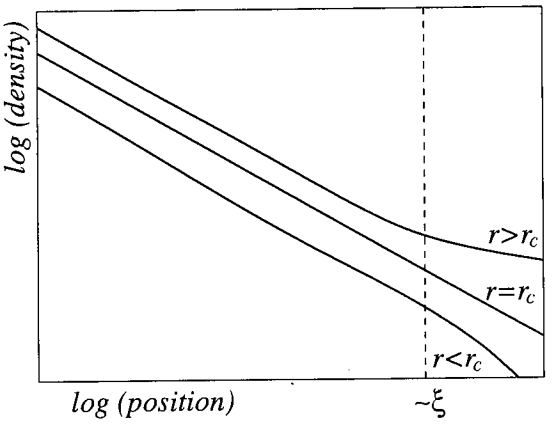

Figure 5.8: Schematic log-log plot of the density profiles in the DACP for wetting rates below, at and above criticality. The dashed line indicates the distance from the left boundary at which one can begin to distinguish the three profiles. This distance is on the order of the correlation length ξ of the model, and thus diverges near criticality.

What this means in practice is that in order to distinguish between density profiles at different r, the system size N must be larger than the correlation length ξ. Figure 5.8 schematically indicates the expected density profiles near criticality, and how they are indistinguishable from straight lines on a log-log plot if the system size is too small. Hence as the critical point is approached, the divergence in ξ implies that one must use ever larger systems in order for the density profiles to be distinguishable.

The fact that larger system sizes must be used near criticality further implies that the simulation run time must be increased considerably. This is because near criticality, the time to reach the steady state increases as a power of the system size: this is a consequence of at least the second largest eigenvalue of the transition rate matrix W approaching zero as $N \to \infty$ (see section 2.4). Furthermore, the correlation time also diverges near criticality. The means that the length of the sequence of generated configurations over which the density profile is sampled must also be increased: in order for a time average to be a reliable estimate of an ensemble average (the quantity of interest), the sample time must span many correlation times.

Once the density profiles have been sampled for sufficiently large system sizes and run times, one estimates the location of the critical wetting rate r_c by invoking a procedure, the method of local slopes [113], that magnifies the differences in the profiles at large distances from the left-boundary evident from figure 5.8. In particular, one seeks to distinguish between the supercritical case ($r > r_c$) where the density profile shows an upturn at large distances and the subcritical case ($r < r_c$) where the profile begins to decay.

The way in which this is done is through the introduction of a *local slope* $\delta(j)$. It is defined as

$$\delta(j) = -\frac{\ln \rho(j) - \ln \rho(j/b)}{\ln b} \tag{5.3.8}$$

where b is the fractional separation between neighbouring measurement points of $\rho(j)$. Then, one plots the graph of $\delta(j)$ against $1/j$ for different values of r. If $r < r_c$, $\delta(j)$ increases as $j \to \infty$ because $\rho(j) \to 0$ in that limit. Conversely, when $r > r_c$, $\delta(j)$ decreases. If $\delta(j)$ remains constant as $j \to \infty$, the corresponding value of r is taken as an estimate the critical wetting rate r_c and the limiting value of $\delta(j)$ as an estimate of the critical exponent δ.

In the analysis of the density profile of the DACP it was found that the integrated quantity

$$m(j) = \frac{1}{j}\sum_{k=0}^{j}\rho(k) \tag{5.3.9}$$

gave more reliable results when used in place of $\rho(j)$ in (5.3.8). This is because $m(j)$ is less susceptible to large fluctuations than $\rho(j)$ in the numerics. Note that in the region of interest (large j) $m(j) \to \rho(j)$ and so the replacement of the density $\rho(j)$ with the integrated quantity $m(j)$ does not affect the analysis of the critical point using the method of local slopes (5.3.8). The local slopes $\delta(j)$ obtained through this analysis are shown in figure 5.9. Using the statements of the previous paragraph, the estimates $r_c \approx 3.305$ and $\delta = 0.163$ are obtained.

These initial estimates can be improved upon, and we also learn something of the correlation length ξ, by using the scaling relation (5.3.7). Specifically, one plots $\rho(j)j^{\delta}$ against $u = j|r-r_c|^{\nu}$ with the initial estimates of δ and r_c and vary ν until all the data lie along two curves (one for $r > r_c$ and one for $r < r_c$). These curves then give the scaling function $f(u)$ in equation (5.3.7). Figure 5.10 shows the result of this procedure. The values of the critical wetting rate and exponents that gave the best fit are

$$r_c = 3.3055(5)\ , \quad \delta = 0.1640(5) \quad \text{and} \quad \nu = 1.7(2) \tag{5.3.10}$$

where the figure in brackets indicates the change in the last quoted digit before the collapse becomes noticeably worse. The estimate of the critical wetting rate r_c is in agreement with that obtained in a similar study of the asymmetric contact process without a boundary drive [122].

5.3.3 Directed percolation exponents in the DACP

I now explain how the directed percolation exponents appear to be manifested in the steady state density profile of the DACP.

First I review the scaling argument that implies that the density decay exponent δ is equal to β/ν where β is the exponent associated with the behaviour of the order parameter near criticality. Recall the scaling form (5.3.7) which read

$$\rho(j) = j^{-\delta} f(j/\xi) \tag{5.3.11}$$

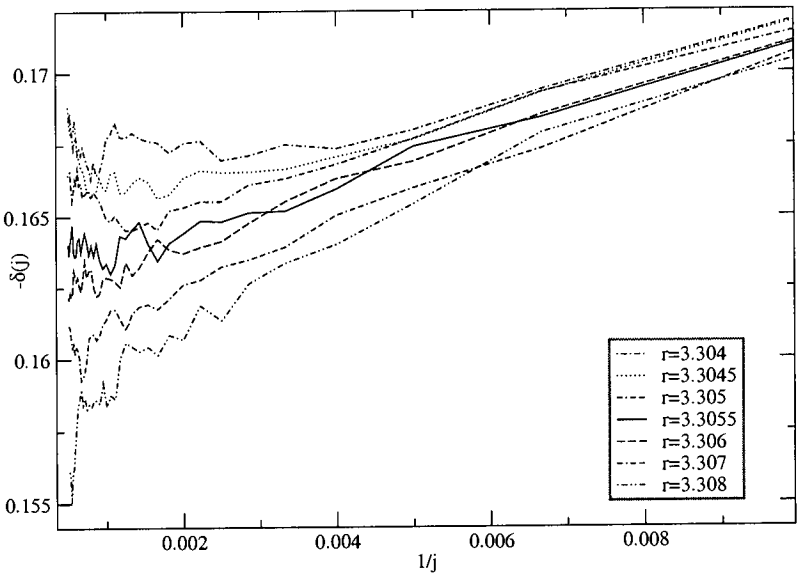

Figure 5.9: Local slopes of the integrated density $m(j)$ at wetting rates r around criticality. The plot corresponding to $r = 3.3055$ approaches the vertical axis nearly horizontally, and so one takes that as the estimate of the critical wetting rate r_c.

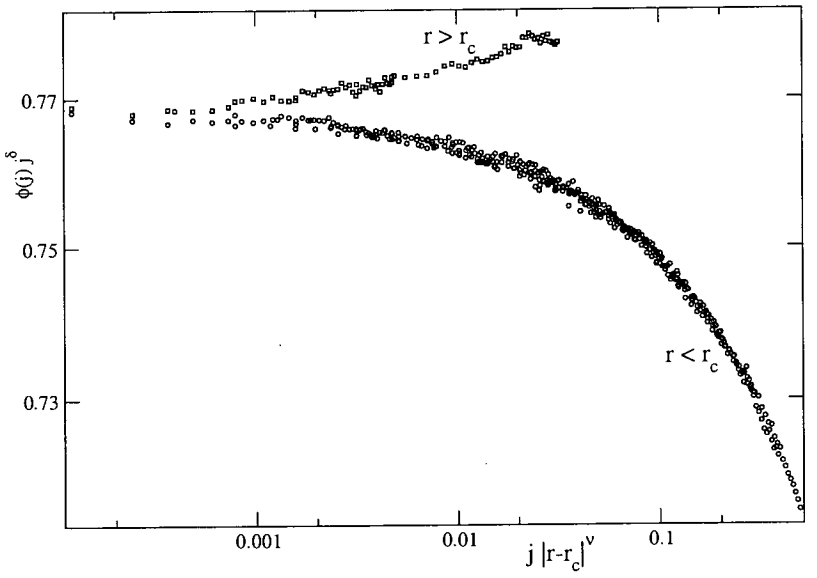

Figure 5.10: Density rescaled according to relation (5.3.7) with $r_c = 3.3055$, $\delta = 0.1640$ and $\nu = 1.7$.

in which $f(u)$ is a scaling function. Now above criticality, $\rho(j)$ approaches a nonzero constant as $j \to \infty$. This implies that $f(u) \sim u^{\delta}$ as $u \to \infty$ so that the j-dependence drops out of (5.3.11). Then

$$\lim_{j\to\infty} \rho(j) \sim \xi^{-\delta} \sim |r - r_c|^{\delta\nu} \tag{5.3.12}$$

because $\xi \sim |r - r_c|^{-\nu}$ near criticality. However, the limit on the left-hand side of the previous equation is none other than the definition of the order parameter ϕ which behaves as

$$\phi = \lim_{j\to\infty} \rho(j) \sim |r - r_c|^{\beta} . \tag{5.3.13}$$

Comparison of the previous two equations implies that $\delta = \beta/\nu$ as claimed.

To compare the numerical results for the critical exponents of the DACP with those of directed percolation, recall the best estimates of the latter from table 5.1. These are

$$\delta = \frac{\beta}{\nu_{||}} = 0.159464(6) , \quad \nu_{\perp} = 1.096854(4) \quad \text{and} \quad \nu_{||} = 1.733847(6) \tag{5.3.14}$$

from which we see that the estimate of $\delta = 0.1640(5)$ for the DACP is in reasonable agreement with that of DP, as is that for the DACP exponent $\nu = 1.7(2)$ if we identify it with the DP temporal correlation exponent $\nu_{||}$.

To understand why the DP spatial correlation length $\xi_{\perp}$ is not observed in the steady-state density profile of the DACP consider figure 5.11 in which a typical space-time plot for the process is presented. On the plot two lengthscales are shown: the longer length indicates the typical lifetime of a cluster which is related to the DP length $\xi_{||}$; the shorter length is the typical width of a cluster which is related to the DP length $\xi_{\perp}$. The density decay length in the DACP can be seen to be a combination of the two DP lengths, i.e. $\xi \approx \xi_{||} \cos\theta + \xi_{\perp} \sin\theta$ in which θ is the angle indicated on figure 5.11.

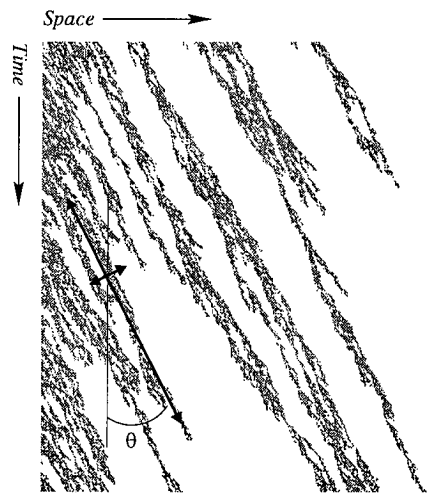

Figure 5.11: A space-time plot of the one-dimensional model just below criticality. The shorter of the two lengths indicated is the DP spatial scaling length; the longer length is the DP temporal scaling length.

Now, as $r \to r_c$, the two lengthscales $\xi_{||}$ and $\xi_\perp$ diverge with different exponents, and so as one takes r closer to r_c, the contribution to ξ from the DP length that diverges less rapidly will become less important. That is $\lim_{r \to r_c} \xi \sim \xi_{||}$ as $\nu_{||} > \nu_\perp$. Hence, the lengthscale associated with the steady-state density decay in the DACP is identified with the directed percolation temporal scaling length $\xi_{||}$. A similar result applies for a directed percolation model with an active (wet) boundary [123].

5.4 The question of the solvability of models with an absorbing state

Through the above mean-field and numerical study of the driven asymmetric contact process I have provided evidence that it possesses a phase transition in the directed percolation universality class. I now return to the observation that no model with DP exponents has yet been solved exactly. In particular I ask whether the special features of the DACP noted at the start of section 5.2, namely that it has a single, unique steady state and does not exhibit any finite-size effects, may make it better poised for exact solution than, for example, the basic contact process.

First I review evidence that has led to the suggestion that the contact process and directed bond percolation models are not exactly solvable. One argument [113] is that the best estimates of the DP exponents are not close to 'reasonable' rational numbers. Stronger evidence comes from exact solutions of the contact process and DP for small system sizes.

In figure 5.12 the real parts of the eigenvalues of the transition rate matrix for the basic contact

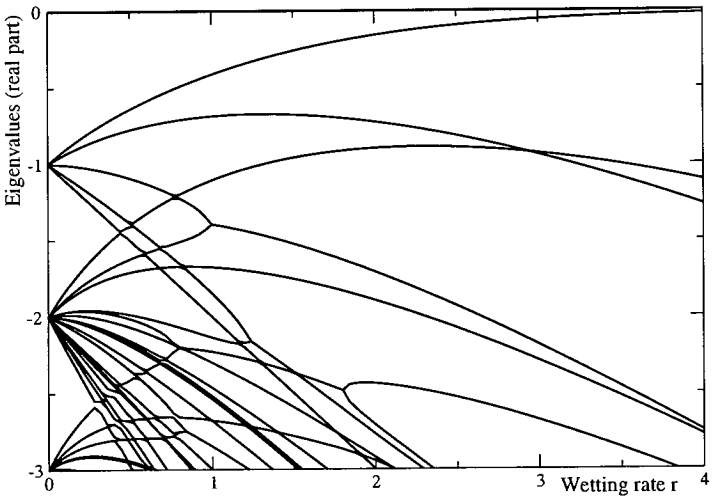

Figure 5.12: Real part of the eigenvalues of the transition rate matrix for the basic contact process with $N = 8$ sites and periodic boundary conditions.

process with $N = 8$ and periodic boundary conditions are shown. Recall from section 2.4 that we expect (at least some) eigenvalues to approach zero at $r = r_c$, indicating the onset of power-law behaviour and two steady states. In the figure the eigenvalue with the second-largest real part approaches zero as $r \to \infty$: this simply reflects the fact that in this limit, the process remains active forever as no drying out of the lattice can occur and so this eigenvalue is not related to the DP transition. The most significant feature of figure 5.12 is the 'spiders-web' of eigenvalue crossings; indeed the presence of bifurcations is reminiscent of the logistic mapping ($x \to ax(1-x)$) well-known in chaos theory. It is this latter observation which further serves as particular reason to doubt the solvability of the contact process.

Meanwhile, a recent study [124] of directed bond percolation (defined in section 5.1.3) at small system sizes has yielded another suggestion of a relationship to chaos. In this study, the probability that a percolating cluster is present in a system comprising N layers was calculated explicitly for $N = 15$. This probability is a polynomial in the single-bond percolation probability p and, in the spirit of the Yang-Lee theory of phase transitions (q.v. section 2.4.1), the polynomial was solved for its zeros in the complex-p plane. The authors of [124] noted a distribution of zeros suggestive of a fractal which they believe to be a signature of nonsolvability. For solvable models, partition function zeros have generally been found to lie along well-defined curves. For example, in section 4.8.4 it was noted that the partition function zeros for an exclusion process lie along hyperbolæ or ellipses depending on a particular control parameter.

I now investigate the structure of eigenvalues of the transition rate matrix as a function of r and zeros of the steady-state normalisation (see section 2.3.2 for a definition) for the driven asymmetric contact process. First recall from the discussion at the start of section 5.2 that the addition of extra sites to an N-site lattice does not affect the statistics of the first N sites

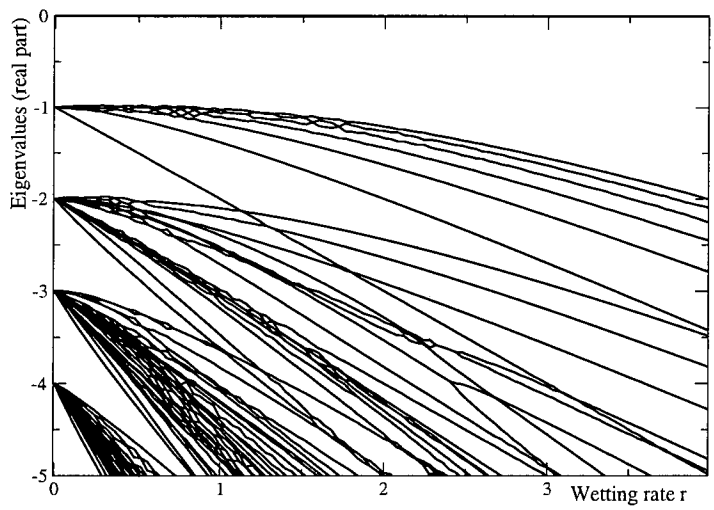

Figure 5.13: Real part of the eigenvalues of the transition rate matrix for the DACP with $N = 7$ sites.

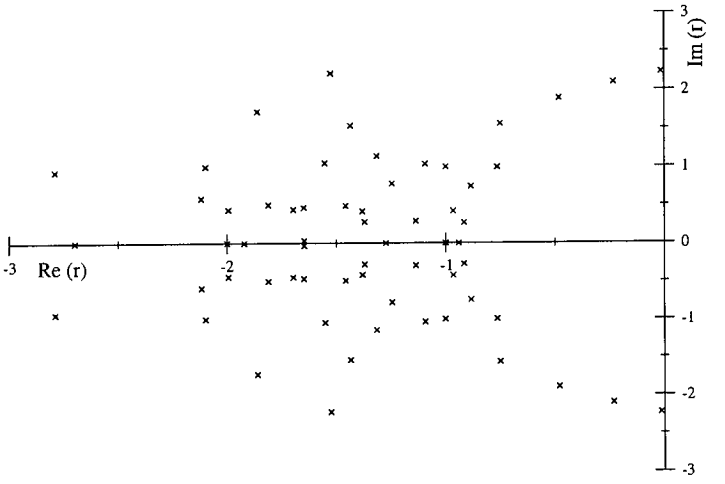

Figure 5.14: Zeros of the steady-state normalisation (partition function) in the complex r plane for the DACP with six sites.

of the system. This also implies that the eigenvalues and zeros of the normalisation for the N-site system are left unchanged as one increases the system size.

Now consider figure 5.13 which shows the real parts of the eigenvalues of the transition rate matrix for the DACP with system size $N = 7$. Note that, in contrast to figure 5.12 for the basic contact process, it appears that the eigenvalues' real parts decrease largely monotonically with wetting rate r. That is, there is no evidence of the continuous phase transition, which implies at least one eigenvalue converging to zero at $r \approx r_c$. Presumably this enters in only at larger N. Note also that there seem to be fewer bifurcations than in figure 5.12; however it is difficult to use this to decide conclusively whether the DACP should be solvable or not.

The zeros of the steady-state normalisation (which is related to the eigenvalue spectrum through equation (2.3.11)) are plotted in the complex plane for $N = 6$ in figure 5.14. Here it is again difficult to decide whether the distribution of zeros is regular or chaotic. On the one hand, some zeros seem to be clustered around the negative real axis in an irregular fashion whereas at the same time there are the beginnings of at least one ellipse curving round towards the real axis.

Again I would argue that this preliminary study of the normalisation zeros is insufficient to be able to make any strong predictions as to the solvability of the DACP. However, to conclude this section I note the suggestion [125] that directed percolation falls into a class of problems which are solvable in the sense that their solutions are expressible in terms of a class of functions ("D-finite functions") which are themselves the solutions of linear ordinary differential equations of finite order.

5.5 Summary of results

In this chapter I have investigated a new variant of the contact process, referred to as the driven asymmetric contact process (DACP), which has asymmetric wetting and a boundary drive. Through a mean-field approximation I demonstrated the existence of a continuous phase transition analogous to that present in simple population models with an absorbing state. Then, through a numerical study, I provided evidence that the critical exponents associated with the phase transition are consistent with those of the directed percolation (DP) universality class, despite the lack of an absorbing state in the model.

It was noted that one of the motivating factors for introducing the DACP is that it has two attractive features from a theoretical point of view. Firstly, the fact that statistical properties of a portion of the system are not affected by altering the overall lattice size should, in principle, make it easier to understand larger systems given the solution for a smaller system. Secondly, the absence of an absorbing state implies that there is a single, unique steady-state which may be beneficial in terms of finding an analytic solution for the model.

In the previous section of this chapter I discussed the possibility of finding an exact solution for a model with DP exponents. Whilst other investigations of the contact process and directed bond percolation have indicated hallmarks of chaos, my preliminary study of the DACP has not revealed that it possesses similar properties. However, the system sizes I was able to access are too small to be able to make any strong claims as to the solvability of the driven asymmetric contact process. Hence the solvability of models in the DP universality class remains, for the time being at least, an important open question.

Chapter 6

Dynamics of wetting

In the previous chapter I introduced a wetting model and studied its static (steady-state) properties. I now turn to its dynamic properties. Specifically, I consider the invasion of the dense, wet bulk into the initially dry lattice, concentrating particularly on the dynamical scaling behaviour of the interface between the wet and dry zones. For the purposes of these investigations, it is useful to introduce a two-dimensional generalisation of the one-dimensional driven asymmetric contact process. I call this new model the wetting model and begin with its definition.

6.1 The wetting model

The wetting model is defined on a two-dimensional lattice, periodic with length L in the x-direction and semi-infinite in the z-direction. Each plane parallel to the z-axis behaves similarly to the driven asymmetric contact process of the previous chapter: wet sites can cause neighbouring sites in the $+z$ direction to become wet with rate r; wet sites can dry out with rate k. The spatial distribution of wet sites can become correlated in the x-direction by allowing wet sites to hop (diffuse) sideways, with rate $D/2$ both to the left and right. The wetting is driven from the lower boundary, i.e. the plane at $z = 0$ is held wet at all times. Figure 6.1 illustrates these dynamics and we note that with $k = 1, L = 1$ the driven asymmetric contact process of chapter 5 is recovered.

In this chapter I wish to understand the way in which the dense wet structure grows from the lower boundary. As before, something of the model is learned from a mean-field approximation. Before I perform these calculations, I will show how the wetting model in the limit $D \to \infty, L \to \infty$ is equivalent to the one-dimensional driven asymmetric contact process in the mean-field.

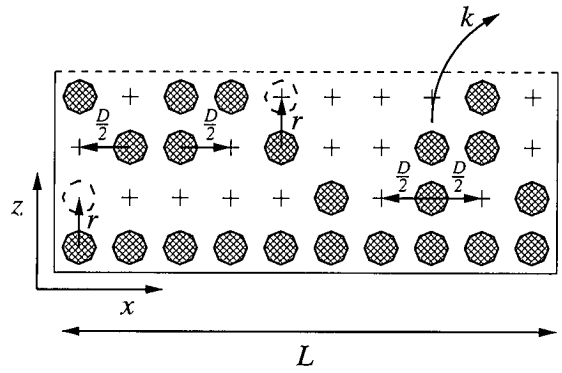

Figure 6.1: Two-dimensional wetting model, derived from the driven asymmetric contact process. Wetting occurs in the $+z$ direction with rate r, sideways hopping (diffusion) at rate $D/2$ and drying out at rate k. The lower boundary ($z = 0$) is wet at all times to drive the system.

6.1.1 Equivalence of the wetting model to the mean-field DACP

Consider the limit $D \to \infty$ in the wetting model. Then, between each wetting and drying-out event, an infinite number of sideways diffusion moves can take place and so each plane reaches a steady state before a particle is added or removed. This sideways diffusion with periodic boundaries and constant particle number is a symmetric exclusion process—see chapter 4. Recall that the steady state probability distribution of this process is one in which all configurations are equally likely.

The average fraction of wet sites $\langle \rho_t(z) \rangle$ in plane z at time t can be shown to obey the exact differential equation

$$\frac{\mathrm{d}}{\mathrm{d}t}\langle \rho_t(z) \rangle = r\langle \rho_t(z-1)[1-\rho_t(z)] \rangle - k\langle \rho_t(z) \rangle \quad z \geq 1 \tag{6.1.1}$$

supplemented by the boundary condition $\rho_t(0) \equiv 1$. This form arises because the locations of wet sites *within* plane z are uncorrelated with those in plane $z-1$; hence the rate at which a site in plane z becomes wet is just $\rho(z-1)[1-\rho(z)]$. For this equation to describe the DACP in the mean-field, it must be demonstrated that $\rho(z-1)$ and $\rho(z)$ are uncorrelated.

To do this I first write down the exact differential equation for two-point density correlations (see section 2.5.2)

$$\begin{aligned}\frac{\mathrm{d}}{\mathrm{d}t}\langle \rho_t(z)\rho_t(z') \rangle = &-2k\langle \rho_t(z)\rho_t(z') \rangle + r\langle \rho_t(z)\rho_t(z'-1)[1-\rho_t(z')] \rangle + \\ &r\langle \rho_t(z-1)[1-\rho_t(z)]\rho_t(z') \rangle + \frac{1}{L}\left(\langle \rho_t(z) \rangle + \langle \rho_t(z-1) \rangle\right)\delta(z,z')\end{aligned} \tag{6.1.2}$$

Now, if the quantities $\rho_t(z)$, $\rho_t(z')$ etc. are uncorrelated, the following is true:

$$\frac{\mathrm{d}}{\mathrm{d}t}\langle \rho_t(z)\rho_t(z') \rangle = \left(\frac{\mathrm{d}}{\mathrm{d}t}\langle \rho_t(z) \rangle\right)\langle \rho_t(z') \rangle + \langle \rho_t(z) \rangle\left(\frac{\mathrm{d}}{\mathrm{d}t}\langle \rho_t(z') \rangle\right). \tag{6.1.3}$$

By inserting (6.1.1) into this equation, and assuming no correlations, (6.1.2) is obtained (if the latter is similarly factorised) *except* when $z = z'$. In that case, the two expressions differ by an amount proportional to $1/L$. Hence, in the limit $L \to \infty$ the wetting model with infinite diffusion obeys

$$\frac{\mathrm{d}}{\mathrm{d}t}\rho_t(z) = r\rho_t(z-1)[1-\rho_t(z)] - k\rho_t(z) \tag{6.1.4}$$

in which the angle brackets have been omitted for clarity. This is none other than the mean-field approximation to equation (5.3.1) which defines the one-dimensional driven asymmetric exclusion process.

6.1.2 Travelling wave solutions in the mean-field

As a starting point in our understanding of the dynamics of wetting model, I will consider travelling wave solutions of equation (6.1.4). As we have seen, this equation describes both the mean-field approximation to the one-dimensional driven asymmetric exclusion process and a particular limit ($D \to \infty, L \to \infty$) of the wetting model.

One finds uniformly translating solutions to (6.1.4) by introducing the function $\psi(z-vt)$ which is related to the density of wet sites in plane z through

$$\begin{aligned} \rho_t(z) &= \psi(z-vt) \\ \rho_t(z-1) &= \psi(z-vt) - \psi'(z-vt) + \tfrac{1}{2}\psi''(z-vt) - \dots \,. \end{aligned} \tag{6.1.5}$$

In these two equations, a prime denotes differentiation with respect to the argument $u = z - vt$ of ψ. Furthermore, the use of the Taylor expansion (which has been truncated at second order) implies the replacement of the discrete set of densities $\rho_t(z)$ with the continuous waveform $\psi(u)$.

Using these relations in (6.1.4) one obtains the ordinary differential equation

$$-v\psi'(u = x - vt) = r(1-\psi)\left(\psi - \psi' + \tfrac{1}{2} + \psi''\right) - \psi \,. \tag{6.1.6}$$

As this equation is nonlinear, we do not necessarily know how to solve it. However, we can obtain some information about the shape of the profile $\psi(u)$ by performing a stability analysis in the *phase-plane* spanned by $q_1 = \psi$ and $q_2 = \psi'$ which we take as independent variables (see e.g. [126] for an account of this standard technique).

It is easy to show that q_1 and q_2 satisfy the coupled first-order differential equations

$$\frac{\mathrm{d}q_1}{\mathrm{d}u} = q_2 \quad \text{and} \quad \frac{\mathrm{d}q_2}{\mathrm{d}u} = \frac{2(q_1 - vq_2)}{r(1-q_1)} - 2(q_1 - q_2) \,. \tag{6.1.7}$$

and the idea now is to specify an initial density $\psi \equiv q_1$ and density gradient $\psi' \equiv q_2$ and use (6.1.7) to determine how the solution to (6.1.6) with this initial condition is represented by a

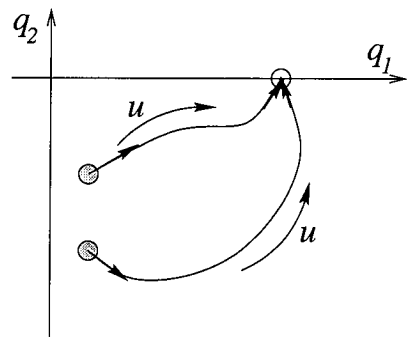

Figure 6.2: Flow in the (q_1, q_2) plane arising from a pair of coupled differential equation such as (6.1.7). The shaded points are an initial condition, and in both cases shown here, the solutions flow towards a fixed point where (q_1, q_2) are constant as the independent variable u is increased.

flow in the phase plane. Figure 6.2 illustrates this for two different initial conditions, shown as shaded circles. As u is increased, the solution of (6.1.6) flows towards, and becomes stationary, at a *fixed point* (q_1^*, q_2^*) which is the same for the two solutions shown. At the fixed points both $\mathrm{d}q_1/\mathrm{d}u$ and $\mathrm{d}q_2/\mathrm{d}u$ are zero and it is the location (and nature) of these fixed points which allow statements to be made about the travelling waves supported by (6.1.4).

It is easily verified that (6.1.7) has two fixed points. One of these has $(q_1^*, q_2^*) = (1 - r^{-1}, 0)$ which corresponds to the homogeneous active phase of the DACP with $r > r_c = 1$ (recall equation (5.3.4) for the steady-state density in the mean field). This is the appropriate choice of q_1 and q_2 at $t = \infty$, or equivalently $u = x - vt = -\infty$. Hence any solution of (6.1.7) beginning at this point should be repelled from this fixed point. This can be determined by linearising (6.1.7) around (q_1^*, q_2^*), i.e. by writing $(q_1, q_2) = (q_1^*, q_2^*) + (\epsilon_1, \epsilon_2)$. Then

$$\frac{\mathrm{d}}{\mathrm{d}u}\begin{bmatrix}\epsilon_1\\ \epsilon_2\end{bmatrix} = \begin{bmatrix}0 & 1\\ 2r(1-\frac{1}{r}) & 2(1-\frac{v}{r})\end{bmatrix}\begin{bmatrix}\epsilon_1\\ \epsilon_2\end{bmatrix} \tag{6.1.8}$$

to first order in ϵ_i. An eigenvector of the matrix in this equation that corresponds to a negative eigenvalue indicates the direction from which a solution (q_1, q_2) is attracted to the fixed point as u is increased. Conversely a positive eigenvalue indicates that a solution is repelled from the fixed point. The above matrix always has one positive and one negative eigenvalue, implying that a solution may initially approach the fixed point but will eventually be repelled away from it. This corresponds to the fact that as $t \to -\infty$ $(u \to \infty)$ we must have $\psi \to 0$ as before the wave passes the system is completely dry.

More is learnt from the fixed point $(q_1^*, q_2^*) = (0, 0)$ which corresponds to the completely dry system. If the solution to (6.1.7) is to exist, this fixed point must be attractive as $u \to \infty$: that is, the eigenvalues of the corresponding matrix must both be negative. In fact, analysis shows that both eigenvalues are

(i) positive if $v < r$ (solutions are repelled from the fixed point);

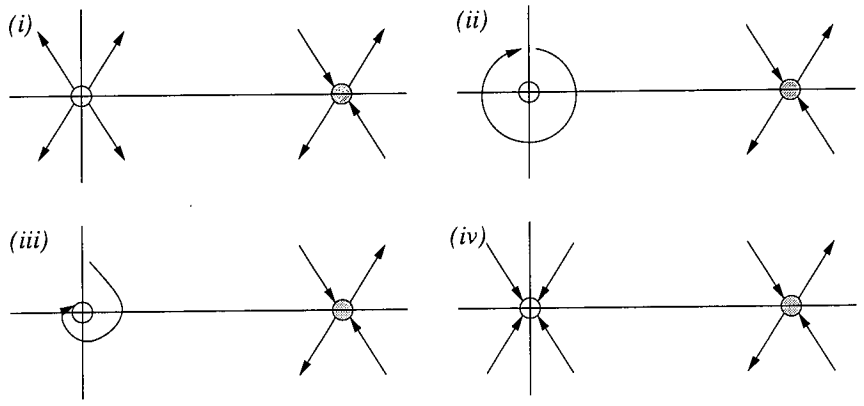

Figure 6.3: Fixed points in the (q_1, q_2) plane for the driven asymmetric exclusion process in the mean field (6.1.7). The shaded circle corresponds to the initial condition $\psi(-\infty) = 1 - 1/r$, $\psi'(-\infty) = 0$. Small fluctuations about this point will invariably cause the solution to be repelled away from the fixed point. For a solution to be physical it must approach the other fixed point (unshaded circle) as $u \to \infty$. In case (i) this is not possible: the fixed point is repulsive. In case (ii) the solution circulates around the fixed point. In case (iii) the solution spirals into the fixed point, yielding unphysical negative densities. In case (iv) the solution may approach the fixed point directly.

(ii) imaginary if $v = r$ (solutions circulate around the fixed point);

(iii) complex with negative real part if $r < v < r + \sqrt{(2r[r-1])}$ (solutions spiral in towards the fixed point);

(iv) negative if $v \geq r + \sqrt{(2r[r-1])}$ (solutions approach the fixed point directly).

These different types of behaviour are illustrated in figure 6.3.

Cases (i) and (ii) are unstable, and so are not suitable travelling wave solutions. Case (iii), although stable, implies negative densities as $u \to \infty$, so is deemed unphysical. Hence, for $\psi(u)$ to be a stable solution of (6.1.6) we must have

$$v \geq v^* = r + \sqrt{(2r[r-1])} \,. \tag{6.1.9}$$

However this result also *seems* unphysical in that the wave velocity exceeds the fundamental wetting rate r. To understand more clearly how this is possible, I simulated the wetting model in the limit of infinite diffusion, which I have shown above (section 6.1.1) to be equivalent to the DACP under a mean-field approximation as $L \to \infty$.

6.1.3 Numerical results

The simulation of the wetting model was performed in a similar way to that described for the DACP described in section 5.3.2. As I now explain, the effects of infinite diffusion are easily incorporated by using the fact that each configuration of wet sites within a particular

horizontal plane of the system is equally likely. Specifically, a Monte Carlo move is performed by first choosing one of the wet sites in the system with equal probability, and obtaining the z-coordinate of this site—this is equivalent to choosing a plane z with a probability proportional to the density of wet sites $\rho(z)$ contained within it. Then one performs an update with a probability as shown below.

Event	Update	Probability
Drying out	$\rho(z) \to \rho(z) - \frac{1}{L}$	$1/Z$
Wetting	$\rho(z+1) \to \rho(z+1) + \frac{1}{L}$	$(r/Z)(1-\rho(z+1))$

In this the normalisation $Z = 1 + r$ (so that neither of the probabilities exceeds one), and the wetting probability is proportional to the probability that a randomly chosen site in plane $z+1$ (and in particular, that with the same x-coordinate as the originally selected particle) is wet. As a microscopic update is initiated by a choice of particle rather than a pair of sites (q.v. the scheme of section 5.3.2), time increases nonlinearly in the simulation. In fact, the real-time 'clock' advances by an amount $1/(MZ)$ where M is the total number of wet sites in all planes before the Monte Carlo move is attempted.

To locate the position of the advancing front one uses the fact that the positions of wet sites within a plane are uncorrelated with those in an adjacent plane. Then, the mean fraction of wet sites in the horizontal plane at z that do not have any other wet sites above them is

$$r(z,t) = \rho_t(z) \prod_{z'=z+1}^{\infty} (1 - \rho_t(z')) \tag{6.1.10}$$

and hence the mean front position is given by

$$\bar{h}(t) = \sum_{z=0}^{\infty} z\, r(z,t)\ . \tag{6.1.11}$$

This quantity was calculated for a sequence of simulation runs at the system size L, and then the mean over these runs was taken to obtain a noise-history (ensemble) averaged estimate of the front position $\langle h(t) \rangle$. It was then a simple matter to estimate the velocity by fitting a quadratic within a given time-window of this function. Figure 6.4(i) shows the front velocity obtained in this way for a variety of system sizes L from 100 to 6400. It can be seen that the time taken to reach a steady velocity does not depend on L and that the asymptotic, steady velocity increases with the system size. Figure 6.4(ii) shows this increase explicitly along with the minimum velocity v^* admitted by the stability analysis of the previous section (dotted horizontal line).

First we note that *all* of the asymptotic velocities shown in figure 6.4 are greater than the wetting rate $r = 3$ used in the simulation. Thus the apparently unphysical result from the

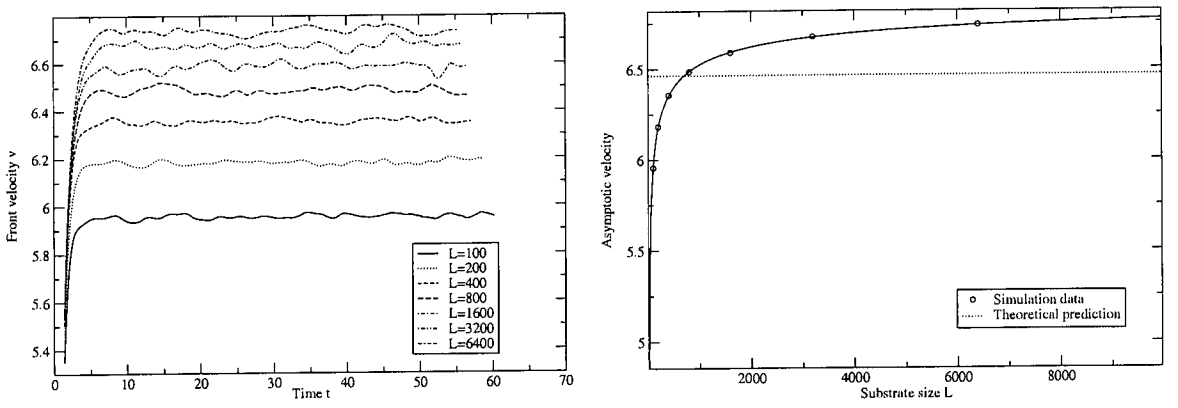

Figure 6.4: (i) Velocities as a function of time in the numerical realisation of the mean-field wetting model with the wetting rate $r = 3$. (ii) Asymptotic (steady) velocities as a function of system size L obtained by fitting to the data of (i). The dotted horizontal line is the prediction from stability analysis and the solid line a fit to the numerical data with the form $v(L) = 6.95 - 5.39L^{-0.37}$.

analysis that $v \geq r$ is indeed correct, and this is because the intuitive idea that the front velocity cannot exceed the wetting rate is wrong.

This notion would be correct if the interface between the advancing dense phase and the empty lattice were flat, in the sense that all the planes up to the interfacial position $z = \bar{h}$ are completely wet, and completely dry above that. However the interface is in fact rough: i.e. it contains peaks and troughs. At an interfacial peak (formed by a fluctuation) wet sites can diffuse sideways. Then, they grow new clusters from a position that was higher than the previous interfacial position away from the local maximum—see figure 6.5. Thus those parts of the interface that are lagging catch the leading parts of the interface up as a consequence of the sideways diffusion process. Although this argument is strictly only true for a finite diffusion rate, the same argument holds in the limit $D \to \infty$ if one thinks not of peaks, but of isolated particles in the leading-edge of the front jumping around instantaneously from one point to another in a plane at constant z before wetting. Alternatively, the fact that the dynamics generate holes in the bulk structure implies that its leading edge must move faster than the rate at which particles are added to the system.

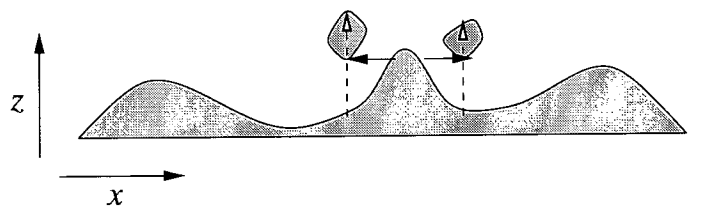

Figure 6.5: Mechanism for a front velocity greater than the wetting rate. At a peak, wet sites can diffuse sideways; then new clusters are formed which causes the interfacial position (as measured e.g. by the most extreme wet site in the z direction) to spring forward faster than the microscopic wetting rate r.

Finally, we note that the fit to the asymptotic velocities suggests that $\lim_{L\to\infty} v(L) = 6.95$, which is greater than the value $v^* = 5.449\ldots$ obtained from the stability analysis (equation (6.1.9)). Although this analysis does not preclude $v > v^*$, it is generally believed [127] that when a stable phase is invading an unstable phase (as is the case with the wetting model) the wavefront proceeds at the smallest velocity allowed by stability analysis (this is referred to as linear marginal stability [127]). I suspect that this is true of the wetting model and the reason that it is not seen in simulation may be due to noncommuting limits. The mean-field predictions for the asymptotic velocity apply in the limits $D \to \infty, L \to \infty$ and $t \to \infty$ taken in that order. In simulation, the limit $D = \infty$ was effected by assuming equilibration of each horizontal plane before wetting and drying-out events, but it was necessary to approach large times at fixed system size. This (roughly) corresponds to the limit $t \to \infty$ followed by $L \to \infty$ and as this order differs from that used to obtain the theoretical predictions, it could be that this is responsible for their difference from the numerical results.

6.2 Theory of interfacial growth: a review

Having seen that the properties of the interface between the dense wet phase and the empty lattice are important in understanding the mean-field description of the driven asymmetric contact process, I now study more closely the interface in the wetting model defined at the start of the chapter. First, in this section, I review some established growth models that will aid our understanding of the wetting model, along with the theory that describes them.

6.2.1 A selection of growth models

It is customary to couch interfacial growth models in terms of a d-dimensional planar *substrate* of dimension L indexed by the coordinate vector $\boldsymbol{r}$. Then the interface itself is described by the *height* above the substrate $h(\boldsymbol{r})$, usually assumed to be periodic (i.e. $h(\boldsymbol{r} + L\boldsymbol{e}_i) = h(\boldsymbol{r})$ where $\boldsymbol{e}_i$ is a unit vector). Using this framework, I describe three growth models with a linear substrate (i.e. $d = 1$ and $\boldsymbol{r} = x$).

Ballistic deposition (BD)

The ballistic deposition model was originally introduced as a description of colloidal aggregation and has since gained the status of a fundamental surface growth model in its own right [128, 129]. In this model, particles are dropped vertically downwards until they come into contact with part of the growing structure. In a lattice model, one normally implements nearest-neighbour sticking rules where a particle comes to rest when it reaches a site whose nearest

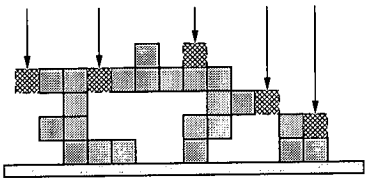

Figure 6.6: Model of ballistic deposition (BD). A randomly chosen substrate position is chosen and a particle dropped vertically until it reaches a site whose nearest-neighbour is occupied. Note the presence of holes in the bulk structure.

neighbour is occupied—see figure 6.6. In terms of the height function, BD corresponds to the update $h(x) \to \max\{h(x)+1, h(x-1), h(x+1)\}$ where x is a randomly chosen coordinate.

Restricted solid-on-solid model (RSOS)

The restricted solid-on-solid model [130] is the same as BD with an additional constraint: at no point is the height at two adjacent sites allowed to differ by more than one unit. That is $|h(x+1) - h(x)| \leq 1$. This constraint prevents the formation of bubbles or overhangs (which will be collectively referred to as *holes*) in the structure—see figure 6.7(i).

It is interesting to consider the case where the rates at which particles stick to the top or side of a growing structure differ [46]. In particular if the ratio of the side-sticking rate to the top-sticking rate is taken to infinity, the surface evolves by troughs becoming 'filled-in' until it is completely flat. We have met this model before. Downward steps (i.e. decreases in the surface height as one looks from left-to-right) move ballistically to the right, upward steps to the left and when two steps meet they 'annihilate'. In other words, this limit corresponds to deterministic ballistic annihilation with equal initial densities of left- and right-moving particles—see figure 6.7(ii).

The generalisation of ballistic annihilation in chapter 3 to include coalescence is unphysical in terms of this growth model. However, the possibility of unequal initial densities corresponds to a final interface that is tilted and a stochastic annihilation probability corresponds to the creation of a terrace when two steps meet. As we saw in chapter 3, these two modifications

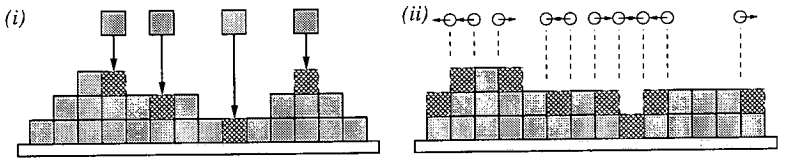

Figure 6.7: Restricted solid-on-solid models. (i) Same as ballistic deposition, but step sizes may not exceed a single particle's height. This prevents the formation of holes in the bulk structure. (ii) Model in which only side-sticking is allowed (new particle positions shown hatched) and mapping to a model of ballistic annihilation: downward steps move to the right and upward steps to the left.

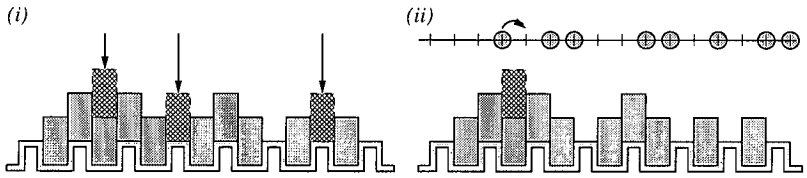

Figure 6.8: (i) Single-step model. Here particles are twice as tall as they are wide, and step-sizes are limited to one particle width. (ii) Mapping to asymmetric exclusion process. Upward steps (looking from left-to-right) are mapped to holes and downward steps to particles. The addition of a particle in the growth model corresponds to the rightward hopping of a particle in the asymmetric exclusion process, as shown.

can give rise to a phase transition from exponential relaxation to exponential multiplied by a power-law.

Single-step model

The single-step model [131, 132] is similar in spirit to the RSOS model in that no holes can be created during the growth process. In this model, particles have a height of two lattice units and the initial condition is constructed so that at even-numbered coordinates x, $h(x)$ is also even whereas at odd-numbered coordinates, $h(x)$ is odd. Particles are then added randomly to the surface subject to the constraint that $|h(x) - h(x+1)| = 1$ at all times—see figure 6.8(i).

This model has also already appeared in this thesis under a different guise, namely the totally asymmetric exclusion process with periodic boundaries (q.v. chapter 4). To see this, one associates particles with downward steps in the single-step model, and holds with upward steps. Then the growth move corresponds to a rightward hopping of a particle—see figure 6.8(ii). The step-size constraint of the single-step model implies that the associated hopping process is an exclusion process with conserved particle number. Also the hopping process must be periodic as the growth model is periodic.

Common features: the dynamical scaling hypothesis

In each of these growth models it is found that an initially flat interface *roughens* in time. This roughening is conveniently studied through examination of the interfacial width $W(t)$

$$W(t)^2 = \langle \Delta h(t)^2 \rangle \quad \text{where} \quad \Delta h(t)^2 = \overline{h(t)^2} - \overline{h(t)}^2 \,. \tag{6.2.1}$$

The angle brackets represent an average over different realisations of the noise (an ensemble average) whereas the overbar denotes an average over substrate coordinates in a given realisation at a time t.

For a wide range of models, including those introduced above, one finds power-law behaviour in the width. At early time $W(t) \sim t^{\beta}$, where β is the early-time growth exponent, and at

late times it saturates at $W(t) \sim L^{\alpha}$ where α is the roughness exponent. In fact, the data for different system sizes can be collapsed onto a single curve according to the dynamical scaling relation [133]

$$W(t) = t^{\beta} f(t/L^{z}) \; . \tag{6.2.2}$$

Here, $f(u)$ is a scaling function and z a further exponent, the dynamic exponent. Note that $\lim_{t\to\infty} W(t) \sim L^{\alpha}$ implies that $\lim_{u\to\infty} f(u) \sim u^{-\beta}$ and therefore that $\alpha = \beta z$. Note also that this dynamical scaling applies only to the interface. That is, the density field of the growing bulk does not, in general, exhibit any scale-invariant (fractal) properties.

The origin of the growth exponents can be understood intuitively (albeit rather qualitatively) as follows. If there were no interactions between neighbouring sites on the substrate (e.g. no nearest neighbour sticking or neighbouring height constraint), the height $h(x)$ increases at each point x independently of any other. Then, from the central limit theorem, the variance of height grows linearly with time and hence $W(t) \sim t^{1/2}$. The presence of interactions implies that the height variables must be correlated at different positions, suggesting that the growth of the width might proceed more slowly than $W(t) \sim t^{1/2}$. How slowly is then quantified by the exponent β.

The saturation of the interfacial width has its origins in the fact that the interfacial correlation length grows with time through a sequence of nearest-neighbour interactions. Then, saturation occurs once this correlation length is of the order of the system size, which takes a time $\sim L^{z}$.

These exponents have been measured numerically for a wide range of models [129, 128]. In table 6.1 I report the results obtained from a variety of studies of the three models introduced above. These reveal that roughly similar exponents have been observed for each: $\alpha \approx 0.5$, $\beta \approx 0.33$ and $z \approx 1.5$ (although looking at the table, one notes that the error-bar on this last figure would be rather large). The similarity of the exponents is suggestive of universal

Model	α	β	z	Reference
BD	0.42	0.30		[133]
BD	0.47	0.30		[131]
BD		0.33		[134]
BD	0.45	0.32	1.40	[135]
RSOS	0.5	0.33		[130]
RSOS	0.489	0.332		[136]
Single-step	0.5	0.33		[131]
Single-step			1.57	[132]

Table 6.1: Growth exponents measured for a range of models in 1+1 dimensions. Based on a similar table presented in [128].

behaviour, and indeed it is believed that all three models fall into the Kardar-Parisi-Zhang universality class.

6.2.2 Kardar-Parisi-Zhang description of interfacial growth

The d+1-dimensional[1] Kardar-Parisi-Zhang (henceforth KPZ) equation is a stochastic differential equation (or Langevin equation, see section 2.5.2) for a height function in continuous d-dimensional space and time. It reads

$$\frac{\partial}{\partial t}h(\mathbf{r},t) = v_0 + \nu\nabla^2 h + \frac{\lambda}{2}(\nabla h)^2 + \eta(\mathbf{r},t)\ . \tag{6.2.3}$$

Here $\eta(\mathbf{r},t)$ is a Gaussian white noise of zero mean and correlator $\langle\eta(\mathbf{r},t)\eta(\mathbf{r}',t')\rangle = \Gamma\delta^d(|\mathbf{r}-\mathbf{r}'|)\delta(t-t')$ and the parameters λ and ν are model-dependent. This equation is extremely well-studied (see [128, 129, 137] for reviews); I recount some of its properties below.

It should first be noted that a rigorous derivation of (6.2.3) from a microscopic growth model is a difficult exercise and has only been achieved for restricted solid-on-solid models [138]. Nevertheless the appropriateness of (6.2.3) as a coarse-grained description is easily justified using phenomenological arguments as follows. Firstly, the constant v_0 is simply the mean rate at which a flat interface would proceed in the absence of noise; the Laplacian represents a tendency of the interface to smoothen (e.g. through surface tension) and the nonlinear term expresses the fact that one generally expects growth to occur normal to an interface. The noise term models fluctuations superposed on deterministic growth rules.

Some understanding of the KPZ equation can be gained by first considering the case $\lambda = 0$. The resulting linear equation, known as the Edwards-Wilkinson (EW) equation, can be solved using Fourier transforms [137]. One finds that the mean interfacial position $\bar{h}(t)$ advances at velocity v_0 and that the interfacial fluctuations $W(t)$ satisfy the dynamical scaling form (6.2.2) with exponents $\alpha = \frac{1}{2}$, $\beta = \frac{1}{4}$ and $z = 2$ for a one-dimensional interface.

When the nonlinear term is present, and growth normal to the interface is enhanced, the mean interfacial position $\bar{h}(t)$ advances at a speed greater than v_0 if the interface is not flat (as it will not be, in general, due to the presence of noise). This is precisely the situation that arose in the study of the wetting model at the beginning of this chapter.

The scaling behaviour of an interface governed by the KPZ equation in 1+1 dimensions is well understood and the exponents known exactly. The roughness exponent α can be calculated by using the fact that if one rewrites the 1+1d KPZ equation (6.2.3) as a Fokker-Planck equation (see section 2.5.2), one finds that (for finite L) it admits a stationary solution that is

[1] In common with the notation for directed percolation, the spatial dimension of the substrate d and the temporal dimension are explicitly separated for clarity.

independent of λ [129]. Hence the exponent α takes on the same value of $\frac{1}{2}$ as the linear EW equation.

One way[2] to determine the dynamic exponent z is through the mapping of the single-step model to the totally asymmetric exclusion process (TASEP) as described in section 6.2.1. In the review of exclusion processes in section 4.8.1 it was noted that the diffusion constant Δ of a single particle in the TASEP with periodic boundaries and a finite density of particles has been shown to behave as $\Delta \sim L^{-1/2}$ where L is the system size. Given that for large times, Δ multiplied by t is the variance of particle hops in the ASEP, and that a particle hop in the ASEP corresponds to a 'unit' of surface growth in the single-step model (see figure 6.8(ii)) we see that $W(t)^2 \sim L^{-1/2}t$ from this result. However the knowledge that $W(t)^2$ eventually saturates to $L^{2\alpha} = L$ implies that the time taken until saturation is $\sim L^{3/2}$. This last exponent is the dynamic exponent z of equation (6.2.2).

Using the relation $\alpha = \beta z$, one finds the 1+1d KPZ exponents exactly as $\alpha = \frac{1}{2}$, $\beta = \frac{1}{3}$ and $z = \frac{3}{2}$. We note that these exact values are in good agreement with the simulation results of 1+1d growth models given in table 6.1. The robustness of these exponents can be traced to a very important property of the 1+1d KPZ equation in that the derivatives of h that appear in (6.2.3) are the only ones whose contribution does not diminish under a rescaling (renormalisation group) transformation. In other words, even if at short length- and timescales, an equation containing other derivatives of h better describes a particular growth model, the *large-scale* interfacial fluctuations of the model in question will be adequately described by the KPZ equation (6.2.3). This statement is not true for d+1-dimensional KPZ growth with $d \geq 2$. Indeed it is believed [129] that there is a critical value of λ below which no interfacial roughening occurs. For larger values of λ (the strong-coupling regime in which the interface is rough) the KPZ equation cannot be treated perturbatively, and there is also controversy as to whether the upper critical dimension (above which the interface is always flat) exists [137, 139]. As I consider only 1+1d growth models in this chapter, these difficulties can be happily avoided.

6.2.3 Scenarios for non-KPZ scaling

Given the applicability of the KPZ description to a large range of systems with a one-dimensional substrate, it is interesting to note those growth processes for which it is known *not* to apply. I briefly review here two cases which will be of importance in the subsequent study of the wetting model.

Firstly, the noise term $\eta(\boldsymbol{r}, t)$ in the KPZ equation (6.2.3) does not exhibit any correlations. In a real (experimental) system, one expects the random forces on an interfacial segment to be

[2]Fuller discussion of the various ways to calculate the 1+1d KPZ exponents can be found in [129, 137]

correlated in both space and time. The effects of correlations have been studied through a range of methods (reviewed in [129]) and the general consensus seems to be that if the correlations are short-ranged, or power-law with a sufficiently large negative exponent (so in some sense also short-ranged) the KPZ scaling exponents are unaffected. For sufficiently long-ranged correlations both the roughness exponent α and the early-time growth exponent β increase with the range of the correlations. This applies to both spatial and temporal correlations.

The other main candidate for non-KPZ scaling is the presence of nonlocal dynamics in the growth model. By nonlocal dynamics it is meant that the velocity of the interface at a point depends on quantities other than the gradient of the height function there. Recall that it has been shown that $1+1$d growth equations involving only gradients of $h(x)$ converge to the KPZ equation on large length scales.

A concrete example of a situation in which nonlocal effects might be important is a growth model in which holes are present. The wetting model introduced at the start of this chapter is a case in point: the drying out of a wet site at the interface causes the height of the interface to recede by the size of the dry patch behind the interface. Intuitively it is understood that there is an interfacial region of some size ξ over which the density of wet sites decays to zero, and that fluctuations of this density in that region can affect the height of the interface. An interface with this property will be referred to as *smeared-out* and is associated with a growing structure that contains holes, or is noncompact.

To understand the relationship between noncompact structures and smeared out interfaces, consider again the RSOS and single-step growth models (section 6.2.1). In both of these models there is a constraint that ensures that no holes are formed in the bulk. Then the interface is *sharp* and thus there is a one-to-one correspondence between an interfacial configuration $h(\boldsymbol{r})$ and the entire density field. In this case, the interfacial dynamics are local, and the KPZ description should apply. Indeed, as table 6.1 shows, the measured growth exponents for RSOS and single-step models are in good agreement with the KPZ values.

There has been some numerical evidence that a growth model qualitatively similar to the wetting model of this chapter exhibits non-KPZ scaling [140, 141]. The model, a stochastic lattice realisation of a Fisher wave [142], is similar to the present wetting model in the sense that it describes a stable phase (the wet bulk) invading an unstable phase (the dry lattice ahead of the interface). Microscopically the model also involves wetting, drying out and diffusion moves and so holes are present in this model too. The claim in [141], supported by analytical work, is that the size of the interfacial region ξ is sufficiently large that one cannot adequately describe the entire system through an interface alone. This idea is manifest in the exponents measured for Fisher waves: in particular the roughness exponent α was found to be close to 0.4 for $d = 1$ in comparison to the KPZ value of 0.5. A value of $\alpha \simeq 0.4$ corresponds to the

KPZ exponent for a two-dimensional substrate, and it has been suggested that, as ξ is large, the correct surface to consider in terms of the scaling behaviour is (a suitable transformation) of the bulk density field which necessarily has a dimension one greater than the interface [143].

In the following section, I shall investigate the scaling exponents of the present wetting model in the context of the foregoing discussion. First, however, I make a final observation about the growth of noncompact structures. Consider again table 6.1 which shows the exponents measured for a range of growth models. Whilst the models involving compact structures (RSOS and single-step) show good agreement with the KPZ exponents, the values obtained for the ballistic deposition model (in which holes are present) are less consistent with the KPZ values. The following study of the asymptotic interfacial behaviour of the wetting model will put us in a position to understand this discrepancy.

6.3 Interfacial growth in the wetting model

To determine the interfacial behaviour of the wetting model (as defined in section 6.1 above) I performed Monte Carlo simulations of the wetting model with finite diffusion rate D. The simulations themselves were conducted in a manner similar to that described for the case of infinite diffusion (section 6.1.3). However, in the present case, the full two-dimensional array of lattice occupancy variables must be manipulated, implying much longer simulation run times than for $D = \infty$. In fact, a back-of-an-envelope calculation reveals that to simulate until the interface saturates requires a run time that grows as L^4. This placed a practical and, as we shall see, crucial upper limit on the system sizes that could be accessed.

6.3.1 Numerical data for interfacial scaling

I consider first the growth of the interface in the case where the drying out process is suppressed, i.e. $k = 0$, and where no holes may be created spontaneously in the bulk. Thus the bulk density is one, and the only holes present are in a small zone (comprising a few lattice sites) near the interface and which are a consequence of the sideways diffusion process. The interface is then considered effectively sharp and one thus expects to find KPZ behaviour, as discussed in section 6.2.3 above.

This expectation is tested by averaging the variance of the height distribution (where the height at substrate coordinate x is the position of the most extreme wet site at that coordinate) over many simulation runs performed at a fixed system size. This then yields an estimate of the function $W(t)$ as given through equation 6.2.1. To extract the scaling exponents, one performs a data collapse according to the dynamical scaling relation (6.2.2)—that is, one plots $W(t)/L^{\alpha}$

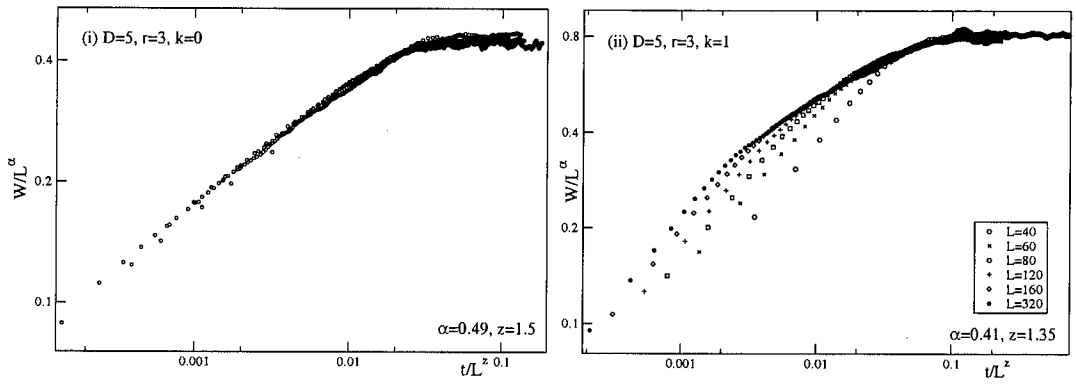

Figure 6.9: The width function $W(t)$ in the wetting model with the parameters $D = 5$, $r = 3$ and the drying-out process (i) inactive and (ii) active. Data from different system sizes L were made to collapse (at least at late times) by scaling W with L^{α} and t with L^{z}. The values of α and z used in each case are shown on the plots.

against t/L^z and varies α and z until the best collapse is seen. The collapse for the case $k = 0$ is shown in figure 6.9(i) for a range of system sizes and representative values of $D = 5$ and $r = 3$. Good data collapse was achieved using exponents consistent with KPZ scaling: $z = 1.50(5)$, $\alpha = 0.49(2)$. Again, the uncertainty in these figures was estimated by varying z and α until the collapse became noticeably worse.

The second plot in figure 6.9 shows the corresponding collapse for the case where k=1, i.e. wet sites can dry out leaving holes in the bulk. The scaling exponents had to be changed to z=1.35(5), α=0.41(2) in order to obtain collapse of the saturation width. Note that in comparison to figure 6.9(i), the collapse is far less convincing. Therefore, although it appears at first sight that the present wetting model exhibits a roughness exponent α consistent with that found for Fisher waves [140, 141] it is not clear that the true asymptotic scaling regime was probed before the required simulation run times rendered increases in L impractical.

However, given the wealth of knowledge surrounding the KPZ equation and the criteria for non-KPZ scaling, other properties of the wetting model can be investigated to determine whether the exponents that were measured for the case $k = 1$ persist as L is increased, or whether in fact they will cross over to the KPZ values.

6.3.2 Extrapolation to the scaling regime

To assess whether the true scaling regime had been accessed in the simulations, I consider how a smeared-out interface might be described through a modified version of (6.2.3). In order to incorporate a generic coupling of the interface to holes in the bulk an additional term is introduced into (6.2.3) to give

$$\frac{\partial}{\partial t}h(x,t) = v_0 + \nu\nabla^2 h + \frac{\lambda}{2}(\nabla h)^2 - kg(x) + \eta\,. \tag{6.3.1}$$

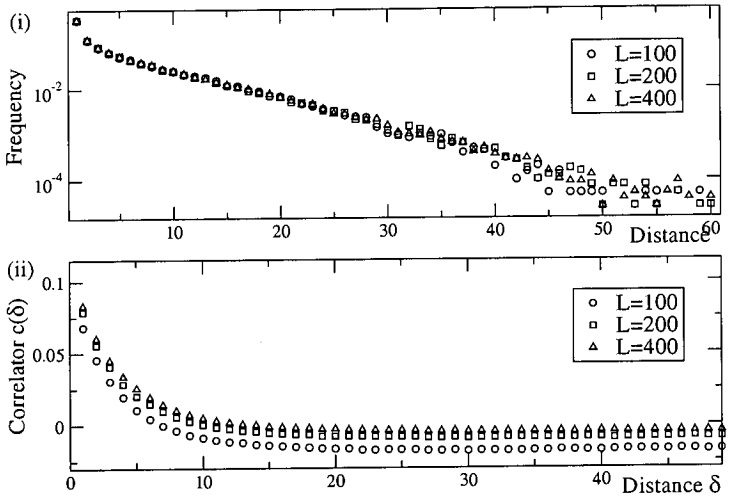

Figure 6.10: (i) The gap-size g probability distribution for the wetting model with L=100, 200, 400. (ii) The gap-size correlation function in the x-direction for the wetting model with L=100, 200, 400. In both cases D=5, r=1.5 and k=1.

An explicit dependence on k has been included in such a way that when k=0 this equation reduces to (6.2.3), as required. The function $g(x)$ is intended to capture any nonlocal dynamics present. In the case of the wetting model, $-kg(x)$ is the rate at which $h(x)$ decreases when a particle at the interface disappears, implying that $g(x)$ is the size of the gap (dry patch) behind the interface. Microscopically, this gap size is defined as the distance between the uppermost and second uppermost particle at position x. It has been checked explicitly that decreases in $h(x)$ follow closely the profile of the gap size $g(x)$ and so I believe that (6.3.1) describes the wetting model adequately, at least on a coarse-grained level.

The question I now proceed to address is whether the extra term in (6.3.1) affects the long-wavelength properties of the interface. To this end the statistics of $g(x)$ in the wetting model were obtained from simulation and it was found that the distribution of gap sizes became stationary rather quickly, at times when the interface was still roughening. In figure 6.10(i) the stationary gap size distribution is shown for three different system sizes. Note in particular that the distribution is system-size independent and decays exponentially. Thus the finite length scale ξ associated with the gap remains constant as the substrate length is increased, and is irrelevant once L is sufficiently large (in the sense that $\xi/L \to 0$ as $L \to \infty$). Of course, at smaller L, $g(x)$ will play some role.

Now I argue that unless $g(x)$ exhibits scale-invariant correlations in the x direction, it will be rescaled into the noise term already present in (6.2.3). Specifically, g can be replaced with $g_0 + \tilde{\eta}$ where g_0 is a constant which can be absorbed into v_0 and $\tilde{\eta}$ is a noise term with irrelevant short-range correlations that can be absorbed into the noise term η in (6.2.3). Recall from the discussion in section 6.2.3 that only sufficiently long-ranged correlations have been found to

affect the asymptotic scaling behaviour.

The possibility of long-range correlations in the wetting model can be excluded by examining the correlation function $c(\delta) = \langle g(0)g(\delta)\rangle / \langle g(0)^2\rangle - 1$, shown in figure 6.10(ii). It is clear that for L=100, 200, 400 the correlation length retains the same (finite) value. The small anticorrelation for large δ vanishes with increased L and can be understood to be a consequence of finite system sizes.

Thus, these results strongly suggest that the term $g(x)$ modifying the KPZ equation from (6.2.3) to (6.3.1) is, for the wetting model, purely cosmetic. In short, I contest that if L were increased beyond the values used in figure 6.9 a crossover towards KPZ scaling would be observed. However, I believe the system sizes that would be required to demonstrate quantitively the crossover in the wetting model are unfeasible, although it has been suggested [144] that a suitable coarse-graining of the density field and redefinition of the height $h(x)$ (see e.g. [141] for details of such a scheme) may help to render the crossover effects less severe.

6.4 Crossover to KPZ scaling in noncompact structures

I now examine the nature of the crossover to KPZ scaling that I believe to be present in the wetting model by devising a simplified version of the system. This reduced model is constructed by replacing the two-dimensional density field of the wetting model with an interface $h(x)$ coupled to a gap of size $g(x)$. A parameter is also included that allows the interface to be varied from being sharp to being smeared-out, thereby retaining the wetting model's most important feature. Thus this new model, which I will refer to as the ballistic deposition and desorption (BDD) model, should also be governed by (6.3.1).

The dynamics of the BDD model are defined as follows—see also figure 6.11. At each time step, one chooses a substrate position x at random then performs either a particle deposition move (at unit rate) or a desorption move (at rate κ). In the former case, a particle is dropped

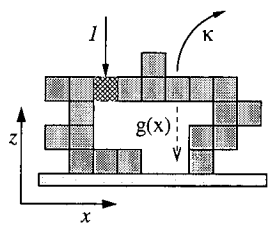

Figure 6.11: Ballistic deposition and desorption (BDD) model. As in ballistic deposition (figure 6.6) particles are added to the structure at unit rate, sticking to nearest-neighbour sites. In the BDD model, particles also desorb from the surface, causing the height to decrease locally by a distance $g(x)$ equal to the size of the gap behind the extremal particle. Note that in the simulation only the interfacial height $h(x)$ and gap size $g(x)$ are stored.

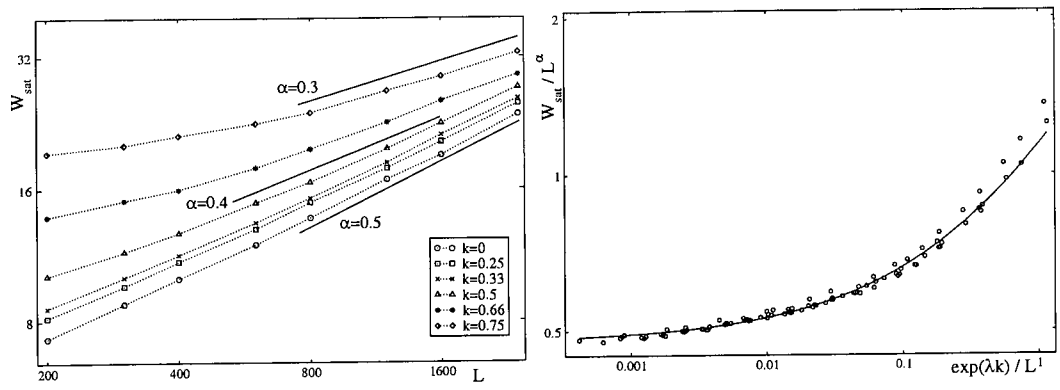

Figure 6.12: (i) Log-log plot of saturation width W_{sat} for different L and κ in the BDD model. The solid lines correspond to α=0.5 (KPZ), 0.4 and 0.3 and illustrate how underestimates in α arise before the scaling regime is reached. (ii) The same data scaled according to equation (6.4.1). The parameters α=0.5 and λ=7.4 were used and the solid line corresponds to A=0.47, B=0.66 and γ=0.55 in (6.4.1).

vertically downwards in column x until it comes to rest at a site whose nearest-neighbour is occupied. This increases the height $h(x)$ by an amount which defines the gap size $g(x)$. A desorption move is implemented by decreasing $h(x)$ by $g(x)$ and replacing $g(x)$ with one of the other gap sizes in the system, chosen at random. The desorption move serves to smear out the interface, and thus the rate κ plays the same role as k in the wetting model (although no numerical equivalence between the two should be assumed).

With the desorption rate κ set to zero, one recovers the ballistic deposition model of section 6.2 which is a widely accepted realization of the KPZ equation (6.2.3) [128]. Thus in this limit, the interface should exhibit KPZ scaling behaviour. With nonzero κ the height $h(x)$ can decrease by an amount specified through the random variable $g(x)$. Simulations have shown that the statistics of $g(x)$ in the BDD model are very similar to those shown in figure 6.10 for the wetting model: both the gap distribution and correlation functions are stationary and have no dependence on system size; the gap size distribution has an exponential tail and the correlation length in the x direction is almost zero. Thus I believe that the BDD model mimics the wetting model in all respects that may affect universal scaling behaviour.

A clear insight into interfacial structure as one approaches the scaling regime is obtained from figure 6.12(i) which shows the saturation width against system size for a range of desorption parameters κ. With both axes logarithmic, one can define an effective exponent $\alpha_\kappa(L)$ as the gradient of the function $\log W_\kappa(\log L)$ with respect to $\log L$. As is evident from the graph, there is a significant range of L, κ for which $\alpha_\kappa(L) \approx 0.4$–$0.45$ (similar to the values of [140, 141]). However the graph also gives evidence for a trend in which the limit $\lim_{L\to\infty} \alpha_\kappa(L)$ coincides for all values of κ. This limit would give the true scaling value for the roughness exponent α for all κ. This can be shown by collapsing all the data of figure 6.12(i) onto a single curve.

To perform the collapse the simple crossover scaling form

$$W_{\rm sat}(L,\kappa) = L^{\alpha}[A + B(\ell(\kappa)/L)^{\gamma}] \tag{6.4.1}$$

was used, in which $\ell(\kappa)$ is some finite length induced by the nonlocal desorption process and γ is the crossover exponent. Remarkably, a reasonable collapse for all L, κ could be effected by taking A, B constant and $\ell(\kappa) = \exp \lambda\kappa$ with λ constant—see figure 6.12(ii). Thus it appears that all the κ-dependence enters through $\ell(\kappa)$ in a very simple way.

The fit to (6.4.1) allows a precise estimate of α=0.50(1) (coincident with the KPZ value) for the BDD model for all κ. Without invoking (6.4.1) the estimation of α would be hampered by the slow power-law crossover to the asymptotic regime and underestimates of α would be obtained (as discussed above). This crossover could explain the discrepancies between previous estimates of α for BD (κ=0) and the KPZ value (see table 6.1).

It is interesting to see how the crossover scaling form (6.4.1) generalises the standard procedure of introducing an intrinsic width. This was first used in the context of Eden growth model [145, 128] where the bulk structure expands by particles sticking to its perimeter. In the intrinsic width approach one writes $W^2_{\rm sat} = AL^{2\alpha} + w_i^2$, i.e. the squares of the scaling width (which grows as L^{α}) and the intrinsic width w_i are added together. The scaling form (6.4.1) coincides with this procedure for large L if $\gamma = 2\alpha = 1$. Moreover, $\gamma = \alpha = \frac{1}{2}$ would correspond to the distinct procedure of the addition of two widths without squaring them first. For the BDD model, I have found that data collapses, all of reasonable quality, could be achieved for values of $\gamma \approx 0.5$–0.7. Thus the present data are more consistent with the linear addition of two widths, although the large error on γ makes it difficult to make any further statements as to the significance of its value.

As the BDD model is believed to be a suitable reduction of the full wetting model, I would expect a similar crossover to 1+1-dimensional KPZ scaling to be exhibited in the wetting model. This result is in contrast to the findings of [141] in which it was proposed that if a stable phase is invading a linearly unstable phase (as is the case of the wetting model) there exists a lengthscale that characterises the decay of the density in the interfacial region that does not diminish as the substrate size is increased. In a later work [143] it was argued that the consequence of this additional lengthscale should lead to d+2-dimensional KPZ scaling. The present work would appear to contradict this suggestion, and so it is necessary to investigate more closely the lattice model studied in [141] (a stochastic realisation of a Fisher wave) for crossover behaviour before reaching any firm conclusion as to the scaling behaviour of a smeared-out interface.

6.5 Summary of results

In this chapter I considered the time-dependent properties of the driven asymmetric contact process (DACP) introduced in the previous chapter and a two-dimensional wetting model derived from it. I have shown that, in a particular limit, the wetting model is equivalent to the mean-field approximation to the one-dimensional DACP and found that travelling waves in this system travel at a speed faster than the fundamental wetting rate. This phenomenon was attributed to the presence of sideways diffusion in the model.

The interfacial properties of the wetting model (for general values of the model parameters) were found to be described by the KPZ equation for large system sizes, although numerical results for small systems were greatly affected by finite-size effects. The data presented for the wetting model provided evidence for the finite-size effects to become slower when the growing structure contains holes.

I conjectured that the wetting model should exhibit a crossover to the KPZ dynamical scaling regime as the system size is increased. This crossover was examined closely in a simplified model (the BDD model) and I was able to determine that the crossover is manifested as a power-law correction to the scaling behaviour of the interfacial width. It would be interesting to see whether the form that was used (and which generalises earlier approaches to the same problem) is applicable to a wider class of models.

In these investigation of the dynamic properties of the wetting model, I have deliberately used only wetting rates above the percolation threshold. Then, indefinite growth of the dense, wet structure was ensured: this of course is necessary to study the interfacial scaling properties. However, we know from the previous chapter that there is a transition to a regime where the extent of the growing phase saturates and it would be interesting to understand how the directed percolation transition is manifested in the wetting model. In fact, the wetting model comprises a contact process and an exclusion process acting at right-angles to one another. Hence it is not obvious whether the wetting transition should be in the 1+1-dimensional or the 2+1-dimensional directed percolation class.

It would also be of interest to see how interfacial properties behave near the critical wetting rate. From the results of this chapter, we understand the crossover to KPZ interfacial scaling to become more pronounced as the density of holes is increased. However, at the wetting transition we also understand that correlation lengths diverge. Presumably, therefore, there is a crossover from KPZ scaling to a regime in which the interfacial properties can be related to directed percolation. There is also the matter of the interfacial velocity near criticality: the mean-field analysis predicts a discontinuity in the velocity at criticality. Preliminary simulation data (not presented here) suggests that there is indeed a rapid increase in the interfacial velocity at the

critical wetting rate. However it is very difficult to tell whether the increase is discontinuous or, e.g. power-law with a large amplitude and so this remains an open question.

Chapter 7

Conclusion and outlook

In this thesis I have considered a number of model systems that are governed by probabilistic microscopic dynamics with the aim of understanding the macroscopic phenomena that emerge. As discussed in chapter 1 the physical motivation for formulating stochastic models lies in the fact that they can be used to describe systems that are not at equilibrium but are relaxing to an equilibrium steady state, or else are driven by their environment. In the absence of a unified framework for studying nonequilibrium systems, I considered separately a range of specific models and in each case established some new results, through both analysis and computer simulation. Towards the end of each chapter I summarised the main results obtained for each model and posed open questions. In this final chapter I wish to comment on the more general common themes that have arisen from this study of distinct model systems.

Firstly, it has been shown that stochastic models admit phase transitions even in one dimension. This is an interesting observation because it is understood that at equilibrium, one-dimensional systems with short-range (e.g. nearest-neighbour) interactions do not exhibit phase transitions. One example of a nonequilibrium phase transition was found to exist in the stochastic particle reaction system of chapter 3. Through an exact solution for the density decay of the system, I showed that different particle density decay forms result as a direct consequence of introducing stochasticity into the reaction dynamics. This is slightly counterintuitive as one might expect that the sole consequence of reducing the probability that two particles react would be to slow down the density decay, rather than change its analytic form.

In chapter 4 an exact solution demonstrated how the interactions between the boundary sites of a lattice and the external environment affect the bulk properties of an exclusion process. In particular, when the rates at which particles can enter and leave the system are varied, the steady-state current exhibits nonanalyticies and the density profiles change from an exponential to a power-law form. This again is in contrast to equilibrium systems in which the effects of boundaries are usually short-ranged (i.e. characterised by a finite lengthscale). The presence of

generic power-law correlations (e.g. in the entirety of the maximal current phase of exclusion processes) is a further interesting feature of nonequilibrium systems: at equilibrium length- and timescales generally diverge only at isolated critical points.

The second main feature shared by several of the systems studied in this thesis is that of robustness of the macroscopic behaviour under certain changes of the microscopic dynamics. A manifestation of this phenomenon familiar from equilibrium statistical physics is universality at critical points which has been seen to apply also to nonequilibrium systems. For example, in chapter 5 I provided evidence for the driven asymmetric contact process to exhibit a continuous wetting transition characterised by directed percolation exponents. This is despite the lack of an absorbing state which is arguably the most important criterion for models to be in the DP universality class. I also argued that a two-dimensional generalisation of this process (the wetting model of chapter 6) exhibits a dynamical scaling regime in the 1+1d Kardar-Parisi-Zhang (KPZ) universality class.

The ubiquity of KPZ interfacial scaling (at least in 1+1 dimensions) can be traced to the fact that when the growth dynamics depend only on the local shape of the interface (as long as the nonlinear term proportional to the square of the gradient is present), the KPZ description is recovered at large length- and timescales. Hence, a wide range of microscopic growth models share the same roughening phenomena once the scaling regime is reached.

The situation regarding the directed percolation (DP) universality class is currently less well understood. As discussed in chapter 5, studies of a range of models all showing DP exponents has led to the directed percolation conjecture. This relates generic properties of a microscopic model (e.g. the nature of the order parameter) to macroscopic critical phenomena and is thus similar in spirit to the identification of an equilibrium model's symmetries with its critical behaviour. Nevertheless, a proof of the directed percolation conjecture is still outstanding (although it should be noted that no data currently contradict it) and one may speculate that finding such a proof could in itself improve our wider understanding of nonequilibrium critical phenomena.

Robustness of collective phenomena is not restricted to universal scaling behaviour. For example, I have shown in chapter 3 that two-species annihilation and coalescence reactions are intimately related for ballistic particle motion in one-dimension which is a result similar to that which has been long established for diffusive annihilation and coalescence systems. It is also interesting to note that the universal $t^{-1/2}$ density decay of diffusive annihilation and coalescence reactions in one dimension also applies to ballistic reaction systems, albeit only for special combinations of the model parameters. Additionally, as discussed in section 4.8, the phase diagram of the partially asymmetric exclusion process has been found to retain a similar form under variation of the microscopic updating scheme, addition of further particle species

and so on.

In the continuing absence of a general theory for nonequilibrium systems this robustness provides a glimmer of hope that it may be possible to formulate such a theory, at least for a restricted class of systems. In this respect, it is also interesting to note emerging connections to general theories of equilibrium systems. One that has been discussed at several points in this thesis is the relationship between partition function zeros and phase behaviour as introduced in section 2.4.1. Zeros of the steady-state normalisation (a nonequilibrium analogue of the partition function) have been investigated in the context of exclusion processes (see the discussion in section 4.8.4) and directed percolation (section 5.4) and it appears that the steady-state normalisation could encode important information about phase behaviour of general stochastic processes in addition to the smaller class of equilibrium systems.

Nevertheless, a general understanding of nonequilibrium systems, and in particular of experimental systems, remains a theoretical challenge. Hopefully, through continued work of the type described in this thesis, it may become possible to meet this challenge.

Bibliography

[1] R. A. Blythe, M. R. Evans, F. Colaiori and F. H. L. Essler (2000) Exact solution of a partially asymmetric exclusion model using a deformed oscillator algebra. *J. Phys. A: Math. Gen.* **33** 2313.

[2] R. A. Blythe, M. R. Evans and Y. Kafri (2000) Stochastic ballistic annihilation and coalescence. *Phys. Rev. Lett.* **85** 3750.

[3] R. A. Blythe and M. R. Evans (2001) Slow crossover to Kardar-Parisi-Zhang scaling. *Phys. Rev. E* **64** 051101.

[4] M. R. Evans and R. A. Blythe (2001) Nonequilibrium dynamics in low dimensional systems. *Preprint* cond-mat/0110630.

[5] K. Huang (1987) *Statistical mechanics* (Wiley, New York), second edition.

[6] J. M. Yeomans (1992) *Statistical mechanics of phase transitions* (Oxford University Press, Oxford).

[7] J. L. Cardy (1996) *Scaling and renormalization in statistical physics.* Cambridge lecture notes in physics (Cambridge University Press, Cambridge).

[8] A. N. Shiryaev (1996) *Probability*, volume 95 of *Graduate Texts in Mathematics* (Springer-Verlag, New York), second edition.

[9] G. R. Grimmett and D. R. Stirzaker (1992) *Probability and random processes.* Oxford Science Publications (Oxford University Press, Oxford), second edition.

[10] S. Karlin (1966) *A first course in stochastic processes* (Academic Press, London).

[11] E. Seneta (1973) *Nonnegative matrices* (Allen and Unwin, London).

[12] P. Lancaster and M. Tismenetsky (1985) *The theory of matrices.* Computer science and applied mathematics (Academic Press, San Diego), second edition.

[13] F. Harary (1969) *Graph theory.* Addison-Wesley series in mathematics (Addison-Wesley, Reading, Mass.).

[14] C. Godsil and G. Royle (2001) *Algebraic graph theory*, volume 207 of *Graduate Texts in Mathematics* (Springer-Verlag, New York).

[15] F. P. Kelly (1979) *Reversibility and stochastic networks* (Wiley, New York).

[16] D. Mukamel (2000) Phase transitions in nonequilibrium systems. In *Soft and fragile matter*, edited by M. E. Cates and M. R. Evans, volume 53 of *Scottish Universities Summer School in Physics*, page 237 (Institute of Physics Publishing, Bristol).

[17] K.-C. Chou (1990) Applications of graph theory to enzyme kinetics and protein folding kinetics: steady and non-steady-state systems. *Biophys. Chem.* **35** 1.

[18] A. N. Jackson (2001) *Structural phase behaviour via Monte Carlo techniques.* Ph.D. thesis, University of Edinburgh.

[19] C. N. Yang and T. Lee (1952) Statistical theory of equations of state and phase transitions. I. Theory of condensation. *Phys. Rev.* **87** 404.

[20] C. N. Yang and T. Lee (1952) Statistical theory of equations of state and phase transitions. II. Lattice gas and Ising model. *Phys. Rev.* **87** 410.

[21] K. Krebs and S. Sandow (1997) Matrix product eigenstates for one-dimensional stochastic models and quantum spin chains. *J. Phys. A: Math. Gen.* **30** 3165.

[22] L.-H. Gwa and H. Spohn (1992) Bethe solution for the dynamical-scaling exponent of the noisy Burgers equation. *Phys. Rev. A* **46** 844.

[23] B. Derrida (1998) An exactly soluble non-equilibrium system: The asymmetric simple exclusion process. *Phys. Rep.* **301** 65.

[24] F. C. Alcaraz and R. Z. Bariev (1999) Exact solution of the asymmetric exclusion model with particles of arbitrary size. *Phys. Rev. E* **60** 79.

[25] N. G. van Kampen (1992) *Stochastic processes in physics and chemistry* (North-Holland, Amsterdam), second edition.

[26] H. Risken (1989) *The Fokker-Planck equation* (Springer-Verlag, Berlin), second edition.

[27] C. W. Gardiner (1985) *Handbook of stochastic methods* (Spinger-Verlag, Berlin), second edition.

[28] A. J. Schofield (1999) Strongly correlated quantum systems. Lecture notes from 1999 EPSRC Theory of Condensed Matter Summer School, Harlech, Wales (unpublished).

[29] M. Doi (1976) Second quantization representation for classical many-particle system. *J. Phys. A: Math. Gen.* **9** 1465.

[30] L. Peliti (1985) Path integral approach to birth-death processes on a lattice. *J. Physique* **46** 1469.

[31] J. L. Cardy (1997) Renormalisation group approach to reaction-diffusion problems. In *The mathematical beauty of physics*, edited by J.-M. Drouffe and J.-B. Zuber, volume 24 of *Advanced series in mathematical physics* (World Scientific, Singapore). See also cond-mat/9607163.

[32] S. R. White (1998) Strongly correlated electron systems and the density matrix renormalization group. *Phys. Rep.* **301** 187.

[33] R. J. Glauber (1963) Time-dependent statistics of the Ising model. *J. Math. Phys.* **4** 294.

[34] T. M. Liggett (1999) *Stochastic interacting systems: contact, voter, and exclusion processes* (Springer-Verlag, Berlin).

[35] S. Redner (1997) Scaling theories of diffusion-controlled and ballistically-controlled bimolecular reactions. In *Nonequilibrium statistical mechanics in one dimension*, edited by V. Privman, chapter 1 (Cambridge University Press, Cambridge).

[36] D. ben Avraham (1997) The coalescence process $A+A \to A$ and the method of interparticle distribution functions. In *Nonequilibrium statistical mechanics in one dimension*, edited by V. Privman, chapter 2 (Cambridge University Press, Cambridge).

[37] T. O. Masser and D. ben Avraham (2001) Correlation functions for diffusion-limited annihilation, $A + A \to \emptyset$. *Preprint* cond-mat/0106306.

[38] J. L. Spouge (1988) Exact solutions for a diffusion-reaction process in one dimension. *Phys. Rev. Lett.* **60** 871.

[39] C. R. Doering and D. ben Avraham (1988) Interparticle distribution functions and rate equations for diffusion-limited reactions. *Phys. Rev. A* **38** 3035.

[40] J. G. Amar and F. Family (1990) Diffusion annihilation in one dimension and kinetics of the Ising model at zero temperature. *Phys. Rev. A* **41** 3258.

[41] L. Peliti (1986) Renomalization of fluctuation effects in the $A+A \to A$ reaction. *J. Phys. A: Math. Gen.* **19** L365.

[42] B. P. Lee (1994) Renormalization group calculation for the reaction $kA \to 0$. *J. Phys. A: Math. Gen.* **27** 2633.

[43] B. P. Lee and J. Cardy (1995) Renormalization group study of the $A + B \to \emptyset$ diffusion-limited reaction. *J. Stat. Phys.* **80** 971.

[44] K. Krebs, M. P. Pfannmuller, B. Wehefritz and H. Hinrichsen (1995) Finite-size-scaling studies of one-dimensional reaction-diffusion systems I: Analytical results. *J. Stat. Phys.* **78** 1429.

[45] G. M. Schütz (1997) Diffusion-limited annihilation in inhomogeneous environments. *Z. Phys. B* **104** 583.

[46] J. Krug and H. Spohn (1988) Universality classes for deterministic surface growth. *Phys. Rev. A* **38** 4271.

[47] M. Gerwinski and J. Krug (1999) Analytic approach to the critical density in cellular automata for traffic flow. *Phys. Rev. E* **60** 188.

[48] S. Ispolatov and P. L. Krapivsky (1998) Ballistic coalescence model. *Physica A* **252** 165.

[49] Y. Elskens and H. L. Frisch (1985) Annihilation kinetics in the one-dimensional ideal gas. *Phys. Rev. A* **31** 3812.

[50] P. L. Krapivsky, S. Redner and F. Leyvraz (1995) Ballistic annihilation kinetics: The case of discrete velocity distributions. *Phys. Rev. E* **51** 3977.

[51] P. L. Krapivsky and C. Sire (2001) Ballistic annihilation with continuous isotropic initial velocity distribution. *Phys. Rev. Lett.* **86** 2494.

[52] P. A. Rey, M. Droz and J. Piasecki (1998) Search for universality in one-dimensional ballistic annihilation kinetics. *Phys. Rev. E* **57** 138.

[53] M. J. E. Richardson (1997) Exact solution of two-species ballistic annhilation with general pair-reaction probability. *J. Stat. Phys.* **89** 777.

[54] Y. Kafri (2000) Exact solution of homogeneous ballistic annihilation with a general reaction probability. *J. Phys. A: Math. Gen.* **33** 2365.

[55] D. E. Knuth (1998) *The art of computer programming*, volume 3: Sorting and searching (Addison-Wesley, Reading, Mass.), second edition.

[56] G. Gasper and M. Rahman (1990) *Basic hypergeometric series*, volume 35 of *Encyclopedia of mathematics and its applications* (Cambridge University Press, Cambridge).

[57] G. Andrews, R. Askey and R. Roy (1999) *Special functions*, volume 71 of *Encyclopedia of mathematics and its applications* (Cambridge University Press, Cambridge).

[58] N. J. Fine (1988) *Basic hypergeometric series and applications.* Number 27 in Mathematical surveys and monographs (American Mathematical Society, Providence).

[59] A. J. Macfarlane (1989) On q-analogues of the quantum harmonic oscillator and the quantum group $SU(2)_q$. *J. Phys. A: Math. Gen.* **22** 4581.

[60] L. J. Rogers (1894) Second memoir on the expansion of certain infinite products. *Proc. London Math. Soc.* **25** 318.

[61] W. R. Allaway (1980) Some properties of the q-Hermite polynomials. *Canad. J. Math.* **32** 686.

[62] S. Katz, J. L. Lebowitz and H. Spohn (1983) Phase transitions in stationary nonequilibrium states of model lattice systems. *Phys. Rev. B* **28** 1655.

[63] S. Katz, J. L. Lebowitz and H. Spohn (1984) Nonequilibrium steady-states of stochastic lattice gas models of fast ionic conductors. *J. Stat. Phys* **34** 497.

[64] M. B. Salamon (editor) (1979) *Physics of superionic conductors*, volume 15 of *Topics in current physics* (Springer-Verlag, Berlin).

[65] B. Schmittmann and R. K. P. Zia (1995) *Statistical mechanics of driven diffusive systems*, volume 17 of *Phase transitions and critical phenomena* (Academic Press, London).

[66] J. Marro and R. Dickman (1999) *Nonequilibrium phase transitions in lattice models* (Cambridge University Press, Cambridge).

[67] G. M. Schütz, R. Ramaswamy and M. Barma (1996) Pairwise balance and invariant measures for generalized exclusion processes. *J. Phys. A: Math. Gen.* **29** 837.

[68] A. J. Bray (1994) Theory of phase-ordering kinetics. *Adv. Phys.* **43** 357.

[69] C. Domb (1996) *The critical point: a historical introduction to the modern theory of critical phenomena* (Taylor and Francis, London).

[70] H. Spohn (1983) Long-range correlations for stochastic lattice gases in a non-equilibrium steady state. *J. Phys. A: Math. Gen.* **16** 4275.

[71] J. Krug (1991) Boundary-induced phase transitions in driven diffusive systems. *Phys. Rev. Lett.* **67** 1882.

[72] F. Spitzer (1970) Interacting Markov processes. *Adv. in Math.* **5** 246.

[73] C. T. MacDonald, J. H. Gibbs and A. C. Pipkin (1968) Kinetics of biopolymerization on nucleic acid templates. *Biopolymers* **6** 1.

[74] T. Chou and D. Lohse (1999) Entropy-driven pumping in zeolites and biological channels. *Phys. Rev. Lett.* **82** 3552.

[75] D. Chowdhury, L. Santen and A. Schadschneider (2000) Statistical physics of vehicular traffic and some related systems. *Phys. Rep.* **329** 199.

[76] S. Sandow (1994) Partially asymmetric exclusion process with open boundaries. *Phys. Rev. E* **50** 2660.

[77] T. Sasamoto, S. Mori and M. Wadati (1996) One-dimensional asymmetric exclusion model with open boundaries. *J. Phys. Soc. Jpn.* **65** 2000.

[78] B. Derrida, E. Domany and D. Mukamel (1992) An exact solution of a one dimensional asymmetric exclusion model with open boundaries. *J. Stat. Phys.* **69** 667.

[79] B. Derrida and M. R. Evans (1993) Exact correlation functions in an asymmetric exclusion model with open boundaries. *J. Physique I* **3** 311.

[80] B. Derrida, M. R. Evans, V. Hakim and V. Pasquier (1993) Exact solution of a 1D asymmetric exclusion model using a matrix formulation. *J. Phys. A: Math. Gen.* **26** 1493.

[81] T. Sasamoto (1999) One-dimensional partially asymmetric simple exclusion process with open boundaries: orthogonal polynomials approach. *J. Phys. A: Math. Gen.* **32** 7109.

[82] F. H. L. Essler and V. Rittenberg (1996) Representations of the quadratic algebra and partially asymmetric diffusion with open boundaries. *J. Phys. A: Math. Gen.* **29** 3375.

[83] R. Askey (1989) Continuous q-Hermite polynomials when $q > 1$. In *q-series and partitions*, edited by D. Stanton, volume 18 of *IMA volumes in mathematics and its applications*, page 151 (Springer-Verlag, New York).

[84] R. Askey and J. Wilson (1985) Some basic hypergeometric orthogonal polynomials that generalize Jacobi polynomials. *Mem. Am. Math. Soc.* **54** 1.

[85] L. Carlitz (1956) Some polynomials related to theta functions. *Annali di matematica pura ed applicata* **41** 359.

[86] T. Sasamoto (2000) Density profile of the one-dimensional partially asymmetric simple exclusion process with open boundaries. *J. Phys. Soc. Jpn.* **69** 1055.

[87] B. Derrida and K. Mallick (1997) Exact diffusion constant for the one-dimensional partially asymmetric exclusion model. *J. Phys. A: Math. Gen.* **30** 1031.

[88] J. S. Hager, J. Krug, V. Popkov and G. M. Schütz (2001) Minimal current phase and universal boundary layers in driven diffusive systems. *Phys. Rev. E* **63** 056110.

[89] B. Derrida, M. R. Evans and K. Mallick (1995) Exact diffusion constant of a one-dimensional asymmetric exclusion model with open boundaries. *J. Stat. Phys* **79** 833.

[90] S. Janowsky and J. L. Lebowitz (1997) Microscopic models of macroscopic shocks. In *Nonequilibrium statistical mechanics in one dimension*, edited by V. Privman, chapter 13 (Cambridge University Press, Cambridge).

[91] R. Ramaswamy and M. Barma (1987) Transport in random networks in a field: interacting particles. *J. Phys. A: Math. Gen.* **20** 2973.

[92] G. M. Schütz and S. Sandow (1994) On $U_q[SU(2)]$-symmetrical driven diffusion. *Europhys. Lett.* **26** 7.

[93] Y. Kafri, D. Biron, M. R. Evans and D. Mukamel (2000) Slow coarsening in a class of driven systems. *Euro. Phys. J. B* **16** 669.

[94] B. Derrida, M. R. Evans and D. Mukamel (1993) Exact diffusion constant for one-dimensional asymmetric exclusion models. *J. Phys. A: Math. Gen.* **26** 4911.

[95] B. Derrida and J. Lebowitz (1998) Exact large deviation function in the asymmetric exclusion process. *Phys. Rev. Lett.* **80** 209.

[96] B. Derrida and C. Appert (1999) Universal large-deviation function of the Kardar-Parisi-Zhang equation in one dimension. *J. Stat. Phys* **94** 1.

[97] B. Derrida, S. A. Janowsky, J. L. Lebowitz and E. R. Speer (1993) Exact solution of the totally asymmetric simple exclusion process: shock profiles. *J. Stat. Phys.* **73** 813.

[98] B. Derrida and M. R. Evans (1999) Bethe ansatz solution for a defect particle in the asymmetric exclusion process. *J. Phys. A: Math. Gen.* **32** 4833.

[99] M. R. Evans, D. P. Foster, C. Godrèche and D. Mukamel (1995) Asymmetric exclusion model with two species: sponteneous symmetry breaking. *J. Stat. Phys.* **80** 69.

[100] M. Clincy, M. R. Evans and D. Mukamel (2001) Symmetry breaking through a sequence of transitions in a driven diffusive system. *Preprint* cond-mat/0108022.

[101] M. R. Evans, Y. Kafri, H. M. Koduvely and D. Mukamel (1998) Phase separation in one-dimensional driven diffusive systems. *Phys. Rev. Lett.* **80** 425.

[102] M. R. Evans, Y. Kafri, H. M. Koduvely and D. Mukamel (1998) Phase separation and coarsening in one-dimensional driven diffusive systems: Local dynamics leading to long-range Hamiltonians. *Phys. Rev. E* **58** 2764.

[103] K. Mallick, S. Mallick and N. Rajewsky (1999) Exact solution of an exclusion process with three classes of particles and vacancies. *J. Phys. A: Math. Gen.* **32** 8399.

[104] K. Nagel and M. Schreckenberg (1992) A cellular automaton model for freeway traffic. *J. Physique I* **2** 2221.

[105] N. Rajewsky, L. Santen, A. Schadschneider and M. Schreckenberg (1998) The asymmetric exclusion process: comparison of update procedures. *J. Stat. Phys.* **92** 151.

[106] M. R. Evans (1996) Bose-Einstein condensation in disordered exclusion models and relation to traffic flow. *Europhys. Lett.* **36** 13.

[107] M. R. Evans (1997) Exact steady states of disordered hopping particle models with parallel and ordered sequential dynamics. *J. Phys. A: Math. Gen.* **30** 5669.

[108] O. J. O'Loan, M. R. Evans and M. E. Cates (1998) Jamming transition in a homogeneous one-dimensional system: The bus route model. *Phys. Rev. E* **58** 1404.

[109] P. F. Arndt (2000) Yang-Lee theory for a nonequilibrium phase transition. *Phys. Rev. Lett.* **84** 814.

[110] B. Derrida, J. L. Lebowitz and E. R. Speer (2001) Exact large deviation functional for the density profile in a stationary nonequilibrium open system. *Preprint* cond-mat/0105110.

[111] T. E. Harris (1974) Contact interactions on the lattice. *Ann. Prob.* **2** 969.

[112] D. Griffeath (1979) *Additive and cancellative interacting particle systems*, volume 724 of *Lecture notes in mathematics* (Springer-Verlag, Berlin).

[113] H. Hinrichsen (2000) Nonequilibrium critical phenomena and phase transitions into absorbing states. *Adv. Phys.* **49** 815.

[114] I. Jensen and R. Dickman (1993) Time-dependent perturbation-theory for nonequilibrium lattice models. *J. Stat. Phys.* **71** 89.

[115] D. Stauffer and A. Aharony (1991) *Introduction to percolation theory* (Taylor and Francis, London), second edition.

[116] S. R. Broadbent and J. M. Hammersley (1957) Percolation processes I: Crystals and mazes. *Proc. Camb. Phil. Soc.* **53** 629.

[117] E. Domany and W. Kinzel (1984) Equivalence of cellular automata to Ising models and directed percolation. *Phys. Rev. Lett.* **53** 311.

[118] H. Takayasu and A. Y. Tretyakov (1992) Extinction, survival, and dynamical phase transition of branching annihilating random walk. *Phys. Rev. Lett.* **68** 3060.

[119] J. L. Cardy and U. Täuber (1998) Field theory of branching and annihilating random walks. *J. Stat. Phys.* **90** 1.

[120] R. H. Schonmann (1986) The asymmetric contact process. *J. Stat. Phys.* **44** 505.

[121] T. Sweet (1997) The asymmetric contact process at its second critical value. *J. Stat. Phys.* **86** 749.

[122] A. Y. Tretyakov, N. Inui and N. Konno (1997) Phase transition for the one-sided contact process. *J. Phys. Soc. Jap.* **66** 3764.

[123] C. C. Chen, H. Park and M. den Nijs (1999) Active width at a slanted active boundary in directed percolation. *Phys. Rev. E* **60** 2496.

[124] S. M. Dammer, S. R. Dahmen and H. Hinrichsen (2001) Directed percolation and the golden ratio. *Preprint* cond-mat/0106396.

[125] A. J. Guttmann and I. G. Enting (1996) Solvability of some statistical mechanical systems. *Phys. Rev. Lett.* **76** 344.

[126] H. Leipholz (1970) *Stability theory: an introduction to the stability of dynamic systems and rigid bodies* (Academic Press, New York). §1.4.1.

[127] U. Ebert and W. van Saarloos (1998) Universal algebraic relaxation of fronts propagating into an unstable state and implications for moving boundary approximations. *Phys. Rev. Lett.* **80** 1650.

[128] A.-L. Barabási and H. E. Stanley (1995) *Fractal concepts in surface growth* (Cambridge University Press, Cambridge).

[129] T. Halpin-Healy and Y. C. Zhang (1995) Kinetic roughening phonomena, stochastic growth, directed polymers and all that. Aspects of multidisciplinary statistical mechanics. *Phys. Rep.* **254** 215.

[130] J. M. Kim and J. M. Kosterlitz (1989) Growth in a restricted solid-on-solid model. *Phys. Rev. Lett.* **62** 2289.

[131] P. Meakin, P. Ramanlal, L. M. Sander and R. C. Ball (1986) Ballistic deposition on surfaces. *Phys. Rev. A* **34** 5091.

[132] M. Plische, Z. Rácz and D. Liu (1987) Time-reversal invariance and universality of two-dimensional growth models. *Phys. Rev. B* **35** 3485.

[133] F. Family and T. Vicsek (1985) Scaling of the active zone in the Eden process on percolation networks and the ballistic deposition model. *J. Phys. A: Math. Gen.* **18** L75.

[134] R. Baiod, D. Kessler, P. Ramanlal, L. Sander and R. Savit (1988) Dynamical scaling of the surface of finite-density ballistic aggregation. *Phys. Rev. A* **38** 3672.

[135] D. Y. K. Ko and F. Seno (1994) Simulations of deposition growth in various dimensions: the possible importance of overhangs. *Phys. Rev. E* **50** R1741.

[136] T. Ala-Nissila, T. Hjelt, J. M. Kosterlitz and O. Venölöinen (1993) Scaling exponents for kinetic roughening in higher dimensions. *J. Stat. Phys.* **72** 207.

[137] J. Krug (1997) Origins of scale invariance in growth processes. *Adv. Phys.* **46** 139.

[138] L. Bertini and G. Giacomin (1997) Stochastic Burgers and KPZ equations from particle systems. *Commun. Math. Phys.* **183** 571.

[139] M. A. Moore. Private communication.

[140] J. Riordan, C. R. Doering and D. ben Avraham (1995) Fluctuations and stability of Fisher waves. *Phys. Rev. Lett.* **75** 565.

[141] G. Tripathy and W. van Saarloos (2000) Fluctuation and relaxation properties of pulled fronts: a scenario for nonstandard Kardar-Parisi-Zhang scaling. *Phys. Rev. Lett.* **85** 3556.

[142] R. A. Fisher (1937) The wave of advance of advantageous genes. *Annals of Eugenics* **7** 355.

[143] G. Tripathy, A. Rocco, J. Casademunt and W. van Saarloos (2001) Univerality class of fluctuating pulled fronts. *Phys. Rev. Lett.* **86** 5215.

[144] J. L. Cardy and A. D. Bruce. Private communication.

[145] J. Kertész and D. E. Wolf (1988) Noise reduction in Eden models II: Surface structure and intrinsic widths. *J. Phys. A: Math. Gen.* **21** 747.