STATISTICAL MECHANICS AND COMBINATORICS OF SOME DISCRETE LATTICE MODELS

BY ARVIND AYYER

A dissertation submitted to the

Graduate School—New Brunswick

Rutgers, The State University of New Jersey

in partial fulfillment of the requirements

for the degree of

Doctor of Philosophy

Graduate Program in Physics

Written under the direction of

Professor Joel L. Lebowitz

and approved by

_____

_____

_____

_____

_____

New Brunswick, New Jersey

October, 2008

ABSTRACT OF THE DISSERTATION

Statistical Mechanics and Combinatorics of Some Discrete Lattice Models

by Arvind Ayyer

Dissertation Director: Professor Joel L. Lebowitz

Many problems in statistical physics involve enumeration of certain objects. In this thesis, we apply ideas from combinatorics and statistical physics to understand three different lattice models.

I We investigate the structure of the nonequilibrium stationary state (NESS) of a system of first and second class particles on L sites of a one-dimensional lattice in contact with first class particle reservoirs at the boundary sites and second class particles constrained to lie the system. The internal dynamics are described by the usual totally asymmetric exclusion process (TASEP) with second class particles. We show in a conceptually simple way how pinned and unpinned (fat) shocks determine the general structure of the phase diagram. We also point out some unexpected features in the microscopic structure of the NESS both for finite L and in the limit $L \to \infty$. In the latter case the local distribution of second class particles is given by an equilibrium pressure ensemble with a pair potential between neighboring particles which grows logarithmically with distance.

II We model a long linear polymer constrained between two plates as a walk on a two-dimensional lattice constrained to lie between two lines, $x = y$ and $x = y + w$, which interacts with these lines via contact parameters s and t. The atomic steps of the walk can be taken to be from an arbitrary but fixed set S with the only condition being that the first coordinate of every element in S is strictly positive. For any such S and any w, we prescribe general algorithms (fully implemented in Maple) for the automated calculation of several mathematical and physical quantities of interest.

III Ferrers (or Young) diagrams are very classical objects in representation theory, whose half-perimeter generating function of Ferrers diagrams is a straightforward rational function. We construct two new classes of Ferrers diagrams, which we call wicketed and gated Ferrers diagrams, which have internal voids in the shape of Ferrers diagrams, and calculate their half-perimeter generating functions, one of which is closely related to the generating function of the Catalan numbers, using a more abstract version of the usual transfer matrix method.

Acknowledgements

First and foremost, I would like to thank my main academic advisors, Prof. Joel Lebowitz and Prof. Doron Zeilberger, without whose guidance this work would be hopelessly far from complete. Their technical expertise, helpfulness, and sense of fun in what they do has been a tremendous motivating factor in trying to better myself. I hope to imbibe these attitudes in my professional life. I would also like to mention Prof. Eugene Speer, whom I consider a "godadvisor", who has gone to great lengths to advise me both academically and nonacademically on a variety of issues. Their attitudes towards the pursuit of science and their differing philosophies have only served to remind me of the value of diversity in science.

I would like to thank the physics department for five years worth of finanicial support. I am grateful to my committee members for their suggestions. Both the graduate directors in my time, Prof. Ted Williams and Prof. Ron Ransome have been very competent and have made the miles of red tape seem like meters. The staff both in the Physics department and in Prof. Lebowitz's office have been very helpful also. I would especially like to mention Yael Goldberg whose passing away was a sad loss.

I am grateful to my coworkers on other projects, Tewodros Amdeberhan, Mikko Stenlund and Carlangelo Liverani, each of whom has helped me understand a different small problem a little better.

The atmosphere at both the physics and mathematics departments here at Rutgers have been very vibrant and my discussions with many people have reinforced the idea that there are infinitely many interesting problems out there.

It is not customary to thank friends for their support and help, but I cannot do

without expressing my gratitude to Diwakar, Vaishnavi, Ashu, Alivia, Kasturi, Shivani, Sonal, Manish, Yesha, Avital, Edinah, Jacob and Madhavi for the pleasure of their company, extended bull sessions and endless cups of tea.

Lastly but not leastly I have to thank Divya for her continuing support and encouragement.

Dedication

To my biological parents

हर चीज के लिये धन्यवाद

and my academic ones

תודה רבה על כל מה שעשיתם עבורי

Thanks for everything!

Table of Contents

List of Figures

List of Tables

Chapter 1

Introduction

Equilibrium statistical mechanics may be considered a closed chapter in the sense that we have a clear formalism and a good understanding of the calculations needed to obtain results about specific systems. There are of course limitations on our abilities to actually carry out computations.

We do not have any comparable general formalism for nonequilibrium systems. In fact, we do not have any example of a *deterministic* microscopic model for which one can compute transport coefficient or other nonequilibrium properties. One can do better with stochastic models. For example, a number of results are known about the hydrodynamic limit of the simple symmetric exclusion process. This yields the diffusion equation and one knows how to calculate the diffusion constant explicitly. The technical tools needed for stochastic models are much simpler than those needed for the deterministic case.

The seminal papers of Zeilberger and coworkers [WZ90b, WZ90a, Zei91, Zei90, PWZ96] on the automated proving of combinatorial identities showed that computers not only helped in understanding large systems by simulations, but also in calculations for small systems which were not feasible by pen-and-paper (computer experimentation). Independently Derrida and coworkers, and other groups, used this idea to solve the one-dimensional totally asymmetric exclusion process (TASEP) [DDM92, DEHP93, DJLS93, SD93]. The TASEP can be considered to be the simplest nontrivial solvable nonequilibrium problem, very much like the Ising model in equilibrium statistical mechanics. These ideas have been applied to model problems in polymer studies and

biology, and have been used extensively to understand nonequilibrium phenomena.

In this thesis, we investigate a variety of nonequilibrium lattice models. In the first chapter, we study a new TASEP on L sites on a line consisting of two kinds of particles, called first and second class, as well as holes, where the dynamics in the bulk are such that first class particles hop to the right with rate 1 exchanging with both second class particles and holes, whereas second class particles only hop to the right with rate 1 exchanging with holes. An example of such a TASEP is shown in Figure 1.1. In other words, first class particles move to the right, holes to the left, and second class particles can move in both directions.

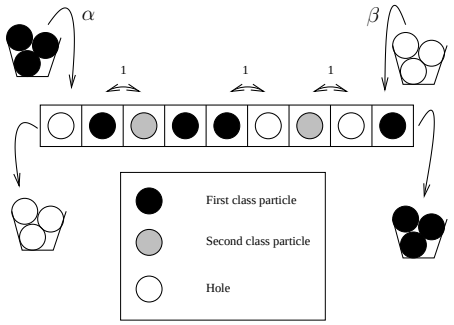


Figure 1.1: An example of the two species TASEP with $L = 9$ and $n = 2$. The locations where an exchange can occur are marked along with the rate of the exchange.

What makes this problem solvable exactly is the dynamics at the boundary. Second class particles cannot enter or leave the system, and thus, their number n is fixed in the system. First class particles can enter on the left at rate α kicking out a hole, and can leave on the right with rate β, being replaced by a hole. Therefore the phase diagram depends on three parameters α, β and $\gamma = n/L$, the fractional number of second class particles in the system.

We used computer experimentation with small systems to guess a closed form expression for the steady state probabilities of the Markov chain (which, unknown to us

at that time, was already known to Arita [Ari06b]) as well as to conjecture some other properties of the system which hold for all finite L and n. We also performed simulations for various values of α and β to guess the phase diagram of the system. We then used the matrix method of Derrida et al. [DEHP93] to prove both the exact results true for any values of L and n as well as the asymptotic results true in the limit of $L, n \to \infty$, $n/L \to \gamma$, which helps determine the phase diagram.

In the second chapter, we are motivated by the problem of polymers held between two fixed parallel plates but consider rather the one dimensional analogue — namely a polymer constrained to lie between two parallel lines, which we take as a matter of convention to be the lines $x = y$ and $x = y + w$, where w is a fixed integer. We model the problem combinatorially as a walk in the two dimensional square lattice which takes steps from a certain set S. A real polymer can be modelled as a walk whose length goes to infinity simultaneously as the lattice spacing approaching zero. We develop a package called *POLYMER* in the computer algebra system Maple which, given the set S satisfying a certain directedness condition and the width w, calculates many important physical quantities.

These quantities include the generating function

$$\phi(z; t, s) = \sum_{\substack{\text{all possible walks} \\ \text{ending on the line } x=y}} z^m t^n s^p, \tag{1.0.1}$$

where

$$m = \text{the length of the walk},$$

$$n = \text{the number of times the walk touches the line } x = y,$$

$$p = \text{the number of times the walk touches the line } x = y + w,$$

the free energy of the polymer,

$$\kappa_w = \lim_{n \to \infty} \frac{1}{n} \log c_w(n), \tag{1.0.2}$$

where $c_w(n)$ is the number of walks from $(0,0)$ to (n,n), and the force exerted by the polymer on the walls,

$$F(w) = \kappa_{w+1} - \kappa_w. \tag{1.0.3}$$

In the third chapter, we consider a generalization of Ferrers diagrams like that in Figure 1.2 (which are called Young diagrams in the physics literature), which can be considered to be subsets of the two dimensional lattice.

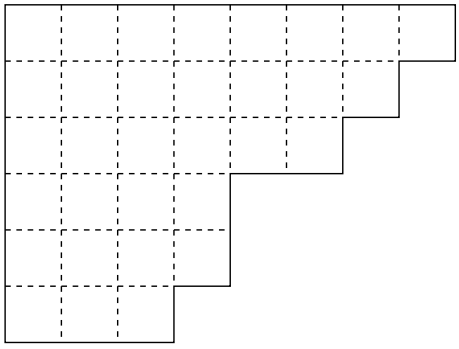


Figure 1.2: A Ferrers diagram with six rows

The generalizations we consider are motivated by the study of self-avoiding polygons and are Ferrers diagrams with internal voids also in the shape of Ferrers diagrams. There are two different generalizations based on where the void lies. We prove a conjectured formula for the half-perimeter generating function of these new Ferrers diagrams using an extension of the well-known transfer matrix method in statistical mechanics. This extension involves constructing a transfer matrix whose entries are operators themselves. Once the formalism is in place, we show that the proof is routine and only involves verification of the equality of algebraic expressions.

Chapter 2

TASEP with Semi-Permeable Boundaries

2.1 Introduction

One of the widely studied stochastic models of nonequilibrium statistical mechanics is the open one-dimensional totally asymmetric exclusion process (TASEP), see [Kru91, DDM92, SD93, DEHP93] as well as the recent review [BE07]. The model consists of a lattice of L sites, each of which is either occupied by a particle or is vacant. Particles try to hop to their right at rate 1. The hop succeeds when the site to the right is empty, otherwise nothing happens. If site 1 is empty, particles enter the site at rate α from a reservoir, and if site L is occupied, the particle exits at rate β to another reservoir.

A configuration of the open TASEP is specified by a binary vector of size L whose ith component is 1(0) if site i is occupied (empty). While analyzing this system, Derrida et al. [DEHP93] found that one could write down the steady state probability of any configuration as a matrix expectation value involving only two kinds of matrices — X_1 for particles and X_0 for vacancies:

$$p([\tau_1 \dots \tau_L]) = \frac{\langle W|X_{\tau_1} \dots X_{\tau_L}|V\rangle}{Z}, \tag{2.1.1}$$

where Z is a normalization factor. For a system, whenever one can write down such a formula (2.1.1) for fixed matrices X_i, the probabilities are said to satisfy a matrix product ansatz. For this model, the two matrices X_0 and X_1 and the vectors $|V\rangle$ and $\langle W|$ satisfy the condition

$$X_1X_0 = X_1 + X_0 \qquad X_1|W\rangle = \frac{1}{\beta}|W\rangle \qquad \langle V|X_0 = \frac{1}{\alpha}\langle V|. \tag{2.1.2}$$

The matrices and vectors can be represented in an infinite dimensional vector space. This property (2.1.1), was used in [DEHP93] to give an explicit formula for the densities and correlation functions and to determine the phase diagram of the system Figure 2.1. This gave a lot of impetus to building other models which could be solved by such an ansatz (see in particular [BE07] for a recent review of matrix methods for the TASEP).

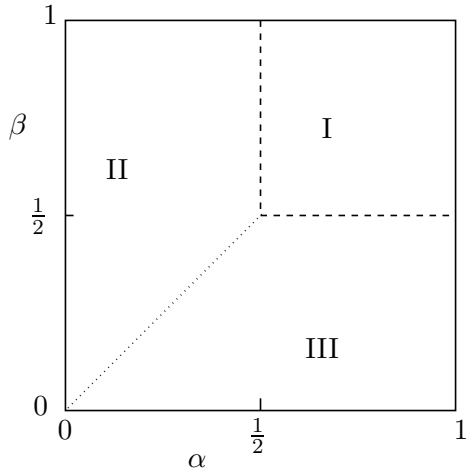


Figure 2.1: The phase diagram for the open TASEP. Region I is the maximal current phase. Regions II and III are the low and high density phases respectively. The boundary between II and III represents the shock line.

The density at site i is given by the steady state probability that site i is occupied by a particle. To study the density, we consider the large L limit and rescale the system to the interval $[0, 1]$. When $\alpha = \beta < 1/2$, [DEHP93] found that the density profile was linear along the system, with value α on the left side and value $1 - \alpha$ on the right side. Typically, such a linear profile is obtained by averaging over different snapshots of density profiles, and is associated with the occurence of *shocks*. These are configurations where the system exhibits a large change in density over a very small region; this change would look like a discontinuity for large system sizes. While studying the same system on the infinite line (and hence without reservoirs), [ABL88] added a single second class particle to the system, a notion first invented in [Spi70]. In the new system, both first and second class particles try to hop to the right at rate 1. For second class particles

the hop succeeds only if the site to the right is empty. For first class particles the hop succeeds if the latter site is empty or is occupied by the second class particle. In this system, when one starts with a distribution with density ρ_0 to the left of the origin and $\rho_1 > \rho_0$ to the right of the origin, one finds that a shock occurs in the system and the second class particle tends to be located near the shock. One natural definition of the location of the shock is exactly the position of the second-class particle.

While studying a system of first-class particles, second-class particles and vacancies on a ring (lattice of L sites with periodic boundary conditions), [DJLS93] showed that the matrix ansatz holds for the system, with three different matrices which satisfy equations that generalize (2.1.2). In this work, we study an open version of the same system in which only first-class particles can enter and leave the system. They enter with rate α only if the first site is empty and exit with rate β just as in the first TASEP model. Second-class particles are confined within the system. The model was first studied by Arita [Ari06a, Ari06b] using a matrix ansatz, and he obtained the density profiles in the bulk and leading boundary corrections to the densities. We analyze the system in detail and explain the density profiles by a structure we call the *fat shock*. We also show that the second-class particles by themselves form an equilibrium system which we study using a pressure (isobaric) ensemble.

2.2 The Model and Summary of Results

The model is defined on a subset of the one dimensional lattice $\mathbb{Z}$ consisting of L sites. Each site i, $i = 1, \ldots, L$, may be occupied by a first class particle, occupied by a second class particle, or vacant; vacant sites are also referred to as holes, and first class particles simply as particles. We shall let these three possible states correspond to the values 1, 2, and 0, respectively, of a random variable τ_i; we also introduce the indicator random variables $\eta_a(i)$, $a = 0, 1, 2$, such that $\eta_a(i) = 1$ if $\tau_i = a$ and $\eta_a(i) = 0$ otherwise.

The internal (bulk) dynamics of the system are given by the usual rules for the

TASEP with second class particles [DJLS93]. The occupation variable τ_i at site $i, i = 1, \ldots, L-1$, attempts when $\tau_i = 1$ or 2 to exchange at rate 1 with τ_{i+1}; when $\tau_i = 1$ the exchange succeeds if $\tau_{i+1} = 0$ or 2, while for $\tau_i = 2$ it only succeeds if $\tau_{i+1} = 0$. In other words, a first class particle at site i jumps to the right by exchanging with either a hole or second class particle at site $i+1$, while a second class particle can only jump if the site on its right is empty. At site $i = 1$, first class particles enter the system at rate α provided that site 1 is vacant ($\tau_1 = 0$); at site $i = L$, first class particles leave the system at rate β provided that site L is occupied by a first class particle ($\tau_L = 1$). Second class particles are thus trapped inside the system. (A similar system, but with a different form of semipermeable boundary, was considered in [KJS03]). An equivalent system is obtained by interchanging first class particles with holes, left with right, and α with β, and this symmetry will be reflected in the structure of the NESS. The latter will be determined by the parameters α, β and the density $\gamma = n/L$ of second class particles, where n is the number of second class particles in the system.

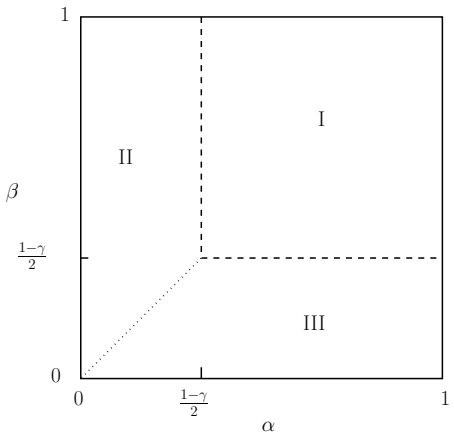


Figure 2.2: The cross section of the phase diagram at a fixed γ.

The phase diagram of this system in the limit $L \to \infty$ is given in Figure 2.2 [Ari06b]. The diagram is determined by the distinct formulas for the (first class) particle current

J_1 in the different regions:

$$J_1 = \begin{cases} \dfrac{1-\gamma^2}{4}, & \text{for } \alpha, \beta \geq \alpha_c \ \text{ (region I)}, \\ \alpha(1-\alpha), & \text{for } \alpha < \alpha_c, \alpha < \beta \ \text{ (region II)}, \\ \beta(1-\beta), & \text{for } \beta < \alpha_c, \beta < \alpha \ \text{ (region III)}; \end{cases} \tag{2.2.1}$$

here the critical value α_c of α and β is

$$\alpha_c = \frac{1-\gamma}{2}. \tag{2.2.2}$$

The current J_2 of the (trapped) second class particles must vanish and the current J_0 of holes satisfies $J_0 = -J_1$. We can think of J_1 as an order parameter which is continuous but not smooth across all the boundaries.

The phase diagram is similar to that of the open one component TASEP [DEHP93, SD93] and indeed in the limit $\gamma \to 0$ reduces to it; moreover, in regions II and III the current is independent of γ and takes the same values as in the one component case, although the size of these regions shrinks as γ increases. As we discuss in Section 2.7, however, there will be residual differences between the local microscopic states of the one species model and $\gamma \to 0$ limit of the two species model; in particular, there remain an infinite number of second class particles near one or both boundaries of the two species system. Note also that there is a discontinuity, equal to γ, in the derivative of J_1 with respect to α (β) on the I/II (I/III) boundary; one might say that the order of the phase transition in J_1 when $\gamma \neq 0$ differs from that when $\gamma = 0$.

The macroscopic density profiles $\rho_a(x)$ in the NESS, $a = 0, 1, 2$, defined by

$$\rho_a(x) = \lim_{L\to\infty, n/L\to\gamma, i/L\to x} \langle \eta_a(i) \rangle, \qquad 0 \leq x \leq 1, \tag{2.2.3}$$

with $\langle \cdot \rangle$ the expected value in the NESS, have been computed in [Ari06b]; the results are summarized in Table 2.1 (but see Remark 2.2.1 below). Knowing any two of these

densities determines the third, via $\sum_a \rho_a(x) = 1$. In fact, knowing $\rho_1(x)$ for all x and all values of α and β determines $\rho_0(x)$, from the particle-hole symmetry, and hence all profiles, but for clarity we give in Table 2.1 both $\rho_1(x)$ and $\rho_0(x)$. In region II the system divides itself into two parts, $x < x_0$ and $x > x_0$, with different formulas for $\rho_0(x)$; similarly in region III there are different formulas for $\rho_1(x)$ for $x < x_1$ and $x > x_1$. Here

$$\begin{aligned} x_0 &= 1 - \frac{\gamma}{1-2\alpha}, \qquad \alpha \le \beta, \quad \alpha < \alpha_c; \\ x_1 &= \frac{\gamma}{1-2\beta}, \qquad \beta \le \alpha, \quad \beta < \alpha_c. \end{aligned} \tag{2.2.4}$$

On the II/III boundary $\alpha = \beta < \alpha_c$, the *shock line*, the profiles include linear regions:

$$\begin{aligned} \rho_0(x) &= \begin{cases} \dfrac{x_0 - x}{x_0}(1-\alpha) + \dfrac{x}{x_0}\,\alpha, & 0 \le x \le x_0, \\ \alpha, & x_0 \le x \le 1 \end{cases} \\ \rho_1(x) &= \begin{cases} \alpha, & 0 \le x \le x_1, \\ \dfrac{1-x}{1-x_1}\,\alpha + \dfrac{x-x_1}{1-x_1}(1-\alpha), & x_1 \le x \le 1. \end{cases} \end{aligned} \tag{2.2.5}$$

These arise from averaging over the position of a shock, as in the one species TASEP; further discussion is given below.

Table 2.1: Density profiles in different regions of the phase plane. Note that x_0 is defined only in region II and on its boundaries, and x_1 only in region III and on its boundaries.

Region	$\rho_1(x)$		$\rho_0(x)$	
I	α_c		α_c	
I/II boundary	α_c		α_c	
I/III boundary	α_c		α_c	
	$x < x_1$	$x > x_1$	$x < x_0$	$x > x_0$
II	α		$1-\alpha$	α
III	β	$1-\beta$	β	
II/III boundary (Shock Line)	$\alpha\,(=\beta)$	linear	linear	$\alpha\,(=\beta)$

Remark 2.2.1 *(a) The density values $\rho_a(x)$ at the boundaries $x = 0, 1$ and at the fixed shocks $x = x_0, x_1$ may depend the way the limit* (2.2.3) *is taken. The boundary cases*

$x = 0, 1$ were discussed in [Ari06b] except on the I/II and I/III boundaries. We discuss the limits at x_0, x_1 in Section 2.6 and 2.7; this gives some further information about limits at the boundaries since $x_0 = 0$ and $x_1 = 1$ on the I/II and I/III boundaries, respectively.

(b) In the one-component model the phase plane regions corresponding to I, II, and III are called the maximum current, low density, and high density regions, respectively. We do not adopt that terminology here since in region III the particle density is low for $x < x_1$.

Some typical profiles, obtained from simulations, are shown in Figure 2.3.

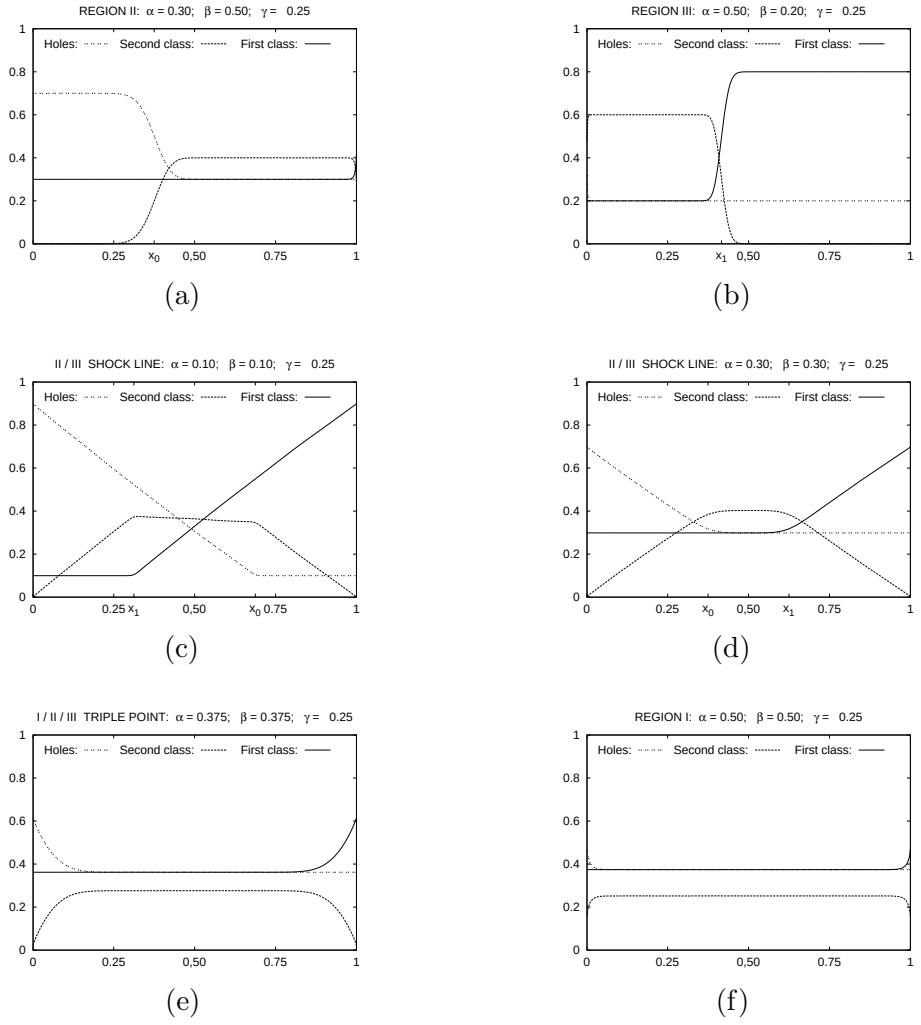


Figure 2.3: Density profiles in a system with $L = 1000$.

Since these are from a finite system, they do not coincide perfectly with the description in Table 2.1: there are boundary effects, and the density transitions in regions II and III, at x_1 and x_0 respectively, have nonzero width of order $\sqrt{L}$. This is related to the nonuniqueness of the limit (2.2.3) at these points, as mentioned above, and is discussed in Section 2.6.

We now give an intuitive discussion of some of the phenomena that give rise to these profiles. Consider a macroscopically uniform portion of our system in the limit $L \to \infty$, with densities of holes, particles and second class particles denoted by ρ_0, ρ_1 and ρ_2, respectively, where $\rho_0 + \rho_1 + \rho_2 = 1$.

In such a region the measure will be the known translation invariant measure, with these densities, for the two species TASEP (see [DJLS93, Spe94, FFK94] and the discussion in Section 2.9); in this measure (which is not a product measure) the first class particles considered separately, and the holes considered separately, are distributed according to product measures, so that $J_1 = \rho_1(1-\rho_1)$ and $J_0 = -\rho_0(1-\rho_0)$. Thus from $J_0 = -J_1$ it follows that in any uniform stretch of the NESS either

$$\rho_1 = \rho_0 = (1-\rho_2)/2 \qquad \text{or} \qquad \rho_2 = 0,\ \rho_1 = 1-\rho_0. \tag{2.2.6}$$

This fact, which may be seen in the results of [Ari06b], is key to understanding the gross structure of the densities in different regions of the phase diagram.

These density profiles differ from those the single-species open TASEP in two notable ways: in regions II and III the density profiles have a point of discontinuity, and on the II/III boundary the linear region occupies only part of the system. These and other properties can be understood in terms of the occurrence of a *fat shock*: a macroscopically uniform interval which contains all the second-class particles, and thus conforms to the first alternative in (2.2.6). To see how this shock arises, recall that if one either identifies first and second class particles by coloring holes black and both kinds of particles white, or else identifies second-class particles and holes by coloring these species red and first-class particles blue, then the black/white particles, as well as the red/blue particles,

form standard two species TASEPs in the bulk. The dynamics at the boundaries is different, since some white "particles" or red "holes" will be trapped in the system. A careful justification of the conclusions below is given in Sections 2.5 and 2.6.

Consider now the behavior of the system on the boundary of regions II and III (the shock line). Then by previous analysis, see e.g. [Lig99], one knows that a typical profile for the one species model contains a shock between a region of density α on the left and $1-\alpha$ on the right; the shock position has mean velocity zero and its (fluctuating) position is uniformly distributed over the system. We see this same behavior for both the black/white and red/blue systems described above, with the black/white shock necessarily located to the left of the red/blue one. The *typical* profile at any given time looks on the macroscopic scale like Figure 2.4, where the convention is that at any point x the height of the region labeled with particle type a is $\rho_a(x)$.

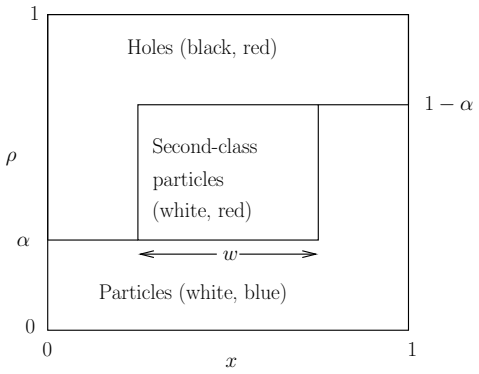


Figure 2.4: Shock interpretation at $\alpha = \beta < \alpha_c$. Densities $\rho_a(x)$ are plotted against x, with the convention that at the height at x of the region labeled with particle type a is $\rho_a(x)$. The fat shock may in fact be located anywhere in the system.

Clearly both shock fronts have mean velocity zero, and since the total number of second-class particles is γL, the macroscopic width w of the fat shock must satisfy $w = \gamma/(1-2\alpha)$. This forces the two shock fronts to move (i.e., fluctuate) in collusion so as to keep the macroscopic shock width fixed; we expect this fluctuation, as for the shock

in the single component TASEP on its shock line, to be on a diffusive time scale growing as L^2. The density profiles $\rho_a(x)$ arise as averages over the shock position, and this gives rise to the linear profiles (2.2.5); in contrast to the situation in the one species case, however, here they occupy only part of the system because the shock can fluctuate only over an interval of width $1-w$. The shock fluctuation is also reflected in the structure of the local measures obtained in the limit $L \to \infty$, which are superpositions of states with different densities (see Section 2.6). The critical value of α occurs when the fat shock fills the system, i.e., when $w = 1$, from which we regain (2.2.2).

The situation in regions II and III is similar. The fat shock width is in general

$$w(\alpha, \beta, \gamma) = \frac{\gamma}{1 - 2\alpha \wedge \beta}, \tag{2.2.7}$$

where $\alpha \wedge \beta = \min\{\alpha, \beta\}$; in region II the shock is pinned to the right boundary, and in region III is pinned to the left boundary. Since the shock is fixed it gives rise to discontinuities in the density profiles; see Figure 2.3 as well as the discussion of a related model, where similar behavior occurs, in Section 2.9 (Figure 2.5). There is no corresponding discontinuity in the single-species TASEP ($\gamma = 0$) because in that case there is a single shock of zero macroscopic width. In region I, w as given in (2.2.7) is greater than 1 and the fat shock fills the system; the density profiles are uniform and conform to the first alternative of (2.2.6).

The outline of the rest of the chapter is as follows. In Section 2.3 we discuss the matrix method for this system. We use a different representation of the matrices from that of [Ari06b], which makes it easier to prove certain features of the NESS discussed later. In Section 2.4 we show that the marginal distribution induced by the NESS on particles in the first n states of the system, and on holes in the last n, is exchangeable, i.e., that the probability of finding r first class particles (holes) on some specified set of r sites among the first (last) n is independent of the choice of sites. In Section 2.5 we establish the fat shock picture described above. In Section 2.6 we determine the local measures, in the bulk, for the infinite volume limit of the system, and in Section 2.7 we

consider the local measures near the boundaries, focusing on a Bernoulli property which is a consequence of the exchangeability established in Section 2.4. In Section 2.8 we show that the second class particles form an equilibrium system, most simply described by a pressure ensemble. This is related, in our case, to the fact that the current of second class particles is zero. For similar situations, see [RPS03, BDSG$^+$08].

In Section 2.9 we make some concluding remarks and, in particular, describe some closely related models. One such model is a generalization of the standard "defect particle" model; another describes a system of first class particles, second class particles, and holes on a ring with one semi-permeable bond which second class particles cannot cross. Several more technical remarks are recorded in the appendices.

2.3 The matrix ansatz

The stochastic system described in Section 2.2 is ergodic in finite volume L and thus there exists a unique invariant measure $\mu_{L,n}^{\alpha,\beta}$ on the configuration space

$$Y_{L,n} \equiv \{\, (\tau_1, \ldots, \tau_L) \mid \tau_i = 0, 1, 2;\ \tau_i = 2 \text{ for } n \text{ values of } i \,\}, \tag{2.3.1}$$

where from now on we will assume that $0 < n < L$. This measure may be obtained from a matrix ansatz [Ari06b], combining the matrix algebra of [DJLS93] (which discussed the system with the same constituents as in the current work, but on a ring) with the treatment of the one species open system via matrix-elements from [DEHP93]. One introduces matrices X_0, X_1, and X_2 and vectors $|V_\beta\rangle$ and $\langle W_\alpha|$ which satisfy

$$X_1 X_0 = X_1 + X_0, \qquad X_1 X_2 = X_2, \qquad X_2 X_0 = X_2, \tag{2.3.2}$$

and

$$\langle W_\alpha | X_0 = \frac{1}{\alpha} \langle W_\alpha |, \qquad X_1 |V_\beta\rangle = \frac{1}{\beta} |V_\beta\rangle. \tag{2.3.3}$$

Then for a configuration $\tau = (\tau_1, \cdots, \tau_L) \in Y_{L,n}$ the probability of τ in the invariant measure is

$$\langle \tau \rangle_{\mu^{\alpha,\beta}_{L,n}} = Z^{\alpha,\beta}(L,n)^{-1} \langle W_\alpha | X_{\tau_1} \cdots X_{\tau_L} | V_\beta \rangle, \tag{2.3.4}$$

where $Z^{\alpha,\beta}(L,n)$ is the normalization factor

$$Z^{\alpha,\beta}(L,n) = \sum_{\tau \in Y_{L,n}} \langle W_\alpha | X_{\tau_1} \cdots X_{\tau_L} | V_\beta \rangle, \tag{2.3.5}$$

which, with a slight misuse of the nomenclature of equilibrium statistical mechanics, we call the *partition function*. We will frequently omit superscripts such as α, β in (2.3.5) when no confusion can arise.

We will work throughout in a realization of (2.3.2)–(2.3.3), different from that of [Ari06b], for which the matrices and vectors have the further properties

$$X_2 = X_1 X_0 - X_0 X_1 = |V_1\rangle\langle W_1|, \qquad X_2 |V_\beta\rangle = |V_1\rangle, \qquad \langle W_\alpha | X_2 = \langle W_1 |, \tag{2.3.6}$$

$$\langle W_\alpha | V_1 \rangle = \langle W_1 | V_\beta \rangle = 1 \qquad \text{for all } \alpha, \beta, \tag{2.3.7}$$

where $|V_1\rangle$ and $\langle W_1|$ are vectors satisfying (2.3.3). Note that X_2 is then a one-dimensional projection operator. The realization is given in Appendix A.1, but we will need no consequences beyond (2.3.6) and (2.3.7). Because of (2.3.7) we make the convention that $Z^{\alpha,1}(0,0) = Z^{1,\beta}(0,0) = 1$.

Remark 2.3.1 *The nature of X_2 in this representation shows that certain distributions obtained from* (2.3.4) *factorize. Let $Q_1, \ldots, Q_n$ denote the (random) positions of the second class particles in the system (note that these can be ordered once and for all). Then the probability that the j_1^{th}, j_2^{th}, …, j_m^{th} second class particles are located on sites $q_{j_1}, \ldots, q_{j_m}$ is*

$$\begin{aligned} \mu^{\alpha,\beta}_{L,n}(Q_{j_i} = q_{j_i}, i = 1, \ldots, m) &= Z^{\alpha,\beta}(L,n)^{-1} Z^{\alpha,1}(q_{j_1} - 1, j_1 - 1) \\ &\quad \times \prod_{i=2}^{m} Z^{1,1}(q_{j_i} - q_{j_{i-1}} - 1, j_i - j_{i-1} - 1)\; Z^{1,\beta}(L - q_{j_m}, n - j_m). \end{aligned} \tag{2.3.8}$$

Moreover, if we condition on the event that $Q_j = q$, i.e., that the j^{th} second class particle is located at site q, then the conditional measure is a product of the measures associated with the sites before and after j, so that if τ is a configuration consistent with this event then

$$\mu_{L,n}^{\alpha,\beta}(\tau_1, \ldots, \tau_L \mid Q_j = q) = \mu_{q-1,j-1}^{\alpha,1}(\tau_1, \ldots, \tau_{q-1}) \mu_{L-q,n-j}^{1,\beta}(\tau_{q+1}, \ldots, \tau_L). \quad (2.3.9)$$

A factorization property of this type is also known for the translation invariant measures for the two species TASEP [DJLS93, FFK94]. Similar expressions are easily obtained when conditioning on the presence of several specified second class particles at specified sites. We will use (2.3.8) *and* (2.3.9) *in Sections 2.5 and 2.6, when we discuss the fat shock and describe local measures in the NESS.*

2.4 Exchangeability of Measures

In this section we demonstrate a remarkable property of the finite-volume NESS with n second class particles: the *exchangeability* [Fel71] of the measure on first class particles within the first n sites, or equivalently on holes in the last n sites. Specifically, this means that for any $r \leq n$ the probability of finding first class particles on the r sites $1 \leq i_1 < i_2 < \cdots < i_r \leq n$ depends only on r, i.e., is independent of the choice of positions $i_1, i_2, \ldots, i_r$. When $r = 1$ this is implicit in (38) of [Ari06b], although it is not emphasized. As a consequence of the ideas of the proof we will also obtain, for any i, j with $i, j \geq 1$ and $i + j - 1 \leq L$, the probability of finding a block of j consecutive first class particles starting at site i; this generalizes the density formula of [Ari06b], which corresponds to $j = 1$.

The key quantity for our arguments is the probability of finding (first class) particles at sites $i_1, \ldots, i_{r-1}$ together with a block of j particles starting at site i_r, where $r \geq 1$, $i_1 < \cdots < i_{r-1}$, and $i_r > i_{r-1} + 1$; we allow $j = 0$, with the interpretation that in this case there is no restriction on what happens at site i_r or any succeeding sites. Thus the

probability in question is $Z(L,n)^{-1}E_r(L,n;i_1,\ldots,i_r;j)$, where E_r is a sum of weights for certain configurations $\tau \in Y_{L,n}$:

$$E_r(L,n;i_1,\ldots,i_r;j) = \sum_{\substack{\tau_{i_1}=\cdots=\tau_{i_{r-1}}=1 \text{ if } r\geq 2 \\ \tau_{i_r}=\tau_{i_r+1}=\cdots=\tau_{i_r+j-1}=1 \text{ if } j\geq 1}} \langle W_\alpha | X_{\tau_1}\ldots X_{\tau_L} | V_\beta\rangle \tag{2.4.1}$$

In (2.4.1) we must have $i_r+j-1 \leq L$, since there are only L sites, and $r+j-1 \leq L-n$, since there are n second class particles. For certain parts of the analysis we will have to consider separately two cases, in which these two inequalities respectively provide the effective bounds on j:

Case 1: $i_r \geq n+r$, so that $0 \leq j \leq L-i_r+1$;

Case 2: $i_r \leq n+r-1$, so that $0 \leq j \leq L-n-r+1$.

We will analyze the E_r using a simple recursion:

$$E_r(L,n;i_1,\ldots,i_r;j) = E_r(L,n;i_1,\ldots,i_r;j+1) + E_r(L-1,n;i_1,\ldots,i_r;j-1). \tag{2.4.2}$$

This holds whenever all terms are defined, which requires that j be positive and satisfy $j \leq L-i_r$ in Case 1 and $j \leq L-n-r$ in Case 2. To derive (2.4.2) we consider the value of τ_{i_r+j} in each term of the sum (2.4.1). Terms with $\tau_{i+j}=1$ sum precisely to $E_r(L,n;i_1,\ldots,i_r;j+1)$, and for terms with $\tau_{i+j}=0$ or $\tau_{i+j}=2$ we use the matrix algebra to reduce

$$X_{\tau_{i_r+j-1}}X_{\tau_{i_r+j}} = \begin{cases} X_1X_0 = X_1+X_0, & \text{if } \tau_{i_r+j}=0, \\ X_1X_2 = X_2, & \text{if } \tau_{i_r+j}=2; \end{cases} \tag{2.4.3}$$

the resulting sum is just $E_r(L-1,n;i_1,\ldots,i_r;j-1)$.

To determine the E_r the recursion (2.4.2) must be supplemented by boundary conditions at the maximum and minimum values of j. When $j=0$, (2.4.1) gives

$$E_r(L,n;i_1,\ldots,i_r;0) = \begin{cases} E_{r-1}(L,n;i_1,\ldots,i_{r-1};1), & \text{if } r\geq 2, \\ Z(L,n), & \text{if } r=1. \end{cases} \tag{2.4.4}$$

The value of E_r for j maximal is case dependent. In Case 1, if j takes its maximum possible value $L - i_r + 1$ then each matrix product in (2.4.1) ends with $X_1^{L-i_r+1}|V_\beta\rangle = \beta^{-(L-i_r+1)}|V_\beta\rangle$, so that

$$E_r(L, n; i_1, \ldots, i_r; L - i_r + 1)$$
$$= \begin{cases} \beta^{-(L-i_r+1)} E_{r-1}(i_r - 1, n; i_1, \ldots, i_{r-1}; 1), & \text{if } r \geq 2, \\ \beta^{-(L-i_r+1)} Z(i_r - 1, n), & \text{if } r = 1, \end{cases} \qquad \text{(Case 1).} \tag{2.4.5}$$

In Case 2, if j takes its maximum possible value $L - n - r + 1$ then the rightmost factor in the matrix product in (2.4.1) is X_2 and there are no factors of X_0; using the matrix algebra relations $X_1 X_2 = X_2$, $X_2^2 = X_2$ and $X_2|V_\beta\rangle = |V_1\rangle$, we have that

$$E_r(L, n; i_1, \ldots, i_r; L - n - r + 1) = 1 \qquad \text{(Case 2).} \tag{2.4.6}$$

Lemma 2.4.1 *The recursion* (2.4.2)*, together with the boundary conditions* (2.4.4) *and either* (2.4.5) *or* (2.4.6)*, determines* $E_r(L, n; i_1, \ldots, i_r; j)$ *by an inductive computation.*

Proof: The primary induction is on increasing values of r, with n held fixed throughout. The inductive assumption that the E_{r-1} are known is needed when (2.4.4) or (2.4.5) is applied, and since for $r = 1$ the right hand side of these equations is a (known) partition function one may treat all values $r \geq 1$ uniformly. Then for fixed r ($r \geq 1$) and $i_1, \ldots, i_r$ we induce on increasing values of L: for the minimum possible value, $L = i_r$, we must be in Case 1 and have $j = 0$ or $j = 1$, so that all $E_r(L, n; i_1, \ldots, i_r; j)$ are determined by (2.4.4) and (2.4.5). For any larger value of L either (2.4.5) or (2.4.6) determines $E_r(L, n; i_1, \ldots, i_r; j)$ for the maximum possible value of j and we may then induce downward on j using (2.4.2). ■

We can now verify exchangeability; we show that $E_r(L, n; i_1, \ldots, i_r; j)$ is equal to the corresponding weight with the sites $i_1, \ldots, i_r$ in standard positions $1, \ldots, r$.

Theorem 2.4.2 *Let $1 \le i_1 < i_2 < \cdots < i_r$ be sites such that $i_r \le n + r - 1$, and let j be an integer less than or equal to $L - n - r + 1$. Then*

$$E_r(L, n; i_1, \ldots, i_r; j) = E_1(L, n; 1; r + j - 1). \tag{2.4.7}$$

Proof: The proof is by induction on r. By Lemma 2.4.1 it suffices to show that $E_1(L, n; 1; r + j - 1)$ satisfies the same recurrence relation (2.4.2) and boundary conditions (2.4.4), (2.4.6) as does $E_r(L, n; i_1, \ldots, i_r; j)$. This follows immediately from the corresponding relations for E_1 and, for $r \ge 2$, the induction assumption. ■

We finally give the explicit formula for E_1 which, by (2.4.7), also provides a formula for $E_r(L, n; i_1, \ldots, i_r; j)$ when $i_r \le n + r - 1$. This result will not be needed in the remainder of the paper.

The formula involves the Catalan triangle numbers (**A009766** of [Slo07])

$$C_n^m = \binom{m+n}{n} \frac{m - n + 1}{m + 1}, \tag{2.4.8}$$

which satisfy the recursion

$$C_n^{m-1} + C_{n-1}^m = C_n^m \tag{2.4.9}$$

and the boundary conditions

$$C_{-1}^m = 0, \qquad C_0^m = 1, \qquad C_m^m = C_{m-1}^m = \frac{1}{m+1}\binom{2m}{m}. \tag{2.4.10}$$

We then define the additional constants

$$c_{j,k} = \begin{cases} C_{k-j}^{k-1}, & \text{if } j \ge 1, \\ \delta_{k,0}, & \text{if } j = 0; \end{cases} \tag{2.4.11}$$

$$d_{m,j,k} = \begin{cases} \binom{m-j+k}{k} - \binom{m-j+k}{k-j}, & \text{if } j \le m, \\ \delta_{k,0}, & \text{if } j = m+1. \end{cases} \tag{2.4.12}$$

In (2.4.11)–(2.4.12) the convention is that $\binom{p}{q} = 0$ for integer p, q with $p \ge 0$, $q < 0$.

Theorem 2.4.3 Case 1: *If* $n+1 \le i \le L$ *and* $0 \le j \le L+1-i$ *then*

$$E_1(L,n;i;j) = \sum_{k=j}^{L-i} c_{j,k} Z(L-k,n) + Z(i-1,n) \sum_{k=0}^{L-i-1} d_{L-i,j,k} \beta^{-L+i+k-1}. \quad (2.4.13)$$

Case 2: *If* $1 \le i \le n$ *and* $0 \le j \le L-n$ *then*

$$E_1(L,n;i;j) = \sum_{k=j}^{L-n} c_{j,k} Z(L-k,n), \quad (2.4.14)$$

Proof: Case 1: We temporarily denote the right hand side of (2.4.13) by $F(L,n;i;j)$. By Lemma 2.4.1 it suffices to verify that F satisfies relations corresponding to (2.4.4), (2.4.5), and (2.4.2). It will be convenient to denote the two terms in (2.4.13) by $F_1(L,n;i;j)$ and $F_2(L,n;i;j)$, respectively.

Since $d_{L-i,0,k} = 0$ and $c_{0,k} = \delta_{k,0}$ we have immediately that $F(L,n;i;0) = Z(L,n)$ (compare (2.4.4)). Moreover, the sum defining $F_1(L,n;i;L-i+1)$ is empty and so from $d_{L-i,L-i+1,k} = \delta_{k,0}$ we have $F(L,n;i;L-i+1) = \beta^{-L+i-1} Z(i-1,n)$ (compare (2.4.5)). It remains to check the equivalent of (2.4.2), which we shall show is satisfied by F_1 and F_2 separately; recall that in (2.4.2), $1 \le j \le L-i$. For $j \ge 2$,

$$\begin{aligned} &F_1(L,n;i;j+1) + F_1(L-1,n;i;j-1) \\ &\quad = \sum_{k=j+1}^{L-i} C_{k-j-1}^{k-1} Z(L-k,n) + \sum_{k=j-1}^{L-1-i} C_{k-j+1}^{k-1} Z(L-1-k,n) \\ &\quad = \sum_{k=j}^{L-i} \left(C_{k-j-1}^{k-1} + C_{k-j}^{k-2} \right) Z(L-k,n) = F_1(L,n;i;j), \end{aligned} \quad (2.4.15)$$

where we have used $C_{-1}^{j-1} = 0$ (see (2.4.10)) and $C_{k-j-1}^{k-1} + C_{k-j}^{k-2} = C_{k-j}^{k-1}$ (see (2.4.9)). The case $j = 1$ is easily checked separately. Similarly one verifies that

$$F_2(L,n;i;j+1) + F_2(L-1,n;i;j-1) = F_2(L,n;i;j) \quad (2.4.16)$$

separately for $j \le L-i-1$ and for $j = L-i$.

Case 2: We denote the right hand side of (2.4.14) by $G(L,n;i;j)$, and show that G satisfies the appropriate boundary conditions and recursion. One checks immediately

that $G(L,n;i;0) = Z(L,n)$ (compare (2.4.4)) and, using (2.4.10), that $G(L,n;i;L-n) = 1$ (compare (2.4.6)). Finally one shows that, for $1 \le j \le L-n-1$,

$$G(L,n;i;j+1) + G(L-1,n;i;j-1) = G(L,n;i;j); \tag{2.4.17}$$

the proof is essentially the same as that of the recursion for F_1 in Case 1. ■

2.5 The fat shock

In this section we give a precise definition and analysis of the fat shock discussed informally in the introduction. The analysis will be used in the next section for the determination of local states in the infinite volume limit. We define the fat shock microscopically as the region between the positions Q_1 and Q_n of the first and last second class particles in the system.

The joint distribution of Q_1 and Q_n was obtained in Remark 2.3.1; it is convenient here to write this, for $j,k,l \ge 0$, as

$$\begin{aligned} \theta^{\alpha,\beta}_{L,n}(j,k,l) &\equiv \mu^{\alpha,\beta}_{L,n}(Q_1 = j+1, Q_n = j+k+2)\,\delta_{j+k+l,L-2} \\ &= \frac{Z^{\alpha,1}(j,0) Z^{1,1}(k,n-2)\, Z^{1,\beta}(l,0)}{Z^{\alpha,\beta}(L,n)}\,\delta_{j+k+l,L-2}. \end{aligned} \tag{2.5.1}$$

We can determine the large-L behavior of θ by replacing the partition functions in (2.5.1) with their asymptotic forms; these can be obtained from [DEHP93] and [Ari06b], and are summarized in Appendix A.2. In some cases it is convenient to further approximate the distribution of k, which represents the fat shock width on the microscopic scale, by a Gaussian. (Recall that a macroscopic width $w = w(\alpha,\beta,\gamma)$ for the fat shock was predicted on heuristic grounds in Section 1 (see (2.2.7)), so we expect that $k \sim wL$ for large L.) We omit details of the computations and summarize the results in the next remark; where the notation $\theta^{\alpha,\beta}_{L,n}(j,k,l) \sim f(\alpha,\beta,\gamma,j,k,l)$ indicates that the ratio of the two quantities approaches 1 as $L \to \infty$ with $n = \lfloor \gamma L \rfloor$, where $\lfloor u \rfloor$ is the greatest integer contained in u.

Remark 2.5.1 *(a) On the boundary of regions II and III ($\alpha = \beta < \alpha_c$),*

$$\theta^{\alpha,\alpha}_{L,n}(j,k,l) \sim \frac{1}{L(1-w)} \sqrt{\frac{A(\alpha)}{\pi L}}\, e^{-A(\alpha)(k-Lw)^2/L}\, \delta_{j+k+l,L-2}, \tag{2.5.2}$$

where $A(\alpha) = (1-2\alpha)^3/(4\gamma\alpha(1-\alpha))$. That is, under $\theta^{\alpha,\alpha}_{L,n}(j,k,l)$, k is approximately Gaussian with mean Lw and variance of order L, j is approximately uniformly distributed on the range $0 \le j \le L-l-2$, and $l = L-2-j-k$. On the macroscopic scale, this means that the width of the fat shock is w and its left endpoint is uniformly distributed on the interval $[0, 1-w]$.

(b) In region II ($\alpha < \alpha_c, \alpha < \beta$),

$$\theta^{\alpha,\beta}_{L,n}(j,k,l) \sim p^{\alpha(1-\alpha),\beta}(l) \sqrt{\frac{A(\alpha)}{\pi L}}\, e^{-A(\alpha)(k-Lw)^2/L}\, \delta_{j+k+l,L-2}, \tag{2.5.3}$$

where we have introduced the (normalized) probability distribution

$$p^{u,\beta}(l) = \frac{\beta(1+\sqrt{1-4u}) - 2u}{2\beta}\, u^l Z^{\beta,1}(l,0), \quad l = 0,1,\ldots, \tag{2.5.4}$$

defined for $u < \beta(1-\beta)$ if $\beta \le 1/2$, $u \le 1/4$ otherwise. $p^{u,\beta}(l)$ decreases exponentially for large l unless $u = 1/4$, when the decrease is as $l^{-3/2}$ (see (A.2.1)*); it follows from* (A.2.2) *that p is normalized. Thus on the microscopic scale l is typically of order 1 and the shock width k is distributed as in (a). On the macroscopic scale the fat shock has width w and is pinned to the right end of the system. The analysis in region III is similar.*

(c) On the boundary of regions I and II ($\alpha_c = \alpha < \beta$),

$$\theta^{\alpha,\beta}_{L,n}(j,k,l) \sim p^{\alpha(1-\alpha),\beta}(l)\, 2\sqrt{\frac{A(\alpha)}{\pi L}}\, e^{-A(\alpha)(k-L)^2/L}\, \delta_{j+k+l,L-2}. \tag{2.5.5}$$

This is as in (b) except that here $w = 1$ and as a result k is distributed as a Gaussian random variable conditioned to have value at most equal to its mean, and there is a corresponding factor of 2 in the normalization. The analysis on the I/III boundary is similar.

(d) At the triple point ($\alpha_c = \alpha = \beta$),

$$\theta_{L,n}^{\alpha,\beta}(j,k,l) \sim \frac{2\gamma^2}{L(1-\gamma)^2}\, e^{-A(\alpha)(k-L)^2/L}\, \delta_{j+k+l,L-2}. \tag{2.5.6}$$

The distribution of k is as in (c) but here j and l are free, subject only to the constraint $j + l = L - 2 - k$.

(e) In region I ($\alpha_c < \alpha, \beta$),

$$\theta_{L,n}^{\alpha,\beta}(j,k,l) \sim p^{(1-\gamma^2)/4,\alpha}(j)\, p^{(1-\gamma^2)/4,\beta}(l)\, \delta_{j+k+l,L-2}; \tag{2.5.7}$$

j and l are both of order 1 (microscopically) and k is determined by the delta function constraint.

Note that the results of Remark 2.5.1 confirm the picture of the fat shock behavior sketched in Section 1.

2.6 Local states in the infinite volume limit in the bulk

In this section we discuss a question inspired by the treatment of the one component system by Liggett [Lig99]: is there a *local state* at position x of the system (in the infinite volume limit), and if so what is it? To formulate a precise question we consider a limit in which n and i increase with L in such a way that $i \to \infty$, $L - i \to \infty$, $i/L \to x \in [0,1]$, and $n/L \to \gamma \in (0,1)$. In this setting we ask about the existence and nature of the weak limit $\lim_{L\to\infty} T^{-i}\mu_{L,n}^{\alpha,\beta}$, where T is translation by one lattice site and so the operator T^{-i} carries site i to the origin; equivalently, we consider the sites of our open system to run from $1-i$ to $L-i$ and look at the probabilities of configurations in the interval from $-K$ to K, take L, i, $L-i$, and n to infinity as above, and then make K arbitrary. The limit (if it exists) is a measure on the configuration space $Y = \{0,1,2\}^{\mathbb{Z}}$; we call it a *local state in the bulk* since (in the $L \to \infty$ limit) it describes a situation infinitely far from each boundary; the *local state at the boundary* will be discussed in Section 2.7.

It will suffice to consider a special class of these limiting procedures; specifically, we will always take

$$n = n_L = \lfloor \gamma L \rfloor \quad \text{and} \quad i = i_L = \lfloor xL \rfloor + c\sqrt{L}; \tag{2.6.1}$$

we must assume that $c > 0$ if $x = 0$ and $c < 0$ if $x = 1$. We then define

$$\mu_{x,c} \equiv \lim_{L\to\infty} T^{-i_L} \mu^{\alpha,\beta}_{L,n_L}. \tag{2.6.2}$$

The limit in (2.6.2) certainly exists along subsequences, by the compactness of the set of measures on Y. To simplify notation we will ignore the necessity of passing to subsequences; since the limiting measure will be found to be unique, the limit of the sequence itself must also exist. For most values of the parameters the limit will in fact be independent of the choice of c, but this is not true when $x = x_0$ in region II or on the I/II boundary, or $x = x_1$ in region III or on the I/III boundary.

We first consider the currents and densities in the state $\mu_{x,c}$. The currents in the finite system, and hence also in the limit, are independent of the site:

$$\langle \eta_1(j)(1 - \eta_1(j+1)) \rangle_{\mu_{x,c}} = \langle (1 - \eta_0(j-1))\eta_0(j) \rangle_{\mu_{x,c}} = J_1 \tag{2.6.3}$$

for any j, with J_1 given in (2.2.1). The limiting densities $\rho_a(x,c)$ for $a = 0, 1, 2$ are defined by

$$\rho_a(x,c) = \lim_{L\to\infty} \langle \eta_a(i_L) \rangle_{\mu^{\alpha,\beta}_{L,n_L}} = \langle \eta_a(0) \rangle_{\mu_{x,c}}, \tag{2.6.4}$$

with the last equality expressing the fact that i_L corresponds to the origin in $\mu_{x,c}$. It is easy to check, from the asymptotic computations of [Ari06b], that the limit in (2.6.4) would be unchanged if i_L were replaced by $i_L + j$ for any fixed j, which implies that $\langle \eta_a(j) \rangle_{\mu_{x,c}} = \rho_a(x,c)$ for any j, i.e., the densities under $\mu_{x,c}$ are translation invariant. Equation (2.6.4) may be viewed as a refined version of (2.2.3), in which the ambiguity in the $L \to \infty$ limit there has been removed.

Noting that in the $L \to \infty$ limit the generator of the dynamics in the neighborhood of i_L does not involve any boundary terms or any constraints on the densities of the

three species beyond $\sum_a \eta_a(j) = 1$, one verifies easily [Lig99] that $\mu_{x,c}$ must be invariant for the infinite-volume two species TASEP dynamics. It then follows that $\mu_{x,c}$ must be a convex combination of the extremal invariant measures for the infinite volume two species TASEP. These measures have been classified in [Spe94]: there is (i) a family of translation invariant measures $\nu^{\lambda_0,\lambda_1}$, defined for $\lambda_0, \lambda_1 \geq 0$, $\lambda_0+\lambda_1 \leq 1$, in which holes, first class particles, and second class particles have densities λ_0, λ_1, and $1-\lambda_0-\lambda_1$, respectively, and (ii) a family of non-translation-invariant "blocking" measures $\hat{\nu}^{m,n}$, where $m, n \in \mathbb{Z} \cup \{-\infty, \infty\}$, $m \leq n$, and m, n are not both infinite: $\hat{\nu}^{m,n}$ is a unit point mass on the configuration $\tau^{m,n}$ given by

$$\tau_i^{m,n} = \begin{cases} 0, & \text{if } i < m, \\ 2, & \text{if } m \leq i < n, \\ 1, & \text{if } n \leq i. \end{cases} \tag{2.6.5}$$

However, the translation invariance of the densities implies that none of the blocking measures can be present in the superposition giving $\mu_{x,c}$.

Thus there exists a probability measure $\kappa_{x,c}(d\lambda_0, d\lambda_1)$ (which depends also on α, β, and γ) that specifies the weights of the different translation invariant extremal measures which enter into the superposition:

$$\mu_{x,c} = \int_{\lambda_0,\lambda_1 \geq 0,\ \lambda_0+\lambda_1 \leq 1} \kappa_{x,c}(d\lambda_0, d\lambda_1)\, \nu^{\lambda_0,\lambda_1}. \tag{2.6.6}$$

We will see that: (i) for most values of the parameters, $\kappa_{x,c}$ is a point mass, so that $\mu_{x,c}$ is one of the measures $\nu^{\lambda_0,\lambda_1}$, (ii) in some cases, in which x may lie to the left of, within, or to the right of the fat shock discussed in the introduction, $\mu_{x,c}$ is a superposition of the two or three measures corresponding to these possibilities, and (iii) no more complicated superposition can occur. Note that, as a consequence, the current J_1 is the same for all elements of the superposition (and the same holds for J_0 and for $J_2 = 0$). Here is a first result in this direction.

Theorem 2.6.1 *If $\mu_{x,c}$ is defined by* (2.6.2) *and the current and densities at x are related by $J_1 = \rho_0(x,c)(1-\rho_0(x,c)) = \rho_1(x,c)(1-\rho_1(x,c))$, then $\mu_{x,c} = \nu^{\rho_0(x,c),\rho_1(x,c)}$.*

The condition in the theorem that $\rho_0(x,c)(1-\rho_0(x,c)) = \rho_1(x,c)(1-\rho_1(x,c))$ corresponds to the zero current of second class particles and leads to the alternatives of (2.2.6). We see from Table 2.1 that this theorem determines $\mu_{x,c}$ completely for most but not all values of α, β, γ, x, and c and that the results are consistent with the intuitive picture sketched in the introduction. In summary:

Remark 2.6.2 *It follows from Theorem 2.6.1 that:*

(a) In region I the local state $\mu_{x,c}$ is ν^{α_c,α_c}; in particular, it is independent of x and c.

(b) In region II the local state $\mu_{x,c}$ is $\nu^{1-\alpha,\alpha}$ for $x < x_0$ and $\nu^{\alpha,\alpha}$ for $x > x_0$, i.e., respectively outside and inside the fat shock. Region III is similar: the local state is $\nu^{\beta,\beta}$ for $x < x_1$ and $\nu^{\beta,1-\beta}$ for $x > x_1$.

Other cases are not determined by the theorem:

(c) The local state is not determined by Theorem 2.6.1 in the interior of the regions where either type 2 or type 0 particles have a linear profile, that is, on the II/III boundary (where the fat shock is not pinned to one or the other end of the system) with $0 < x < x_0$ or $1 > x > x_1$. See Figure 2.3(c,d).

(d) The local state is not determined by Theorem 2.6.1 (i) at $x = x_0$ in region II and on the I/II boundary, where $x_0 = 0$; (ii) at $x = x_1$ in region III and on the I/III boundary, where $x_1 = 1$, or (iii) at $x = x_0 = 0$ and $x = x_1 = 1$ at the triple point. All of these points are edges of the (pinned) fat shock; see Figure 2.3(a,b,e).

We will determine $\mu_{x,c}$ in cases (c) and (d) below.

Proof of Theorem 2.6.1: Using (2.6.6) together with the relations $\langle \eta_1(i) \rangle_{\nu^{\lambda_0,\lambda_1}} = \lambda_1$, $\langle \eta_1(i)(1-\eta_1(i+1)) \rangle_{\nu^{\lambda_0,\lambda_1}} = \lambda_1(1-\lambda_1)$ (which hold for all i because the marginal of

$\nu^{\lambda_0,\lambda_1}$ on configurations of first class particles is a product measure), we find that

$$\rho_1(x,c) = \langle \eta_1(i) \rangle_{\mu_{x,c}} = \int \lambda_1 \, \kappa_{x,c}(d\lambda_0, d\lambda_1) = \langle \lambda_1 \rangle_{\kappa_{x,c}} , \tag{2.6.7}$$

and

$$\begin{aligned} J_1 &= \langle \eta_1(i)(1-\eta_1(i+1)) \rangle_{\mu_{x,c}} \\ &= \int \lambda_1(1-\lambda_1) \, \kappa_{x,c}(d\lambda_0, d\lambda_1) = \langle \lambda_1(1-\lambda_1) \rangle_{\kappa_{x,c}} . \end{aligned} \tag{2.6.8}$$

From $J_1 = \rho_1(x,c)(1-\rho_1(x,c))$, then, we see that $\langle \lambda_1^2 \rangle_{\kappa_{x,c}} = \langle \lambda_1 \rangle_{\kappa_{x,c}}^2$, so that $\lambda_1 = \langle \lambda_1 \rangle_{\kappa_{x,c}} = \rho_1(x,c)$ almost surely with respect to $\kappa_{x,c}$. Similarly, $\lambda_0 = \rho_0(x,c)$ almost surely with respect to $\kappa_{x,c}$, so that $\mu_{x,c} = \nu^{\rho_0(x,c),\rho_1(x,c)}$. □

We now turn to the determination of the local measure $\mu_{x,c}$ at a point x where the densities are varying linearly or are discontinuous—case (c) or (d) of Remark 2.6.2. Recall that in Section 2.5 we have determined the probability $\theta_{L,n}^{\alpha,\beta}(j,k,l)$ that $Q_1 = j+1$ and $Q_n = j+k+2$, where $j+k+l = L-2$ and Q_1 and Q_n are the position of the first and last second class particles. Now let $\mu_{L,n,j,k,l}^{\alpha,\beta}$ denote the measure $\mu_{L,n}^{\alpha,\beta}$ conditioned on $Q_1 = j+1$, $Q_n = j+k+2$. The key observation we will use follows from Remark 2.3.1, specifically, from (2.3.9) or a simple generalization thereof: if G is a function on $Y_{L,n}$ which depends on the τ_i only for $m_0 \le i \le m_1$, then

$$\langle G \rangle_{\mu_{L,n,j,k,l}^{\alpha,\beta}} = \begin{cases} \langle G \rangle_{\mu_{j,0}^{\alpha,1}} , & \text{if } m_1 \le j, \\ \langle T^{-j-1} G \rangle_{\mu_{k,n-2}^{1,1}} , & \text{if } j+2 \le m_0, \ m_1 \le i+k+1, \\ \langle T^{-j-k-2} G \rangle_{\mu_{l,0}^{1,\beta}} , & \text{if } j+k+3 \le m_0. \end{cases} \tag{2.6.9}$$

Now let (r_L) be a sequence of integers with $r_L \to \infty$ and $r_L/\sqrt{L} \to 0$. For any function F on Y depending only on finitely many spins we may write

$$\begin{aligned} \langle F \rangle_{\mu_{x,c}} &= \lim_{L\to\infty} \sum_{j+k+l=L-2} \theta_{L,n_L}^{\alpha,\beta}(j,k,l) \, \langle T^{i_L} F \rangle_{\mu_{L,n_L,j,k,l}^{\alpha,\beta}} \\ &= \lim_{L\to\infty} \big[\Xi_L^{(1)} \langle F \rangle_{\mu_L^{(1)}} + \Xi_L^{(2)} \langle F \rangle_{\mu_L^{(2)}} + \Xi_L^{(3)} \langle F \rangle_{\mu_L^{(3)}} + \text{remainder} \big]. \end{aligned} \tag{2.6.10}$$

Here $\mu_L^{(p)}$ is for $p = 1, 2, 3$ the probability measure defined by

$$\mu_L^{(p)} = \Xi_L^{(p)-1} \sum\nolimits_{j,k,l}^{(p)} \theta_{L,n_L}^{\alpha,\beta}(j,k,l) T^{-i_L} \mu_{L,n_L,j,k,l}^{\alpha,\beta}, \tag{2.6.11}$$

where $\sum_{j,k,l}^{(1)}$ ranges over values satisfying $j > i_L + r_L$, $\sum_{j,k,l}^{(2)}$ over $j < i_L - r_L$ and $k > i_L + r_L$, $\sum_{j,k,l}^{(3)}$ over $k < i_L - r_L$, and

$$\Xi_L^{(p)} = \sum\nolimits_{j,k,l}^{(p)} \theta_{L,n_L}^{\alpha,\beta}(j,k,l) = \begin{cases} \mu_{n,L}^{\alpha,\beta}(Q_1 > i_L + r_L), & \text{if } p = 1, \\ \mu_{n,L}^{\alpha,\beta}(Q_1 < i_L - r_L,\ Q_n > i_L + r_L), & \text{if } p = 2, \\ \mu_{n,L}^{\alpha,\beta}(Q_n < I_L - r_L), & \text{if } p = 3. \end{cases} \tag{2.6.12}$$

The remainder in (2.6.11) contains those terms from the full sum over i and k which are omitted from $\sum^{(1)}$, $\sum^{(2)}$, and $\sum^{(3)}$. We have suppressed the dependence of these various entities on α, β, γ, x, and c.

We now take the $L \to \infty$ limit in (2.6.10). It follows from Remark 2.5.1 and the fact that r_L grows more slowly than $\sqrt{L}$ that the remainder vanishes in this limit. The $\Xi_L^{(p)}$ are expressed as probabilities in (2.6.12) and their limiting values $\Xi_{x,c}^{(p)} = \lim_{L\to\infty} \Xi_L^{(p)}$ may be determined from Remark 2.5.1; these limits will be discussed on a case by case basis below.

Finally, to study $\lim_{L\to\infty} \mu_L^{(p)}$ we observe that for sufficiently large L (if F depends on τ_i only for $|i| \le m$ then $r_L > m$ suffices) we have by (2.6.9) that

$$\langle F \rangle_{\mu_L^{(p)}} = \begin{cases} \Xi_L^{(1)-1} \sum_{j,k,l}^{(1)} \theta_{L,n_L}^{\alpha,\beta}(j,k,l) \left\langle T^{i_L} F \right\rangle_{\mu_{j,0}^{\alpha,1}}, & \text{if } p = 1, \\ \Xi_L^{(2)-1} \sum_{j,k,l}^{(2)} \theta_{L,n_L}^{\alpha,\beta}(j,k,l) \left\langle T^{i_L - j - 1} F \right\rangle_{\mu_{k,n-2}^{1,1}}, & \text{if } p = 2, \\ \Xi_L^{(3)-1} \sum_{j,k,l}^{(3)} \theta_{L,n_L}^{\alpha,\beta}(j,k,l) \left\langle T^{i_L - j - k - 2} F \right\rangle_{\mu_{l,0}^{1,\beta}}, & \text{if } p = 3. \end{cases} \tag{2.6.13}$$

The limits $\lim_{L\to\infty} \mu_L^{(p)}$ for $p = 1, 2, 3$ are all treated similarly; let us discuss the case $p = 2$ in detail. Equation (2.6.13) displays $\mu_L^{(2)}$ as a convex combination of the probability measures $T^{-(i_L - j - 1)} \mu_{k,n-2}^{1,1}$. Each of these measures, for large L, is by Remark 2.6.2(a) approximately equal to $\nu^{\alpha\wedge\beta,\alpha\wedge\beta}$ (recall that $\alpha \wedge \beta = \min\{\alpha, \beta\}$), since the critical value

α_c^* of α for a system with $k \simeq wL = \gamma L/(1-2\alpha\wedge\beta)$ sites and $n \simeq \gamma L$ second class particles—and thus an effective value $\gamma^* = n/k = 1-2\alpha\wedge\beta$ of γ—is $(1-\gamma^*)/2 = \alpha\wedge\beta$. The same should be true of $\mu_L^{(2)}$. The corresponding evaluations for $p = 1, 3$ come from the results of [Lig99] for the local measures in the one species open system. We conclude that

$$\lim_{L\to\infty} \mu_L^{(p)} = \begin{cases} \nu^{1-\alpha\wedge(1/2),\alpha\wedge(1/2)}, & \text{if } p = 1, \\ \nu^{\alpha\wedge\beta,\alpha\wedge\beta}, & \text{if } p = 2, \\ \nu^{\beta\wedge(1/2),1-\beta\wedge(1/2)}, & \text{if } p = 3, \end{cases} \tag{2.6.14}$$

and so

$$\mu_{x,c} = \Xi_{x,c}^{(1)}\nu^{1-\alpha\wedge(1/2),\alpha\wedge(1/2)} + \Xi_{x,c}^{(2)}\nu^{\alpha\wedge\beta,\alpha\wedge\beta} + \Xi_{x,c}^{(3)}\nu^{\beta\wedge(1/2),1-\beta\wedge(1/2)}. \tag{2.6.15}$$

Remark 2.6.3 *The heuristic argument for* (2.6.14) *given above could be made precise by justifying the implicit exchange of limits; we sketch instead an alternate proof, again for* $p = 2$. *We know that the limiting current for the measures* $\mu_{k,n-2}^{1,1}$, *as* L *and hence* $k \simeq wL$ *goes to infinity, is* $\alpha(1-\alpha)$, *and the limiting densities of holes, particles, and second class particles are* α, α, *and* $1-2\alpha$, *respectively. One can in fact show further that these limits are obtained with error which goes to zero uniformly at sites* i *satisfying* $r_L \le i \le k - r_L$; *from this, it follows that the measures* $\mu_L^{(2)}$ *have the same limiting current and densities. Then an argument as in the proof of Theorem 2.6.1 establishes* (2.6.14).

To complete our discussion of the local states in the bulk we must supplement (2.6.15) with a determination of the weights $\Xi_{x,c}^{(p)} \equiv \lim_{L\to\infty} \Xi_L^{(p)}$ for cases (c) and (d) of Remark 2.6.2. The cases in the next remark are parallel to those of Remark 2.5.1.

Remark 2.6.4 *(a) On the boundary of regions II and III (the shock line, case (c) of*

Remark 2.6.2) we find from Remark 2.5.1(a) and (2.6.13) *that*

$$\Xi^{(p)}_{x,c} = \frac{1}{1-w} \times \begin{cases} (1-w-x)_+, & \text{if } p=1, \\ 1-w-(1-w-x)_+-(x-w)_+, & \text{if } p=2, \\ (x-w)_+, & \text{if } p=3. \end{cases} \tag{2.6.16}$$

Here $u_+ = u$ *if* $u \geq 0$ *and* $u_+ = 0$ *if* $u < 0$. *Note that these coefficients, and hence the local measure* $\mu_{x,c}$ *given by* (2.6.15), *are independent of* c.

(b) In region II ($\alpha < \alpha_c, \alpha < \beta$), *at the fixed shock* x_0, *the* $\Xi^{(p)}_{x,c}$ *do depend on* c:

$$\Xi^{(1)}_{x_0,c} = 1 - \Phi\left(c\sqrt{A(\alpha)}\right), \qquad \Xi^{(2)}_{x_0,c} = \Phi\left(c\sqrt{A(\alpha)}\right), \qquad \Xi^{(3)}_{x_0,c} = 0. \tag{2.6.17}$$

Here Φ *is the error function defined by*

$$\Phi(t) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{t} e^{-\tau^2/2}\, d\tau. \tag{2.6.18}$$

The analysis in region III is similar.

(c) On the boundary of regions I and II ($\alpha_c = \alpha < \beta$), *where* $x_0 = 0$, *we find that for* $c > 0$,

$$\Xi^{(1)}_{0,c} = 2 - 2\Phi\left(c\sqrt{A(\alpha)}\right), \qquad \Xi^{(2)}_{0,c} = 2\Phi\left(c\sqrt{A(\alpha)}\right) - 1, \qquad \Xi^{(3)}_{0,c} = 0. \tag{2.6.19}$$

The analysis on the I/III boundary is similar: for $c < 0$,

$$\Xi^{(1)}_{1,c} = 0, \qquad \Xi^{(2)}_{1,c} = 1 - 2\Phi\left(c\sqrt{A(\alpha)}\right), \qquad \Xi^{(3)}_{1,c} = 2\Phi\left(c\sqrt{A(\alpha)}\right). \tag{2.6.20}$$

(d) At the triple point ($\alpha_c = \alpha = \beta$), *where* $x_0 = 0$ *and* $x_1 = 1$, (2.6.19) *and* (2.6.20) *again hold (with* $c > 0$ *and* $c < 0$ *respectively).*

We finally note that the results of this section yield density profiles as well as the finite volume corrections to these profiles at the fixed shocks (see Remark 2.2.1(a)), since $\rho_a(x, c)$ may be calculated from (2.6.4) and (2.6.15). For example, on the II/III border (shock line) we find in this way that

$$\rho_0(x) = \Xi^{(1)}_{x,c}(1-\alpha) + (\Xi^{(2)}_{x,c} + \Xi^{(3)}_{x,c})\alpha, \quad \rho_1(x) = (\Xi^{(1)}_{x,c} + \Xi^{(2)}_{x,c})\alpha + \Xi^{(3)}_{x,c}(1-\alpha), \tag{2.6.21}$$

with the weights $\Xi^{(p)}$ given by (2.6.16); it is easy to see that (2.6.21) reproduces (2.2.5). Here we have used the notation $\rho_a(x)$ of (2.2.3), rather than writing $\rho_a(x,c)$ as in (2.6.4), since the densities do not depend on c. In region II we have, from (2.6.15) and (2.6.16), that

$$\rho_0(x_0,c) = 1-\alpha-\Phi\left(c\sqrt{A(\alpha)}\right)(1-2\alpha). \tag{2.6.22}$$

Similar results hold in region III at the point x_1.

2.7 Local states in the infinite volume limit at the boundaries

In this section we study limiting measures $\lim_{L\to\infty} T^{-i_L}\mu^{\alpha,\beta}_{L,n_L}$ as in (2.6.2), still taking $n_L = \lfloor \gamma L \rfloor$ but now assuming that either i_L or $L-i_L$ is fixed. Without loss of generality we can assume that $i_L = 1$ or $i_L = L$ (the measure as seen from site j or site $L-j$ can be recovered from these limits) and thus define

$$\mu_{\text{left}} \equiv \lim_{L\to\infty} T^{-1}\mu^{\alpha,\beta}_{L,\lfloor\gamma L\rfloor}, \qquad \mu_{\text{right}} \equiv \lim_{L\to\infty} T^{-L}\mu^{\alpha,\beta}_{L,\lfloor\gamma L\rfloor}. \tag{2.7.1}$$

Note that μ_{left} (respectively μ_{right}) does not coincide with any of the measures $\mu_{0,c}$, $c>0$, (respectively $\mu_{1,c}$, $c<0$,) studied in Section 2.6. The densities under μ_{left} and μ_{right} were studied in [Ari06b]; for example, $\langle \eta_a(j)\rangle_{\mu_{\text{left}}}$ is denoted $\rho^a_{\text{left},j}$ in [Ari06b].

By the particle hole symmetry it suffices to consider μ_{left}, which is a measure on the semi-infinite configuration space $\{0,1,2\}^{\{0,1,2,3,\dots\}}$. In general we do not have a proof that the limit defining μ_{left} exists (except along subsequences), although we expect this to be true; see also the comment below Theorem 2.7.1. The next result, however, gives a somewhat surprising property which any (subsequence) limit must satisfy; to simplify notation, we will speak as if the limit itself exists.

Theorem 2.7.1 *The distribution of first class particles under the measure* μ_{left} *is*

Bernoulli with a constant density ρ, where ρ is given by

$$\rho = \begin{cases} \alpha_c, & \textit{in region I,} \\ \alpha, & \textit{in region II,} \\ \beta, & \textit{in region III.} \end{cases} \tag{2.7.2}$$

Proof: By Theorem 2.4.2, we know that the (marginal) distribution of the variables $\eta_1(i)$ under μ_{left} is exchangeable, so that by de Finetti's theorem [Fel71] this distribution is a superposition of Bernoulli distributions. From [Ari06b] we know that for any $i \geq 0$, $\rho \equiv \langle \eta_1(i) \rangle_{\mu_{\text{left}}} = \lim_{L\to\infty} \langle \eta_1(i) \rangle_{\mu^{\alpha,\beta}_{L,\lfloor \gamma L \rfloor}}$ is given by (2.7.2) and that $\lim_{L\to\infty} \langle \eta_1(i)(1-\eta_1(i+1) \rangle_{\mu^{\alpha,\beta}_{L,\lfloor \gamma L \rfloor}} = J_1$ (see (2.2.1)). In each region of the phase plane these limits satisfy the relation $J = \rho(1-\rho)$. Then from an argument as in the proof of Theorem 2.6.1 it follows that the $\eta_1(i)$ distribution is the product measure ν^ρ, where ρ is given by (2.7.2). ■

Note that in region II the density of second class particles any finite distance from the left boundary goes to zero as $L \to \infty$ [Ari06b], so that knowing that the distribution of particles is Bernoulli completely determines any limiting measure to be a Bernoulli measure on particles and holes only. It follows that the limiting measure exists without passing to subsequences.

We discuss briefly one aspect of the measure μ_{left} in the limit $\gamma \to 0$ (note that we are taking this limit *after* the $L \to \infty$ limit). Consider first region I; from Remark 2.5.1(e) we see that the position Q_1 of the first second class particle in the system is distributed according to $p^{1/4,\alpha}(q_1)$; this is a normalizable distribution which decreases as $q_1^{-3/2}$ for large q_1, so that there remains a second class particle in the system, but $\langle Q_1 \rangle_{\mu_{\text{left}}} = \infty$. In fact more is true; by a calculation similar to that of Remark 2.5.1 one can show that all $Q_j - Q_{j-1}$, $j = 2, 3, , \ldots$, have this same distribution (see also the discussion of the pressure ensemble in Section 2.8) so that there remain an infinite number of second class particles in the system under μ_{left}. The same is true in region III, but there by

Remark 2.5.1(b) Q_1 is distributed according to $p^{\beta(1-\beta),\alpha}$, so that $\langle Q_1 \rangle_{\mu_{\rm left}} < \infty$; the distribution of the $Q_j - Q_{j-1}$, $j = 2, 3, , \ldots$, is the same as in region I.

Remark 2.7.2 *One may compare Theorem 2.7.1 with result in [DJLS93] for the infinite volume limit of a two-component TASEP system on a ring: that the distribution of first class particles to the right of a second class particle, and that of holes to the left of such a particle, is Bernoulli. The two results are closely related, because if we set* $\alpha = \beta = 1$ *in the open system then the matrix element* $\langle W_1 | X_{\tau_1} \cdots X_{\tau_L} | V_1 \rangle$ *giving the weight of the configuration* $\tau_1, \ldots, \tau_L$ *is [DJLS93] exactly the weight of the configuration* $2, \tau_1, \ldots, \tau_L$ *on a ring. Because the numbers of first class particles and of holes on the ring is fixed, and these numbers fluctuate in the open system, this does not establish an exact equivalence of the* $\alpha = \beta = 1$ *case of Theorem 2.7.1 to the result of [DJLS93]; nevertheless, it is clear that the former is in some sense a generalization of the latter to values of* α *and* β *other than 1. But the result of [DJLS93] is in another sense more general than Theorem 2.7.1, since the infinite volume limit of the open system has zero current of second class particles, but this is not true for the system of [DJLS93].*

2.8 The pressure ensemble for second class particles

We here consider the steady state distribution of the second class particles only, so that one may think of identifying the first class particles and holes as a new type of hole. For d a positive integer we define

$$\phi_\alpha(d) = -\log(4^{-d} Z^{\alpha,1}(d-1,0)) = -\log(4^{-d} \langle W_\alpha | (X_0 + X_1)^{d-1} | V_1 \rangle) \qquad (2.8.1)$$

It follows from the (α, β) symmetry of $Z^{\alpha,\beta}(L,n)$ that $\phi_\beta(d)$ is also equal to $-\log(4^{-d} Z^{1,\beta}(d-1,0))$. Using (2.3.8) we find that the probability that the n second

class particles in the systems are located at sites $q_1 < q_2 < \cdots < q_n$ is

$$\mu_{L,n}^{\alpha,\beta}(Q_1 = q_1, \ldots, Q_n = q_n) \\ = (4^{-L} Z_{\alpha,\beta}(L,n))^{-1} e^{-\phi_\alpha(q_1) - \sum_{i=2}^{n} \phi(q_i - q_{i-1}) - \phi_\beta(L - q_n)}, \quad (2.8.2)$$

where we have denoted $\phi_1(d)$ by $\phi(d)$. The motivation for the factors 4^{-d} in (2.8.1) will be discussed below; with this normalization $\phi(d) \sim (3/2) \log d$ for $d \to \infty$ [DJLS93, FFK94].

We note that (2.8.2) has precisely the form of the canonical distribution for a system in a domain of length L with particles interacting with their nearest neighbor only via a pair potential $\phi(d)$. (Such an interaction is rather unphysical; one may think of any intervening particle as screening the interaction of particles that it separates.). There is also a potential $\phi_\alpha(d)$ ($\phi_\beta(d)$) representing the interaction of the first (last) particle with the left (right) boundary.

The TASEP dynamics for the full system gives rise in a natural way to a dynamics on the system of the second class particles which satisfies detailed balance with respect to this equilibrium measure. In the state in which the second class particles are at $q_1, \ldots, q_n$ the i^{th} second class particle moves to site $q_i + 1$ at rate 1 whenever that site is empty (in the original sense), an event which by a simple generalization of (2.3.9) occurs in the NESS with probability

$$\frac{\langle W_1 | X_0 (X_0 + X_1)^{q_{i+1} - q_i - 2} | V_1 \rangle}{Z^{1,1}(q_{i+1} - q_i - 1, 0)} = \frac{e^{-\phi(q_{i+1} - q_i - 1)}}{e^{-\phi(q_{i+1} - q_i)}}, \qquad \text{if } i < n, \quad (2.8.3)$$

and with probability

$$\frac{\langle W_1 | X_0 (X_0 + X_1)^{L - q_i - 1} | V_\beta \rangle}{Z^{1,\beta}(L - q_i, 0)} = \frac{e^{-\phi_\beta(L - q_i - 1)}}{e^{-\phi_\beta(L - q_i)}}, \qquad \text{if } i = n. \quad (2.8.4)$$

One finds similarly that the probability that the site $q_i - 1$ is occupied by a first class particle is

$$\frac{e^{-\phi(q_i - q_{i-1} - 1)}}{e^{-\phi(q_i - q_{i-1})}}, \quad \text{if } i > 1, \qquad \frac{e^{-\phi_\alpha(q_1 - 1)}}{e^{-\phi_\alpha(q_i)}}, \quad \text{if } i = 1. \quad (2.8.5)$$

The dynamics in which $q_i \to q_i + 1$ when $q_{i+1} - q_i \geq 2$, with rate given by (2.8.3), and $q_i \to q_i - 1$ when $q_i - q_{i-1} \geq 2$, with rate given by (2.8.5), is easily seen to satisfy detailed balance with respect to the measure (2.8.2).

To obtain the properties of the system described by (2.8.2) in the thermodynamic limit, $L \to \infty, n/L \to \gamma$, it is most convenient to consider the *pressure* or *isobaric ensemble* $\pi_{p,n}^{\alpha,\beta}$ [Hil56, Per87], in which instead of keeping the volume L of the system fixed we imagine that the right wall is in contact with a reservoir of pressure p. The value of p can be chosen so as to make the average volume equal to L, as discussed below. More precisely, we let the position of the right boundary, which we denote q_{n+1}, fluctuate, and add a term involving the pressure p to the measure. This yields the probability distribution in the pressure ensemble:

$$\begin{aligned}&\pi_{p,n}^{\alpha,\beta}(q_1, \ldots, q_n, q_{n+1}) \\ &\quad = \mathcal{Z}^{\alpha,\beta}(p,n)^{-1} \exp\left(-\phi_\alpha(q_1) - \sum_{i=2}^{n} \phi(q_i - q_{i-1}) - \phi_\beta(L - q_n) - p q_{n+1} \right).\end{aligned} \tag{2.8.6}$$

The partition function has the form $\mathcal{Z}^{\alpha,\beta}(p,n) = \mathcal{Z}_1(\alpha,p)\mathcal{Z}_2(p)^n\mathcal{Z}_1(\beta,p)$, where $\mathcal{Z}_1$ and $\mathcal{Z}_2$ are readily found, for $z = \sqrt{1 - e^{-p}}$ satisfying

$$1 \geq z \geq \max\{0, 1 - 2\alpha, 1 - 2\beta\}, \tag{2.8.7}$$

to be given by

$$\mathcal{Z}_1(\alpha,p) = \frac{\alpha(1-z)}{z + 2\alpha - 1}, \qquad \mathcal{Z}_2(p) = \frac{1-z}{1+z}. \tag{2.8.8}$$

Thus (2.8.6) becomes

$$\begin{aligned}\pi_{p,n}^{\alpha,\beta}(q_1, \ldots, q_n, q_{n+1}) &= \mathcal{Z}_1(\alpha,p)^{-1} e^{-\phi_\alpha(q_1) - p q_1} \\ &\times \left[\prod_{i=2}^{n} \mathcal{Z}_2(p)^{-1} e^{-\phi_\alpha(q_i - q_{i-1}) - p(q_i - q_{i-1})} \right] \mathcal{Z}_1(\beta,p)^{-1} e^{-\phi_\beta(q_{n+1} - q_n) - p(q_{n+1} - q_n)}.\end{aligned} \tag{2.8.9}$$

The convenient factorization property of the probability $\pi_{p,n}^{\alpha,\beta}(q_1, \ldots, q_n, q_{n+1})$ displayed in (2.8.9), which implies that the variables q_1 and $q_j - q_{j-1}$, $j = 2, \ldots, n+1$, are

independent, has prompted the use of the pressure ensemble for equilibrium systems, without any reference to dynamics. The requirement that particles only interact with their first neighbors is usually imposed artificially (see, however, [KUH63]). In our model this condition arises naturally from the dynamics. Note that $q_n - q_1$, the width of the fat shock, is thus represented as a sum of independent random variables; this is consistent with its Gaussian distribution in regions I and II of the fixed volume ensemble (see Remark 2.5.1).

One easily checks that (writing now $\pi^{\alpha,\beta}_{p,n} = \pi$)

$$\begin{aligned} \langle q_1 \rangle_\pi &= -\frac{d}{dp} \log \mathcal{Z}_1(\alpha) = \frac{\alpha(1+z)}{z(z+2\alpha-1)}, \\ \langle q_j - q_{j-1} \rangle_\pi &= -\frac{d}{dp} \log \mathcal{Z}_2 = \frac{1}{z}, \qquad j = 2, \dots, n, \\ \langle q_{n+1} - q_n \rangle_\pi &= -\frac{d}{dp} \log \mathcal{Z}_1(\beta) = \frac{\beta(1+z)}{z(z+2\beta-1)}. \end{aligned} \tag{2.8.10}$$

Note that when z approaches its lower bound, which is 0 if $\alpha, \beta \geq 1/2$ and $\max\{1 - 2\alpha, 1 - 2\beta\}$ otherwise, the average size $\langle q_{n+1} \rangle_\pi$ of the system goes to infinity for every $n \geq 1$; there is simply not enough pressure to confine the system. To agree with standard definitions we have defined the potentials ϕ_α in (2.8.1) so the the size of the system in the absence of boundary terms, that is, $\langle q_n - q_1 \rangle_\pi$, goes to infinity when $p \to 0$ ($z \to 0$).

To find the appropriate pressure corresponding to the canonical ensemble with $L = n/\gamma$ studied above we must set the expected system length

$$\langle q_{n+1} \rangle_\pi = \langle q_1 \rangle_\pi + \sum_{j=2}^{n} \langle q_j - q_{j-1} \rangle_\pi + \langle q_{n+1} - q_n \rangle_\pi = -\frac{d}{dp} \log \mathcal{Z}^{\alpha,\beta}(p, n), \tag{2.8.11}$$

equal to L and solve for p (or z), subject to the restrictions (2.8.7). With (2.8.10) the equation to be solved becomes

$$\frac{\alpha(1+z)}{z(z+2\alpha-1)} + \frac{n}{z} + \frac{\beta(1+z)}{z(z+2\beta-1)} = \frac{n}{\gamma}. \tag{2.8.12}$$

We will discuss the solution of this equation in various regions of the phase plane; it is useful to bear in mind that each of the three terms on the left hand side increases as z

decreases from 1 to its lower limit $\max\{0, 1-2\alpha, 1-2\beta\}$.

Consider first region I, where $\alpha, \beta > (1-\gamma)/2$. Since for $z = \gamma$ the left hand side of (2.8.12) is $n/\gamma + O(1)$, where the $O(1)$ term is positive, the solution must be of the form $z = \gamma + O(1/n)$. In the limit $n \to \infty$ we thus have $z = \gamma$ or $p = -\log(1-\gamma^2)$. In this case $\langle q_1 \rangle_\pi$ and all $\langle q_j - q_{j-1} \rangle_\pi$, $j = 2, \ldots, n+1$, are of order unity, so the bulk of the system has, in the limit $n \to \infty$, the same structure as that obtained from our NESS when $L \to \infty$ in region I.

In region II, where $\alpha < \beta$ and $\alpha < (1-\gamma)/2$, we have from $z > 1-2\alpha > 1-2\beta$ that the third term in (2.8.12) is $O(1)$, and from $z > 1-2\alpha > \gamma$ that the first term must be $O(n)$, i.e., we must have $z = 1-2\alpha+O(1/n)$. In fact we find easily that for large n,

$$z \simeq 1 - 2\alpha + \frac{2\alpha(1-\alpha)}{1-2\alpha-\gamma}\,\frac{1}{n}. \tag{2.8.13}$$

Now $\langle q_1 \rangle_\pi$ is of order n but $\langle q_{n+1} - q_1 \rangle_\pi$ is still of order 1; this corresponds to the fat shock being fixed to the right wall, i.e., to what we see in region II. The situation in region III is of course obtained by interchanging α with β and left with right. In the case when $\alpha = \beta < (1-2\gamma)/2$ one sees again that $z \simeq 1 - 2\alpha + c/n$ and that $\langle q_1 \rangle_\pi = \langle q_{n+1} - q_n \rangle_\pi$ are both of order n; this simply means that the average position of the fat shock is in the middle.

One may of course analyze the pressure ensemble directly, rather than looking for the correspondence with the open system of fixed length; the key question is how one allows p or equivalently z to vary with the number n of (second class) particles. If z is held fixed (necessarily in the range (2.8.7)) then the behavior of the system corresponds to that of the open system in region I. If $\alpha < 1/2$ and $\alpha < \beta$, and one takes $z = 1-2\alpha+c/n$ then the behavior is as in region II; similarly if $\beta < 1/2$ and $\beta < \alpha$ one obtains region III behavior by taking $z = 1-2\beta+c/n$, and if $\alpha = \beta < 1/2$ such a z value gives behavior corresponding to the II/III boundary. A detailed analysis of the ensemble, for example of the shape of the profiles of the fat shock, would essentially repeat the analysis in the fixed L, i.e., fixed overall density γ, ensemble studied earlier, and we will not take up

these questions again here.

Note that for almost all permissible values of the pressure our system is in region I; only by fine tuning the pressure to change with n in a range of order $1/n$ do we get configurations as in regions II or III. This is reminiscent of what happens when one goes from a fixed magnetization to a fixed external magnetic field h in the Ising model at low temperatures, in dimension two or higher. The whole coexistence region, corresponding to the average magnetization being smaller than the spontaneous magnetization, corresponds to the single value $h = 0$.

Other choices of z can lead to regimes different from those considered in the present work. For example, if we again suppose that $\alpha = \beta < 1/2$, but now take z closer than order $1/n$ to $1 - 2\alpha$—to be specific, say $z = 1 - 2\alpha + c/n^2$—then $\langle q_{n+1} \rangle_\pi$ is of order n^2 and hence the density of second class particles is zero.

2.9 Concluding Remarks

1. As noted already in several places above, the local properties of our system away from the boundaries approach, as $L \to \infty$, those of the states of the two species TASEP on the lattice $\mathbb{Z}$. Because of this it will be useful to describe here some known properties of the (extremal, translation invariant) NESS's of that system, i.e., of the states ν^{ρ_0,ρ_1} introduced in Section 2.6. These states differ from those of other models for which the NESS of the finite open system can be solved exactly, such as the one species simple exclusion process and the zero range processes [DJLS93, EH05, BDSG$^+$07], in that they are not product measures; this is so despite the fact that their projections (marginals) on the configurations of first class particles alone, or on the configurations of holes alone, are in fact Bernoulli. The states ν^{ρ_0,ρ_1} may be obtained [DJLS93] as the $N \to \infty$ limits of states of a two component TASEP on a ring of N sites, with $N_\alpha = \rho_\alpha N$ particles of type α, $\alpha = 0, 1, 2$, where $\rho_2 = 1 - \rho_0 - \rho_1$; see also [Spe94].

As noted in [DJLS93], the structure of the states ν^{ρ_0,ρ_1} is quite intricate, containing

several mysterious features which we still do not understand in any direct intuitive way. They are not even Gibbs measures [Spe94], even though all the truncated correlation functions involving a finite number of sites decay exponentially fast. This decay follows from the (mysterious) fact that, conditioned on the presence of a second class particle at site i, the measure factorizes: the left and right sides of i become independent. For the corresponding property for the open system studied in this paper see Remark 2.3.1. Another (mysterious) fact is that if one conditions on the presence of a second class particle at i then the particles to the right of i, and the holes to the left of i, have this Bernoulli property [DJLS93].

Another related property of the states ν^{ρ_0,ρ_1} is that if one conditions on there being a first class particle (respectively a hole) at site i then the measure to the left (respectively right) of i is the same as if there was no conditioning at all, i.e., the same as that described in the first paragraph of this section. (This may be expressed colloquially by saying that if one observes that the fastest horse is in front then one gains no information about the rest.) The property has in fact been established in not only the two species but also the n-species TASEP (see [FM07], Proposition 6.2), using a representation of the stationary measure based on queuing theory; a direct proof for the two species model may be given using the two properties of second class particles noted in the previous paragraph. We remark that the property of factorization around a second class particle does not extend in a direct way to the n-species model [FM07].

2. The fact that the measures ν^{ρ_0,ρ_1} are not product measures gives extra structure to the local states $\mu_{x,c}$ discussed in Section 2.6, which are superpositions of such measures. We note here however that as in the case of the one component TASEP, when such a superposition occurs only on the shock line $\alpha = \beta < 1/2$, the translation invariant measures which enter into the superposition (and which correspond to the measures on one side or another of a shock) all have the same current. This can be interpreted as saying that the properly averaged local current does not fluctuate. These averages can

be obtained either as long time averages of the stochastic flux across a single bond, or as spatial averages over an interval of length K, with $K \to \infty$ *after* $L \to \infty$. We believe that the convergence of the average total flux across an open system to a deterministic value, as $L \to \infty$, is a general property of the NESS of systems like those discussed here, but do not know how to prove this directly at the present time. It seems reasonable to expect similar behavior in higher dimensions and different settings, e.g., for driven diffusive systems on a torus [KLS84].

3. It follows from the "separating" property of conditioning on the presence of a second class particle at a given site that the distribution under ν^{ρ_0,ρ_1} of the second class particles alone is given by a renewal process [DJLS93, FFK94]. When the current J_2 vanishes, i.e., when $\rho_0 = \rho_1 = (1-\rho_2)/2$, then (as noted in Section 2.8) the distribution of the distance between nearest neighbor particles in this process has a simple exponential dependence on ρ_2 which can be obtained from a pressure ensemble, with $p = -\log(1-\rho_2^2)$, as in region I of the open system.

Combining this expression for the pressure as a function of the density with standard thermodynamic relations we can obtain expressions for the chemical potential λ and Helmholtz free energy a in the uniform infinite system of second class particles with density ρ_2:

$$\begin{aligned} \lambda(\rho_2) &= -\log\left(\frac{1-\rho_2}{1+\rho_2}\right), \\ a(\rho_2) &= (1-\rho_2)\log(1-\rho_2) + (1+\rho_2)\log(1+\rho_2). \end{aligned} \tag{2.9.1}$$

From (2.9.1) we can obtain the large deviation function for the probability of finding $n_2\mathcal{L}$ particles in an interval of $\mathcal{L}$ lattice sites [DLS03]. The large deviation for first class particles or holes alone is of course given by the properties of the Bernoulli measure. Large deviation properties of the full measure have not, so far as we know, been determined for the two species system.

4. Even knowing fully the properties of the infinite system still leaves open the problem

of how fast the correlations in the vicinity of a site $\lfloor xL \rfloor$ approach those in the local measure $\mu_{x,c}$. This may be of particular interest in the case when $\mu_{x,c}$ is a superposition of extremal infinite volume measures ν^{ρ_0,ρ_1} as discussed in Section 2.6. We might expect the L dependence in that case, where typical density profiles differ from average ones, to be different from that where the two coincide. We leave this as an open problem.

5. We now briefly describe two related model systems, containing both first and second class particles on a ring, which are intermediate between those studied in [DJLS93] and in this paper.

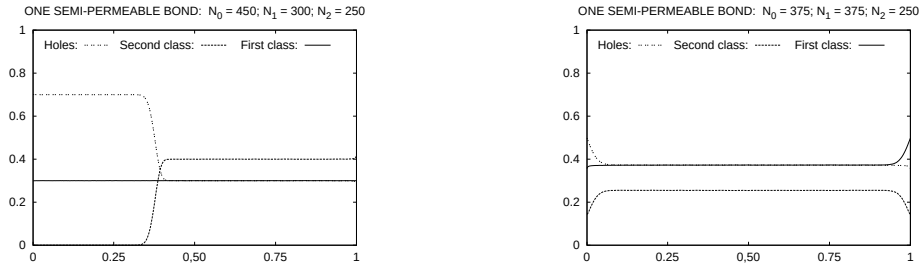
5.1 Recall the "truck" or "defect particle" model [Der96, Mal96, DE99, BE07], a standard two species TASEP system: a single defect particle together with (first class) particles and holes, on a ring of $L+1$ sites, can exchange with a hole ahead of it (clockwise) at rate α and with a particle behind it at rate β. Let us add to the ring also n (standard) second class particles, which make the same exchanges as does the defect particle but at rate 1 in each case, and which do not exchange at all with the defect particle. To be definite let us say that there are n_1 first class particles and n_0 holes on the ring, with $n + n_1 + n_0 = L$. Then the stationary measure for this system is almost the same as that of our open system: using the matrices X_0, X_1, and X_2 of Section 2.3, and introducing also $X_\delta = |V_\beta\rangle\langle W_\alpha|$, we find that a configuration $\delta, \tau_1, \ldots, \tau_L$, where δ represents the defect particle, has weight:

$$\mathrm{Tr}(X_\delta X_{\tau_1} \cdots X_{\tau_L}) = \langle W_\alpha | X_{\tau_1} \cdots X_{\tau_L} | V_\beta \rangle, \tag{2.9.2}$$

(compare (2.3.4)). The difference, of course, is that this is a canonical ensemble and the partition function must be obtained by summing the weights over only those configurations with the proper numbers of all species. This relation between this truck model on a ring of $L+1$ sites and the two species open system studied in this paper is completely parallel to that between the usual defect particle model and the open one species TASEP. We expect that details of the stationary state could be worked out in

parallel to that of the usual defect particle model, but we have not done so.

5.2 In the second model, the ring has N sites labeled by $i \in [-N/2+1, N/2]$ and contains $N_1 = \bar{\rho}_1 N$ first class particles, $N_2 = \bar{\rho}_2 N$ second class particles, and $N_0 = N - N_1 - N_2 = \bar{\rho}_0 N$ holes. The particles jump clockwise according to the TASEP rules given in section 2.2, *except* at one specified semi-permeable bond, say between site 0 and site 1, which prohibits the passage of second class particles. (We can think of this bond as a restriction in a channel).



(a) (b)

Figure 2.5: Density profiles in a system with one semi-permeable bond: $L = 1000$.

Unfortunately we do not have an exact solution for this system. To see what happens, however, we note that, as in the open system, the current J_2 of second class particles will vanish in the stationary state. On the other hand, since we would have $J_2 = \bar{\rho}_2(\bar{\rho}_0 - \bar{\rho}_1)$ if the system were uniform, a uniform state is possible only if $\bar{\rho}_1 = \bar{\rho}_0$. If $\bar{\rho}_1 < \bar{\rho}_0$ then J_2 would be positive in the uniform system and second class particles would drift to the right; the upshot is that there will be a fat shock of width $w = \bar{\rho}_2 N/(1 - 2\bar{\rho}_1)$ containing all second class particles at density $\rho_2 = 1 - 2\rho_1$ pinned to the back of the barrier. If $\bar{\rho}_1 > \bar{\rho}_2$ then the fat shock of width $w = \bar{\rho}_2 N/(1 - 2\bar{\rho}_0)$ will be pinned to the front of the barrier. In the case $\bar{\rho}_0 = \bar{\rho}_1 = (1 - \bar{\rho}_2)/2$ the system will be uniform. See Figure 2.5 for some typical profiles in this system; note that N_0, N_1, and N_2 have been chosen so that the bulk densities in Figures (a) and (b) here are the same as those in Figures (a) and (e) of Figure 2.3, but that the boundary effects and finite density shock transition are noticeably different in the two models.

Letting $N \to \infty$ with $\bar{\rho}_1, \bar{\rho}_0$ fixed would yield an infinite system with a barrier at the bond $(0, 1)$. Consider first the case $\bar{\rho}_1 < \bar{\rho}_0$. To the right of the origin there would be no second class particles and a uniform density of first class particles described by a product measure. Far to the left of the barrier the state would be $\nu^{\bar{\rho}_0, \bar{\rho}_0}$, i.e., a uniform state with $\rho_1 = \rho_0 = \bar{\rho}_1$ and $\rho_2 = 1 - 2\bar{\rho}_1$. We do not know, however, the structure of the system just to the left of the barrier. Similar conclusions hold for $\bar{\rho}_1 > \bar{\rho}_0$.

Chapter 3

Directed Lattice Walks

3.1 Introduction

A directed walk on a lattice is a walk consisting of steps which have a strictly positive inner product with a given vector. In other words, all steps move toward a fixed direction. Directed walks are very useful models in statistical mechanics. They have been used to study various kinds of problems. In the presence of disorder they have been used to study the DNA denaturation transition [PS66b, PS66a] and the wetting transition [DHV92]. Without disorder, they have been used to study the critical temperature of the two dimensional Ising model [Tem56]. Our motivation is the study of polymers held between two close plates [vR00].

Consider a long linear polymer in dilute solution constrained between two plates. Naïvely one would expect the polymer to exert a force on the plate, simply because such a configuration is not entropically favorable. However, this can change if there is an adsorptive interaction between the plates and the polymer. In the latter case, as we tune the strength of this interaction, a phase transition occurs which flips the overall interaction. Calculation of the point at which the phase transition occurs is one of the major concerns of this work.

One of the advantages of modelling the polymer by a directed walk is that the latter is self-avoiding by definition and can therefore represent the rough configuration of the polymer even though the microscopic details may differ. So far this framework has been used to study walks with a specific set of steps given by $\{(1,1),(1,-1)\}$ [BORW05]. In

general, this is not quite satisfactory for a couple of reasons. First, this set of steps allows the polymer to move at an angle of $\pm 45^\circ$ only and thus severely restricts its configuration space. For example, a better model would be $\{(1,2),(2,1),(1,-2),(2,-1)\}$. Secondly, one would like to consider polymers with different bond lengths. In other words, suppose that the polymer has various molecules $\{X_i\}$ and the values of the bond lengths between the various species is $\{l_{ij} = |X_i - X_j|\}$, where $|X_i - X_j|$ denotes the bond length between X_i and X_j. Then one would need to examine a different set of steps.

Suppose that an experimentalist is working on a specific physical polymer wedged between two plates and would like to understand the behavior of certain physical quantities as certain other parameters are varied. One way she can understand the system is to model the polymer as a directed walk. Apriori, there could be many different sets of steps one could consider depending on the nature of the polymer, and each one would lead to a different numerical estimate of these physical quantities. She would like to have a black box which would take as input a certain set of steps and the width w between the plates and get as output physically measurably quantities. One of the motivations behind this work is to create such a black box.

Until now, the study of the constrained walk with an arbitrary set of bond lengths has been daunting because there are no general results in this direction. We present a toolbox of algorithms (implemented in a Maple package named *POLYMER*) which can be used to calculate various quantities of interest, such as the generating function of the number of constrained walks starting and ending on one of the plates and the free energy of the polymer, as well as to plot quantities such as the force exerted by the polymer on the plates in different regimes. We emphasize that these calculations can be done for *any* choice of steps, at least in principle. In practice, of course, one is restricted by the limited resources of the largest computers.

We should mention that, unknown to us, the equations for infinite width walks were already written down, albeit in slightly different language, in [Duc00]. While this work

was in preparation, there also appeared [BM07] where the topic of constrained walks with arbitrary steps is also treated. The flavor of the work is somewhat different there. While the results there are of much greater generality, they are existence results. The results here are more algorithmic in nature.

3.2 The Setting

Consider a walk constrained in the two dimensional square lattice $\mathbb{Z}^2$ by $0 \leq y \leq w$ where w is the width. In Fig. 3.1 we show a walk with width $w = 4$ involving the steps $(1, 1)$ and $(1, -1)$.

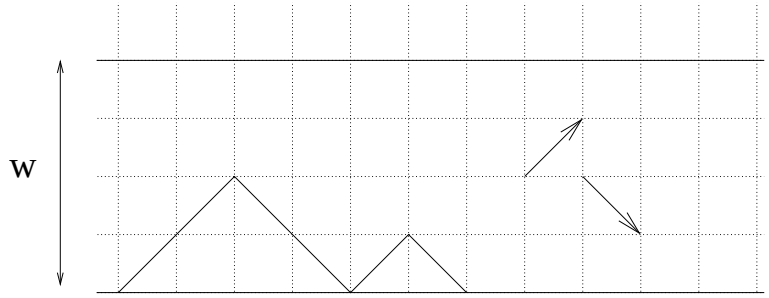


Figure 3.1: A walk with width $w = 4$ involving the steps $(1, 1)$ and $(1, -1)$. The steps are shown with arrows on the right.

Such a walk can always be reinterpreted as a walk in the region given by $0 \leq x - y \leq w, y \geq 0$. This is done simply by rotating and reflecting the above figure in the appropriate way. The line $x - y = 0$ is the same as $y = 0$ above and the line $x - y = w$ becomes $y = w$. In Fig. 3.2 the same walk is redrawn for this situation. Notice that the steps have also been rotated. They have become $[1, 0]$ and $[0, 1]$. The former walks are known as *Dyck paths* and because the latter represent the so-called ballot problem, we call them *ballot paths*. For convenience, we represent Dyck steps with the usual parentheses $(,)$ and ballot steps with square parentheses $[,]$. Points on the lattice are always denoted by the usual parentheses $(,)$.

Since we are interested in directed walks, the steps should have a positive inner

product with the vector $(1,0)$ for Dyck paths and the vector $[1,1]$ for ballot paths.

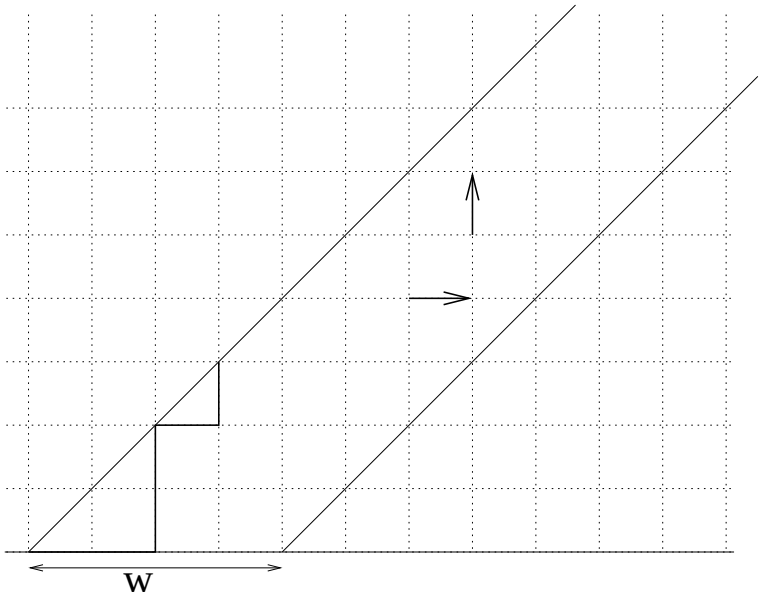


Figure 3.2: The same walk shown in Figure 3.1 with $w = 4$ and steps $[1,0]$ and $[0,1]$ interpreted as a ballot walk. The steps are shown with arrows on the right.

3.3 Calculating Walks

The main idea in counting the number of walks with a general set of steps is *recursion*.

3.3.1 Simple Walks

Consider the number of ballot paths from $(0,0)$ to the point (m,n) which we denote as $c(m,n)$. Let us first consider simple steps. For example, consider the steps $[1,0]$ and $[0,1]$. Then $c(m,n)$ satisfies the simple recurrence relation

$$\begin{aligned} c(m,n) &= 0 \quad \text{if } m<0, n<0, m-n<0 \text{ or } m-n>w, \\ c(m,n) &= c(m,n-1)+c(m-1,n) \quad \text{if } m>n>0, \end{aligned} \tag{3.3.1}$$

because, if $m > n > 0$, one can arrive at the point (m,n) either by the step $[1,0]$ from $(m-1,n)$ or by the $[0,1]$ steps from $(m,n-1)$ and hence the number of such walks

simply adds. On the other hand, one can never reach (m, n) if $m < 0, n < 0, m - n < 0$ or $m - n > w$. Thus, we only need the initial condition $c(0, 0) = 1$ (the null walk) to determine all walks. For example,

$$\begin{aligned} c(2,1) &= c(2,0) + c(1,1) \\ &= (c(1,0) + c(2,-1)) + (c(1,0) + c(0,1)) \\ &= (c(0,0) + 0) + (c(0,0) + 0) \\ &= 2 \end{aligned} \tag{3.3.2}$$

and one can easily check that there are only two walks from $(0, 0)$ to $(2, 1)$. Let us take a more complicated example. Suppose the steps are $[1, 0], [0, 2]$ and $[1, 1]$. Then the recurrence becomes more complicated and is given by

$$\begin{aligned} c(m,n) &= 0 \quad \text{if } m < 0, n < 0, m - n < 0 \text{ or } m - n > w, \\ c(m,n) &= c(m, n-2) + c(m-1, n) + c(m-1, n-1) \quad \text{if } m > n. \end{aligned} \tag{3.3.3}$$

When there is yet another constraint given by $x - y \leq w$, we simply need to put in another "if" condition and the main recurrence relation remains unchanged.

This idea can be implemented as an algorithm as follows: Let *Steps* represent the set of all possible steps and let $Steps[i]$ denote the ith step. Then the number of walks $c(m, n)$ from $(0, 0)$ to (m, n) with width w is calculated as follows:

if $m = n = 0$ then

RETURN 1

if $m + n < 0$ then

RETURN 0

if $m - n < 0$ or $m - n > w$ then

RETURN 0

if $m > n > 0$ and $m - n < w$ then

$Prev = \{(m, n) - Steps[i] | i = 1, 2, ...\}$

$$RETURN \sum_i c(Prev[i])$$

3.3.2 Walks with Boundary Interactions

In combinatorics, boundary interactions can be implemented with variables t, s which "record" the number of times the walk touches the left and right walls respectively. More precisely, each walk is assigned a monomial in t and s where the degree of the monomial in $s(t)$ is the number of times the walk touches the left wall (right wall). The initial point does not count.

The reason for doing this is one can see the effect of this boundary interaction on physical quantities like the force exerted by the polymer on the plates. For example, if the polymer interacts very strongly with the right wall and weakly with the left wall, meaning s is very large and t very small, then walks which touch the right wall count for much more than those which just touch the left wall. This in turn increases the probability that the polymer is close to the right wall. Hence, physically, the polymer exerts very little force on the left wall. For example, consider walks from $(0, 0)$ to $(2, 2)$ with steps $[1, 0], [0, 1]$. There are only two possibilities as shown.

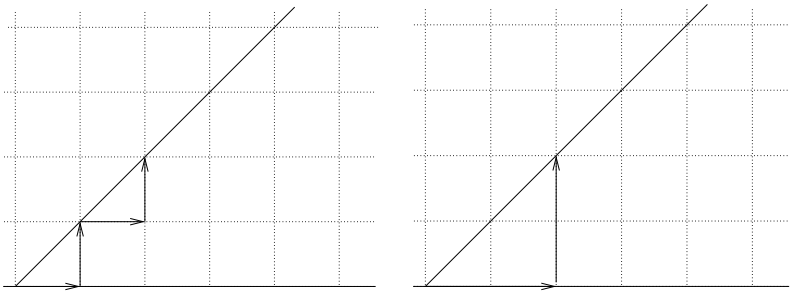


Figure 3.3: The only possible walks from $(0, 0)$ to $(2, 2)$ using the steps $[0, 1]$ and $[1, 0]$

When $w = 1$, only the left walk is permitted and carries a weight of t^2s^2 because it touches both walls twice. When $w = 2$, both walks are permitted but carry different weights. The left one carries a weight of t^2 and the right one, a weight of ts. When $w = 3$, the left one still carries a weight of t^2 but the right one does not touch the right

wall at all and therefore carries a weight of t. Thus,

$$c_1(t,s) = t^2 s^2 \tag{3.3.4}$$

$$c_2(t,s) = t^2 + ts \tag{3.3.5}$$

$$c_3(t,s) = t^2 + t \tag{3.3.6}$$

Setting $t = s = 1$ gives the number of such walks. The algorithm representing these walks closely resembles simple walks.

if $m = n = 0$ then

RETURN 1

if $m + n < 0$ then

RETURN 0

if $m - n < 0$ or $m - n > w$ then

RETURN 0

if $m - n = 0$ then

$$Prev = \{(m,n) - Steps[i] | i = 1, 2, ...\}$$

$$RETURN \sum_i (t \cdot c(Prev[i]))$$

if $m - n = w$ then

$$Prev = \{(m,n) - Steps[i] | i = 1, 2, ...\}$$

$$RETURN \sum_i (s \cdot c(Prev[i]))$$

if $m > n > 0$ and $m - n < w$ then

$$Prev = \{(m,n) - Steps[i] | i = 1, 2, ...\}$$

$$RETURN \sum_i (c(Prev[i]))$$

3.4 Generating Functions

The generating function is an important tool in combinatorics. It is another way to package the same information as brute-force counting. It is a formal power series whose

coefficients in the Taylor series give precisely the count. More specifically, if $c_w(n)$ is the number of ballot walks from $(0,0)$ to (n,n) with width w, the generating function $\phi_w(z)$ is given by

$$\phi_w(z) = \sum_{n=0}^{\infty} c_w(n) z^n \tag{3.4.1}$$

We demonstrate the calculation of the generating function for both the case of finite width and infinite width. The ideas involved in both are quite different and so they will be treated differently. However, both involve a set of tricks commonly used in combinatorics. We describe them in detail in subsequent sections.

3.4.1 Finite width

We use the same ideas described above to calculate the generating function of the number of walks with any set of steps.

We first start with a nontrivial example. We will spell out all the details here. Consider the steps $[1,0],[0,2],[1,1]$ and width, $w = 2$. We define three generating functions $\phi_i(z), i = 0,1,2$ for this problem, where $\phi_i(z)$ counts the number of walks from $(0,0)$ to $(n+i,n)$.

In terms of initial conditions, only ϕ_0 has a nontrivial condition, namely the zero-step walk from $(0,0)$ to $(0,0)$. Now, let us consider each of the generating functions one at a time.

For a walk ending at (n,n), the last step can have two possibilities. It can either end with a $[1,1]$ step from $(n-1,n-1)$ (which is also described in terms of ϕ_0) or it can end with a $[0,2]$ step from $(n,n-2)$ (which is described by ϕ_2). Since the weight of the walk depends only on the y-coordinate, we have the following equation for ϕ_0

$$\phi_0 = 1 + z\phi_0 + z^2\phi_2 \tag{3.4.2}$$

Now, for a walk ending at $(n+1,n)$, there are two ending possibilities. The final step can be $[1,0]$ from (n,n) (described by ϕ_0) or it can be $[1,1]$ from $(n,n-1)$ (described

by ϕ_1). The $[0,2]$ step is not possible because the walk would have to start outside the prescribed strip. Thus,

$$\phi_1 = z\phi_1 + \phi_0 \tag{3.4.3}$$

Finally, a walk ending at $(n+2, n)$ also has two possibilities. The final step can be $[1,0]$ from $(n+1, n)$ (described by ϕ_1) or it can be $[1,1]$ from $(n+1, n-1)$ (described by ϕ_2).

$$\phi_2 = \phi_1 + z\phi_2 \tag{3.4.4}$$

These are now three linear equations in three variables. These, when solved, for ϕ_0 gives the rational function

$$\phi_0 = \frac{1 - 2z + z^2}{1 - 3z + 2z^2 - z^3} \tag{3.4.5}$$

This example contains the essence of the argument to follow. For any set of steps and width w, we will always have $w+1$ linear equations in the variables $\phi_0, \cdots, \phi_w$ independent of the number and kinds of steps involved. These equations will be linear precisely because the ending of each walk contributing to ϕ_i takes its last step starting from some other walk contributing to, say, ϕ_j. To be precise, let us denote the set of steps simply as $Steps$. We will let $Steps[i]_x$ and $Steps[i]_y$ denote the x and y components of the ith element of $Steps$. Then, for $i = 0, \cdots, w$,

$$\phi_i \quad = \quad \delta_{i,0} + \sum_{\substack{j \\ 0 \le i + Steps[j]_y - Steps[j]_x \le w}} z^{Steps[j]_y} \phi_{i + Steps[j]_y - Steps[j]_x} \tag{3.4.6}$$

Solving this system will always lead to *rational* solutions for each of the generating functions ϕ_i. And modulo unexpected cancellations, they all have the same denominator - the determinant of the corresponding matrix in the linear system!

3.4.2 Finite Width with Boundary Interactions

We can also calculate the generating function of walks with variables t, s (called *weight enumerators*) using essentially the same idea.

Let us consider the same example with steps $[1,0],[0,2],[1,1]$ and width, $w = 2$. We again have three generating functions $\phi_i(z), i = 0,1,2$. Since we have a factor of t everytime we touch the left wall, the equation for ϕ_0 is

$$\phi_0 = 1 + tz\phi_0 + tz^2\phi_2 \tag{3.4.7}$$

The equation for ϕ_1 is unchanged because walks contributing to it are touching neither of the two walls at the last step.

$$\phi_1 = z\phi_1 + \phi_0 \tag{3.4.8}$$

The equation for ϕ_2 is affected because walks contributing to it are exactly on the right wall at the final step.

$$\phi_2 = s\phi_1 + sz\phi_2 \tag{3.4.9}$$

Solving this gives

$$\phi_0 = \frac{(1-z)(1-sz)}{1-z-sz-tz+sz^2+tz^2-stz^3} \tag{3.4.10}$$

As this example shows, the generating functions for these walks can also be calculated exactly using the same technique as in the previous section. In fact, the equations are quite similar unweighted enumeration. For the extreme generating functions, the modified equations look like

$$\phi_0 = 1 + \sum_{\substack{j \\ 0 \le Steps[j]_y - Steps[j]_x \le w}} tz^{Steps[j]_y}\phi_{Steps[j]_y - Steps[j]_x} \tag{3.4.11}$$

$$\phi_w = \sum_{\substack{j \\ 0 \le w + Steps[j]_y - Steps[j]_x \le w}} sz^{Steps[j]_y}\phi_{w+Steps[j]_y - Steps[j]_x} \tag{3.4.12}$$

while for the remainder, ie $i = 1, \cdots, w-1$, the equations remain the same.

$$\phi_i = \sum_{\substack{j \\ 0 \le i + Steps[j]_y - Steps[j]_x \le w}} z^{Steps[j]_y}\phi_{i+Steps[j]_y - Steps[j]_x} \tag{3.4.13}$$

3.4.3 Infinite Width

We have to manipulate generating functions in a different way to calculate them for the infinite width walks. We will need a number of definitions for this purpose.

First, define an $[ij]$ *walk* as one which starts on the line $x - y = i$ and ends on the line $x - y = j$. Since we have infinite width, both i, j are unbounded. Denote the generating function of such walks by $f^{[ij]}(z)$.

Define an *irreducible* $[ij]$ *walk* as a walk which, as before, starts on the line $x-y = i$ and ends on the line $x-y = j$ with the restriction that it touches the line corresponding to the minimum of i and j only at the corresponding endpoint. Denote the generating function of such an irreducible walk by $g^{[ij]}(z)$.

Now the idea is to relate these generating functions for different values of i and j where both range from 0 to a certain finite value depending on the kind of steps.

Consider the following set of steps: $\{[0, 1], [1, 0], [2, 0], [0, 2]\}$. First, a [00] walk is either the empty walk or it is composed of an irreducible [00] walk followed by a smaller [00] walk.

$$f^{[00]} = 1 + f^{[00]}g^{[00]} \tag{3.4.14}$$

Next, a [01] walk is always uniquely composed of an arbitrary [00] walk followed by an irreducible [01] walk. Similarly, a [10] walk is uniquely composed of an irreducible [10] walk followed by an arbitrary [00] walk.

$$f^{[01]} = g^{[01]}f^{[00]} \tag{3.4.15}$$

$$f^{[10]} = g^{[10]}f^{[00]} \tag{3.4.16}$$

A [11] walk either never goes below the first level, in which case it is simply the same as a [00] walk, or if it does, it is composed of an irreducible [10] walk followed by

an arbitrary [01] walk.

$$f^{[11]} = f^{[00]} + g^{[10]} f^{[01]} \tag{3.4.17}$$

Now, we go on to describe the irreducible walks. In each case, we have to consider different cases for the starting step and the ending step. First, an irreducible [00] walk can begin with either the $[1,0]$ or $[2,0]$ step and end with either the $[0,1]$ or $[0,2]$ step. If the walk starts with $[1,0]$ and ends with $[0,1]$, then there could be an arbitrary [00] walk in between. If the walk starts with $[1,0]$ and ends with $[0,2]$, there has to be an arbitrary [01] walk in between. If the walk starts with $[2,0]$ and ends with $[0,1]$, there has to be an arbitrary [10] walk in between. And finally, if the walk starts with $[2,0]$ and ends with $[0,2]$, there is a [11] walk in between. For each of these cases only the y-coordinate of the steps give the corresponding powers of z.

$$g^{[00]} = zf^{[00]} + z^2 f^{[01]} + zf^{[10]} + z^2 f^{[11]} \tag{3.4.18}$$

For an irreducible [01] walk, we just need to consider the starting steps. If it starts with $[1,0]$, the remainder is an arbitrary [00] walk. If it starts with $[2,0]$, the remainder is again an arbitrary [10] walk. A very similar argument on the ending step yields the equation for an irreducible [10] walk.

$$g^{[01]} = f^{[00]} + f^{[10]} \tag{3.4.19}$$

$$g^{[10]} = zf^{[00]} + z^2 f^{[01]} \tag{3.4.20}$$

This finally gives the desired seven equations in seven variables. Notice that all the equations are algebraic and therefore the solution of this system will also be algebraic, ie, the solution of a polynomial equation. We are ultimately interested in $f^{[00]}$ and in

this case, eliminating the other variables and replacing $f^{[00]}$ by F gives

$$z^4F^4 - 2z^3F^3 - z^2F^3 + 2z^2F^2 + 3zF^2 - 2zF + 1 - F = 0 \tag{3.4.21}$$

For the generic set of steps, we describe the algorithm. Let

$$m = \max_i(|Steps[i]_x - Steps[i]_y|) \tag{3.4.22}$$

Then we define $f^{[ij]}$ and $g^{[ij]}$ for $i, j = 0, 1, \cdots, m-1$. It is these generating functions we will write down equations for. It turns out the equations for $f^{[ij]}$ are almost completely independent of the kind of steps. In particular, only $f^{[00]}$ depends on whether there are any steps of the form $[k, k]$.

$$f^{[00]} = 1 + g^{[00]}f^{[00]} + f^{[00]} \cdot \sum_{(k,k)\in Steps} z^k \tag{3.4.23}$$

For $f^{[ll]}, l = 1, \cdots, m-1$, the equation is

$$f^{[ll]} = f^{[00]} + \sum_{k=1}^{l} g^{[l-k,0]}f^{[0,l-k]} \tag{3.4.24}$$

and for the remainder,

$$f^{[ij]} = \begin{cases} \sum_{k=1}^{i} f^{[i-k,0]}g^{[k,j]} & \text{for } i < j \\ \sum_{k=1}^{j} f^{[0,j-k]}g^{[i,k]} & \text{for } i \geq j \end{cases} \tag{3.4.25}$$

For the irreducible generating functions, we need to specify only $g^{[0j]}$ and $g^{[j0]}$ because

$$g^{[ij]} = \begin{cases} g^{[0,j-i]} & \text{for } i < j \\ g^{[i-j,0]} & \text{for } i \geq j \end{cases} \tag{3.4.26}$$

simply by definition. For $g^{[00]}$, we need to consider both starting and ending steps and sum on all possible combinations of these.

$$g^{[00]} = \sum_{\substack{(i,j) \\ X[i],Y[j]>0}} z^{Steps[i]_y + Steps[j]_y} f^{[X[i],Y[j]]} \tag{3.4.27}$$

where $X[i] = Steps[i]_x - Steps[i]_y - 1, Y[j] = Steps[j]_y - Steps[j]_x - 1$.

For the remaining irreducible generating functions, we will need to sum over either on the starting steps or on the ending steps depending on whether we are considering $g^{[0i]}$ or $g^{[i0]}$.

$$g^{[0i]} = \sum_{\substack{k \\ Steps[k]_x - Steps[k]_y > 0}} z^{Steps[k]_y} f^{[Steps[k]_x - Steps[k]_y - 1, i-1]} \tag{3.4.28}$$

$$g^{[i0]} = \sum_{\substack{k \\ Steps[k]_y - Steps[k]_x > 0}} z^{Steps[k]_y} f^{[i-1, Steps[k]_y - Steps[k]_x - 1]} \tag{3.4.29}$$

Thus we have, for a given m, $2m^2$ algebraic equations in as many variables and we should be able to eliminate everything except $f^{[00]}$ and obtain a single polynomial equation in $f^{[00]}$.

3.4.4 Infinite Width with Boundary Interactions

Just as for generating functions of finite width, it is possible to automate the calculation of weight enumerators with one parameter t counting the number of times the walk touches the line $y = x$.

The algorithm here is similar to that of the previous section and in fact, we will need the generating functions $f^{[ij]}$ calculated earlier for the same set of steps.

We begin with the same example as in the previous section, namely with steps $\{[0,1],[1,0],[0,2],[2,0]\}$. Now, let $H(z,t)$ be the weight enumerator. Notice that the only difference between H and $f^{[00]}$ is this extra parameter t in the former. We need only one extra equation for H apart from the other ones. This one is almost exactly like the one for $f^{[00]}$. Namely,

$$H = 1 + t\left(zf^{[00]} + z^2 f^{[01]} + zf^{[10]} + z^2 f^{[11]}\right) H. \tag{3.4.30}$$

All this is saying is that H is either the empty walk or it is composed of all possible starting and ending steps with a factor of t contributing towards the ending step followed

by a smaller such walk. The calculations of $f^{[ij]}$ have to be done using the same equations as in the previous section.

This gives the following algebraic equation for H.

$$\begin{aligned}
&1 + 3t^2H^2 - 2tzH + 3tH - 4H + t^3H^3 + 3t^3z^2H^3 + 2t^2z^2H^2 + 6tzH^2 \\
&-3t^2zH^2 - 3t^3z^2H^4 + t^4z^2H^4 + 2t^2z^2H^4 + 2t^3z^3H^4 - 2t^4z^3H^4 - 2t^3z^3H^3 \\
&+t^4z^4H^4 + t^4zH^4 - 6H^3tz - 3t^2H^4z - 9tH^2 + 6H^2 + 2H^4tz - 6t^2H^3 \\
&+H^4 + 9H^3t - 4H^3 - t^3H^4 + 3H^4t^2 - 3H^4t + 6t^2zH^3 - 4t^2z^2H^3 = 0
\end{aligned} \tag{3.4.31}$$

which is of fourth order in H, just as before.

When there is a general set of steps, we have the same equations as before with an additional equation for H,

$$H = 1 + t \cdot \left(\sum_{(k,k) \in Steps} z^k \right) H + tg^{[00]}H. \tag{3.4.32}$$

Solving this system will give us the required algebraic equation for H.

3.5 Empirical Guessing

We say that a sequence of rational functions $\{F_w(z)\}_{w=0}^{\infty}$ belongs to $\mathcal{A}_L(w,z)$ if each function $F_w(z)$ can be written in the form

$$F_w(z) = \frac{P_w(z)}{P_{w+1}(z)} \tag{3.5.1}$$

where the polynomials $\{P_w(z)\}_{w=0}^{\infty}$ satisfy a recurrence of order L in w with constant (in w) coefficients,

$$\sum_{i=0}^{L} a_i(z) P_{w+i}(z) = 0 \tag{3.5.2}$$

For any given set of steps, one can empirically check if the generating functions $\phi_w(z)$ belong to the class $\mathcal{A}_L(w,z)$. This is done by using the holonomic ansatz [Zei07b] and searching for a recurrence of order L among the numerators $P_w(z)$.

For most steps, it does turn out that the numerator of $\phi_w(z)$ is precisely the denominator of $\phi_{w-1}(z)$. The generating functions of many classes of steps do turn out to belong to $\mathcal{A}_L(w,z)$ for some L. For example, steps of the form $\{[0,1],[n,0]\}$ always lead to generating functions which belong to $\mathcal{A}_{n+1}(w,z)$. Yet another nontrivial example is as follows: Steps of the form $\{[0,1],[1,0],[n,n+2],[n+2,n]\}$ with $n \geq 0$ belong to $\mathcal{A}_5(w,z)$ [AZ07a]!

It is an open problem whether the generating functions of all set of steps belong to $\mathcal{A}(w,z)$.

3.6 Free Energy

One can define the free energy for this system as follows. Let $c_w(n)$ denote the number of walks from $(0,0)$ to (n,n). Then the free energy κ_w is defined by

$$\kappa_w = \lim_{n\to\infty} \frac{1}{n} \log c_w(n) \tag{3.6.1}$$

Since the generating function for the walk with any fixed w is rational, this is the negative of the logarithm of the smallest real positive zero of the denominator.

There are well-established algorithms [SZ94, Com64], to find the recurrence relation satisfied by the sequence $\{a_n\}$ given the generating function $\phi(z) = \sum a_n z^n$. In Maple, this is implemented in the package titled GFUN [SZ94]. Now, given this recurrence relation, there exists an algorithm [Bir30, BT33, WZ85] to find the asymptotic behaviour of the generating function. This is implemented in the package GuessHolo2 [Zei07b]. One can therefore calculate the free energy for walks with a given set of steps at any finite width as well as for infinite width.

One can also calculate the free energy as a function of the variables t, s. Unfortunately there is no explicit formula for the free energy as a function of t, s, but given any specific values of t and s, these can be calculated exactly. This is mainly because we do not have a general formula for the roots of polynomials of degree ≥ 5. For the same

reason, one cannot calculate the free energy for the infinite width case as a function of the parameter t.

3.7 Force

The force exerted on the plates is given, in the discrete case, by

$$F(w) = \kappa_{w+1} - \kappa_w \tag{3.7.1}$$

In the limit when w is very large, the force is defined by the derivative of the free energy with respect to w. One can also calculate the force for any specific values of the variables t, s. For the same reason as in the previous section, this calculation cannot be done keeping t and s arbitrary.

The structure of the phase diagram is most clearly seen by plotting the force as a function of t, s. The region where the force is positive is the desorbing region and where the force is negative is the adsorbing region.

3.8 Examples

We use the algorithms outlined above and some others to calculate quantities of interest for different sets of steps. We emphasize that these are not individually calculated for these particular set of steps but are simply outputs of algorithms described in previous sections. These algorithms are implemented in Maple.

3.8.1 $\{(1,1),(1,-1)\}$ Steps

We repeat some of the calculations in [BORW05] with standard Dyck steps to demonstrate the power of this approach. The corresponding ballot steps are $[0,1],[1,0]$.

The generating function at any finite width can be calculated for even reasonably large widths in short enough times. For example, when $w = 3$,

$$\phi_3(z) = \frac{1-2z}{1-3z+z^2} \tag{3.8.1}$$

We can calculate the equation satisfied by the infinite width generating function.

$$-F(z) + 1 + zF(z)^2 = 0 \tag{3.8.2}$$

Let $P_w(z), Q_w(z)$ be the numerators and denominators of $\phi_w(z)$. One can conjecture a recurrence relation in w for the Q_w's. For these steps, it turns out that

$$zQ_w(z) - Q_{w+1}(z) + Q_{w+2}(z) = 0 \tag{3.8.3}$$

Figure 3.4 shows the free energy plotted for a range of widths. We also plot the free energy for a particular width for a set of weight-enumerating parameters in Figure 3.5. Notice that even such a small value of w shows the characteristics of the phase diagram in Figure 7 of [BORW05]. For $0 \leq t, s \leq 2$, the value of the free energy is more or less constant and outside, it seems to grow more or less linearly and we can clearly notice the non-analyticity at the line $t = s$.

We can also plot the free energy of the infinite width case as a function of the parameter t which counts the number of times the walk touches the diagonal $x = y$. This is done in Figure 3.6. To see how the phase diagram looks, one can also plot the force for a reasonable large value of the width. As seen in Figure 3.7, this looks like the derivative of Figure 3.5. There is a strong positive force in the region $0 \leq t, s \leq 2$, which is the desorbed region. There is also the clear attraction regime around the line $t = s$ as shown in Figure 8 of [BORW05].

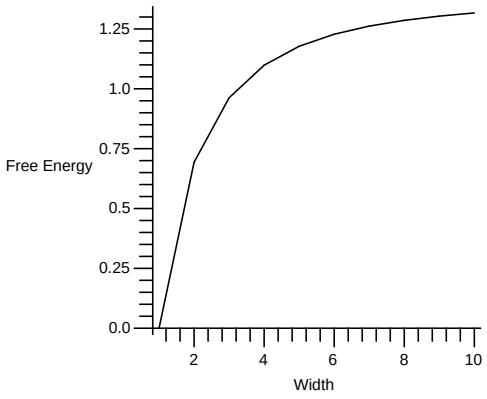


Figure 3.4: Plot of free energy for steps $\{[0,1],[1,0]\}$ for different widths

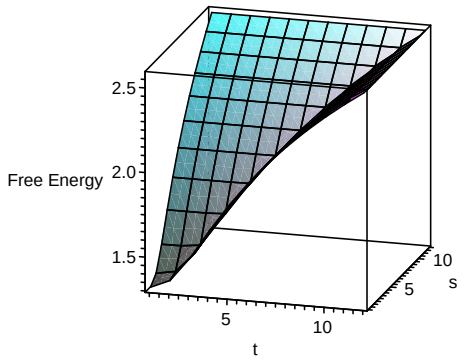


Figure 3.5: Plot of free energy for steps $\{[0,1],[1,0]\}$ at a fixed width as a function of boundary parameters

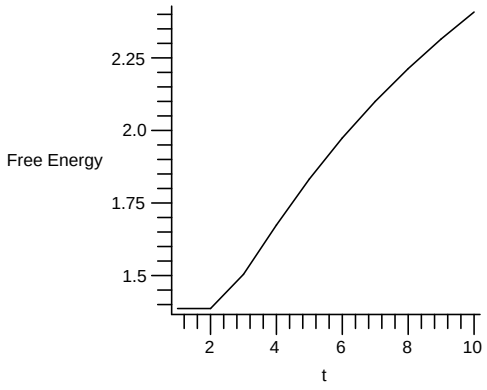


Figure 3.6: Plot of free energy for steps $\{[0,1],[1,0]\}$ as a function of the boundary parameter

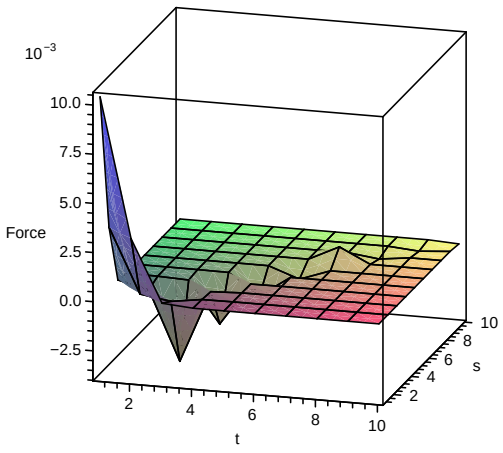


Figure 3.7: Plot of force for steps $\{[0,1],[1,0]\}$ at a fixed width as a function of boundary parameters

3.8.2 $\{(1,1),(2,2),(1,-1),(2,-2)\}$ Steps

For a more nontrivial example, consider the following ballot steps: $\{[1,0],[2,0],[0,1],[0,2]\}$ [AZ07a]. Figure 3.8 shows the free energy plotted for a range of widths. Notice that the behaviour is very similar to that of the previous example in Figure 3.4. We know that the behavior must be monotonically increasing with the width and is a concave function, which is true for both figures.

We also plot the free energy for a particular width for a set of weight-enumerating parameters in Figure 3.9. Here too, the free energy increases with increasing values of t, s. However, unlike Figure 3.5, there is no apparent loss of analyticity. That might be because the width is too small here.

To see how the phase diagram looks, one can also plot the force for a reasonable large value of the width. Figure 3.10 shows this. The sheet in this figure is smoother than in the analogous sheet in the previous example Figure 3.7. And there is again a similarity between the two figures simply because there is a similarity between the corresponding plots for the free energy.

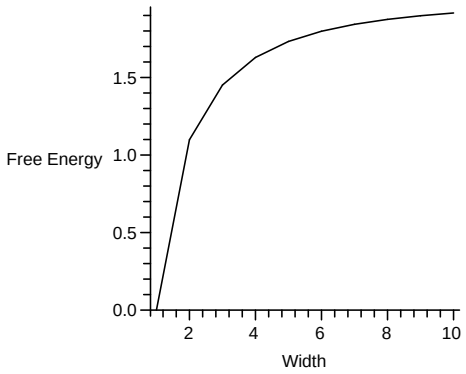


Figure 3.8: Plot of free energy for steps $\{[0,1],[1,0],[2,0],[0,2]\}$ as a function of the width

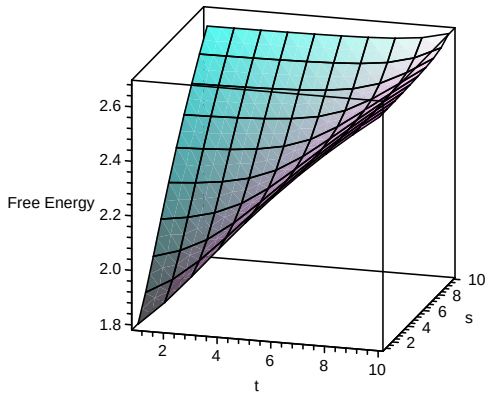


Figure 3.9: Plot of free energy for steps $\{[0,1],[1,0],[2,0],[0,2]\}$ at a fixed width as a function of boundary parameters

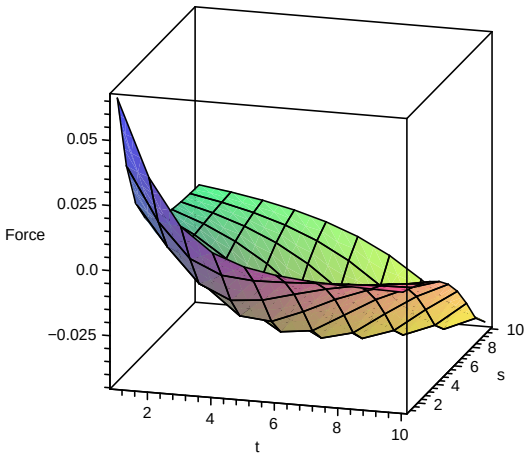


Figure 3.10: Plot of force for steps $\{[0,1],[1,0],[2,0],[0,2]\}$ at a fixed width as a function of boundary parameters

Chapter 4

Gated and Wicketed Ferrers diagrams

4.1 Introduction

Both the self-avoiding walk and the self-avoiding polygon are mathematical idealizations motivated by problems in the sciences [MS93]. They are useful as models of chain-like objects such as polymers and helps understand the conformational structure of proteins in living organisms.

The self-avoiding walk is one of those problems which is very simple to state, has generated enormous interest and tons of conjectures, and yet has resisted attempts at a solution for several decades. Very simply, it is a path on a lattice which never visits itself. This simplest non-Markovian condition has made the problem extremely resistant to attacks. Even though no closed-form solution is known for the number of such walks for finite number of steps, there are very precise conjectures on their asymptotic behavior[1]. A close cousin of the self-avoiding walk is the self-avoiding polygon, which is a closed self-avoiding walk. This seems to be just as intractable as the self-avoiding walk.

In attempts to generalize this problem, and motivated by the physical problem of cross-linked membranes and polymer sheets, people have modelled the three-dimensional version, self-avoiding polyhedra and self-avoiding vesicles (which enclose a volume of space) by considering random non-intersecting surfaces[LSF87, FGW91]. These vesicles also arise in the study of percolation clusters in statistical mechanical models [Car00].

[1]See `http://mathworld.wolfram.com/Self-AvoidingWalkConnectiveConstant.html` for more on this problem.

One now notices that the cross-section of a generic self-avoiding vesicle is in fact an outermost self-avoiding polygon with any number of self-avoiding polygons nested inside the outer polygon and hence, the study of self-avoiding polygons with self-avoiding holes [vRW89, GJWE00]. Any exact enumeration in this area tends to be useful because it helps in obtaining bounds on the asymptotic growth exponents of self-avoiding walks and polygons.

Staircase polygons are a subset of self-avoiding polygons that can be enumerated exactly. In fact, both the perimeter and area generating functions have a particularly nice form — they are algebraic, meaning the generating function satisfies a polynomial equation. Motivated by the recent interest in the enumeration of staircase polygons with a single staircase puncture (see A057410 in [Slo07]) and conjectures of a holonomic solution [GJ06, RJG07], we investigate the simpler problem of Ferrers diagrams with Ferrers punctures of different kinds - wickets and gates. Since Ferrers diagrams form a subset of staircase polygons, we simple-mindedly expect a "nice" solution here too. Fortunately for us, this naïve expectation turns out to be true and the generating function in both cases is not only algebraic, the degree of the polynomial equation that the generating function satisfies is two!

The technique we use is called the umbral transfer matrix method [Zei00], which is a generalization of the usual transfer matrix method in statistical mechanics, but where the entries are themselves operators called umbral operators. Our problem is simple enough to be solvable by a one-dimensional umbral transfer matrix.

For a given combinatorial building, the umbral operator is essentially the architectural plan for the structure. At any stage of the construction, it tells the builder exactly how to proceed from there on. The power of the method is twofold: it helps in generating terms in the sequence and secondly, once an ansatz is in place, it is very easy to prove or disprove the ansatz. One simply has to check the equality of algebraic expressions.

We first introduce the method by applying it to the simple case of standard Ferrers diagrams and then go on to apply it to the case of interest defining the various umbral operators that are needed and proving the main theorems.

4.2 Standard Ferrers Diagrams

Definition 1 *A* Ferrers diagram *is a collection of n rows of blocks, the ith row of which contains m_i blocks, the first row at the bottom and the last row on top. All the rows are left aligned and such that if $1 \leq i < j \leq n$ then $m_1 \leq m_i \leq m_j \leq m_n$.*

Since a picture is worth a thousand words, and because we want our convention to be clear, we augment the definition with an example.

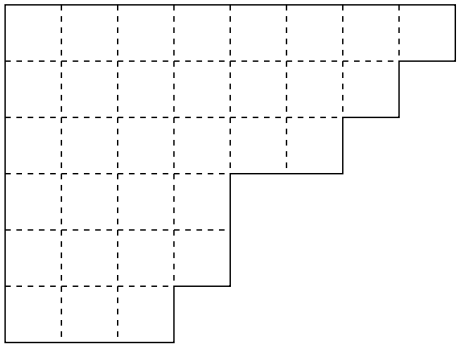


Figure 4.1: A Ferrers diagram with six rows, $m_1 = 3, m_2 = m_3 = 4, m_4 = 6, m_5 = 7, m_6 = 8$.

We first illustrate the methodology by applying it to the simple and almost trivial case of the half-perimeter generating function of usual Ferrers diagrams. Recall that this means we want to calculate the sum

$$\sum_{F \text{ a Ferrers diagram}} t^{P(F)/2}, \tag{4.2.1}$$

where $P(F)$ is the perimeter of the Ferrers diagram F. For example, when F is the Ferrers diagram in Figure 4.1, $P(F) = 28$. The basic idea is to assign a catalytic variable to keep track of the width of the topmost segment and consider all possible

Ferrers diagrams whose height is one more, which can be built by adding blocks to the original Ferrers diagram. This is represented by the so-called umbral evolution operator. In other words, if $h(F)$ represents the height of the Ferrers diagram, then we are representing the sum in (4.2.1) as

$$\sum_{F} = \sum_{F|h(F)=1} + \sum_{n=2}^{\infty} \sum_{F|h(F)=n} , \tag{4.2.2}$$

and using the evolution operator to represent every Ferrers diagram in the second sum in terms of the some element of the first sum.

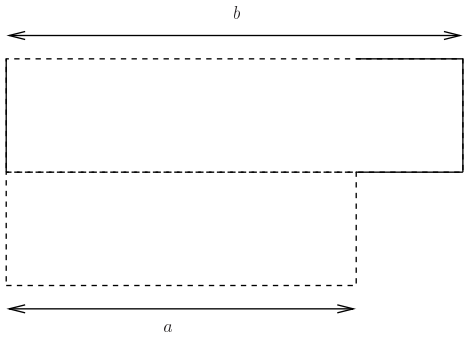


Figure 4.2: The umbral evolution operator for Ferrers diagrams. The darkened lines contribute to the perimeter.

Suppose, the top of the Ferrers diagram has height a as shown in Figure 4.2. Then, we construct all possible legal Ferrers diagrams which can be built from it. Let x be the catalytic variable counting the width of the topmost segment and t be the variable counting the half-perimeter. The umbral evolution acts by adding the contributions of the top width and the extra half-perimeter for all such legal Ferrers diagrams ,

$$x^a \mapsto \sum_{b=a}^{\infty} x^b t^{b-a+1} \tag{4.2.3}$$

$$= \frac{x^a t}{1-xt}. \tag{4.2.4}$$

This gives the evolution for the monomial x^a. Since we need the operator evolution,

we need the evolution for a formal power series $p(x)$:

$$U_0(p(x)) = \frac{tp(x)}{1-xt}. \tag{4.2.5}$$

Now we need initial conditions. Obviously, they correspond to Ferrers diagrams of height one. Thus, as a formal power series, the initial condition is

$$\begin{aligned} I(x) &= \sum_{a=1}^{\infty} x^a t^{a+1} && (4.2.6)\\ &= \frac{xt^2}{1-xt}. && (4.2.7) \end{aligned}$$

Then the generating function of Ferrers diagrams $F(x)$, where the variable x keeps track of the height of the topmost segment, satisfies the equation

$$\begin{aligned} F(x) &= I(x) + U_0(F(x)) && (4.2.8)\\ &= \frac{xt^2}{1-xt} + \frac{tF(x)}{1-xt}, && (4.2.9) \end{aligned}$$

which has the simple rational solution

$$F(x) = \frac{xt^2}{1-t-xt}, \tag{4.2.10}$$

Setting $x = 1$ gives the well-known formula

$$F(1) = \frac{t^2}{1-2t}, \tag{4.2.11}$$

the expansion of which gives the terms in A000079 [Slo07].

4.3 Gated and Wicketed Ferrers diagrams

Definition 2 *A* gated Ferrers diagram *is a Ferrers diagram from which another Ferrers diagram with the same conventions is removed from the top.*

Definition 3 *A* wicketed Ferrers diagram *is a Ferrers diagram from which another Ferrers diagram with the same conventions is removed from the interior.*

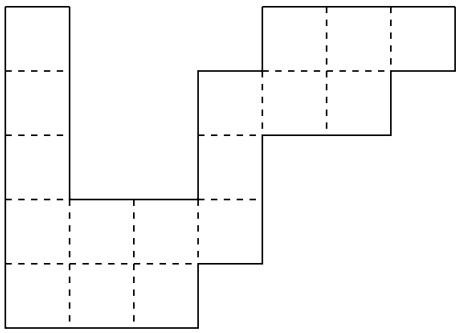


Figure 4.3: A gated Ferrers diagram with five rows containing a three-rowed Ferrers gate.

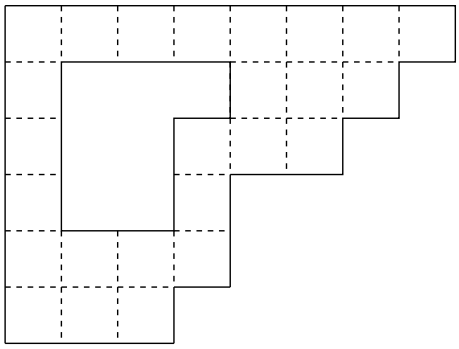


Figure 4.4: A wicketed Ferrers diagram with six rows containing a three-rowed Ferrers wicket .

The terminology is inspired by the figures. A gated Ferrers diagram like the one in Figure 4.3, when inverted, looks as if a door or a gate is cut out from the Ferrers diagram because the cut is at the bottom. A wicketed Ferrers diagram like the one in Figure 4.4 looks as if there is a wicket or window cut out because the cut is in the interior.

We use the umbral transfer matrix method again to count these objects. Unlike the simple case before, however, we need three and five different umbral operators for gated and wicketed diagrams respectively. Further complications arise from the fact that there are different number and types of catalytic variables in these different regions.

4.3.1 Construction of the gate/wicket

Suppose that the construction before the wicket is complete. In other words, we have a Ferrers diagram so far. We now begin creating the wicket in Figure 4.5. We take an object with one catalytic variable x counting the top width and add to it a different object with three catalytic variables x_1, x_2, x_3 marking the start of the wicket, the width of the wicket and the remaining width respectively also at the top.

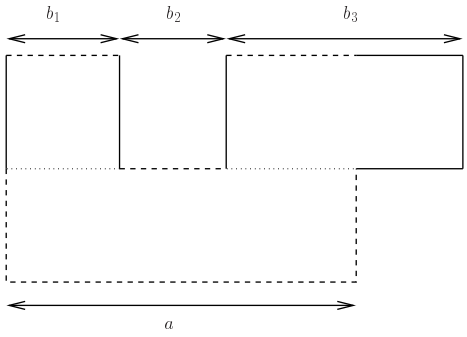


Figure 4.5: The umbral evolution operator for creating the gate or wicket. The darkened lines contribute to the perimeter.

Since we want the wicket to be strictly within the diagram, we can see that $b_1 < a-1$ and $b_1 + b_2 < a$. The extra contribution to the half-perimeter can be seen by looking at the darkened lines in Figure 4.5. Thus the evolution operator U_1 acts on the monomial by

$$\begin{aligned} x^a \mapsto \quad & \sum_{b_1=1}^{a-2} \sum_{b_2=1}^{a-1-b_1} \sum_{b_3=a-b_1-b_2}^{\infty} x_1^{b_1} x_2^{b_2} x_3^{b_3} t^{2+b_1+b_2+b_3-a} \\ & = \frac{x_1^{a-1} x_2^2 x_3 t^2}{(x_1-x_2)(1-x_3t)(x_2-x_3)} - \frac{x_1^{a-1} x_2 x_3^2 t^2}{(x_1-x_3)(1-x_3t)(x_2-x_3)} \\ & - \frac{x_1 x_2^a x_3 t^2}{(x_1-x_2)(1-x_3t)(x_2-x_3)} + \frac{x_1 x_2 x_3^a t^2}{(x_1-x_3)(1-x_3t)(x_2-x_3)} \end{aligned} \tag{4.3.1}$$

Therefore, U_1 acts on a formal power series $p(x)$ as follows:

$$\begin{aligned} U_1(p(x)) = &\frac{x_2 x_3 t^2 p(x_1)}{(x_1 - x_2)(x_1 - x_3)(1 - x_3 t)} \\ &+ \frac{x_1 t^2 \left[p(x_3) x_2 (x_1 - x_2) - p(x_2) x_3 (x_1 - x_3)\right]}{(x_1 - x_2)(x_1 - x_3)(x_2 - x_3)(1 - x_3 t)}. \end{aligned} \tag{4.3.2}$$

4.3.2 Extension of the gate/wicket

The operator which extends the gate or the wicket takes as input a formal power series with three catalytic variables and returns a formal power series of the same kind.

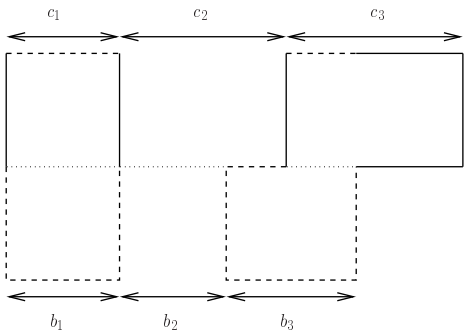


Figure 4.6: The umbral evolution operator for extending the gate or wicket. The darkened lines contribute to the perimeter.

The nature of the Ferrers diagram forces $c_1 = b_1$. The figure shows that $c_2 \geq b_2$ and $c_2 < b_2 + b_3$. Thus the umbral operator U_2 acts as

$$\begin{aligned} x_1^{c_1} x_2^{b_2} x_3^{b_3} \mapsto & x_1^{c_1} \sum_{c_2 = b_2}^{b_2 + b_3 - 1} \sum_{c_3 = b_2 + b_3 - c_2}^{\infty} x_2^{c_2} x_3^{c_3} t^{2 + c_2 + c_3 - b_2 - b_3} \\ &= \frac{t^2 x_1^{c_1} x_2^{b_2} x_3 (x_2^{b_3} - x_3^{b_3})}{(1 - x_3 t)(x_2 - x_3)} \end{aligned} \tag{4.3.3}$$

Thus, U_2 acts on formal power series according to

$$U_2(p(x_1, x_2, x_3)) = \frac{x_3 t^2}{(1 - x_3 t)(x_2 - x_3)} [p(x_1, x_2, x_2) - p(x_1, x_2, x_3)] \tag{4.3.4}$$

Since gated Ferrers diagrams are essentially unfinished wicketed Ferrers diagrams

we are in a position to describe the former. We can now express the generating function for these objects completely.

Theorem 4.3.1 *Let $\phi(t)$ be the half-perimeter generating function of gated Ferrers diagrams. Then ϕ satisfies the following quadratic equation:*

$$\left[t^2(1-2t)^4(1-3t+t^2)\right]\phi^2 + \left[t^4(1-3t+t^2)(1-2t)^2\right]\phi - t^{10} = 0 \tag{4.3.5}$$

Before going on to the proof, it might be helpful to say a few words on how one could arrive at this answer. Using the umbral transfer matrix method and sufficient computing time, one simply generates polynomials in x_1, x_2, x_3 which correspond to a fixed half-perimeter value. Once one has a sufficient number of terms, one checks for either a P-recursive ansatz, or in this case, an algebraic ansatz.

In this case, there are three symbolic variables and therefore both generating of the data as well as checking the ansatz takes very long and therefore, one plugs in various integer values of the x_i's and finds that for all of them, the equation satisfied by the conjectured generating function factorizes beautifully. Then a simple polynomial interpolation algorithm does the trick.

As far as we can tell, the sequence enumerated by ϕ have not been studied before and has been entered in the integer sequence database as A133106 [Slo07].

Proof of Theorem 4.3.1: Let $\phi_{123}(x_1, x_2, x_3, t)$ be the generating function for gated Ferrers diagrams where t counts the half-perimeter, x_1 counts the width before the start of the wicket, x_2 counts the width of the wicket and x_3 counts the width after the end of the wicket. We claim that $\phi_{123}(x_1, x_2, x_3, t)$ satisfies the equation below.

$$\begin{aligned}&\left[(x_2 - x_3 + x_3(x_3 - x_2)t + x_3t^2)(1 - x_1t - t)^2(1 - x_2t - 2t + t^2)\right.\\ &\left.\times(1 - x_3t - t)^2\right]\phi_{123}^2 + \left[t^4(x_1x_2x_3)(1 - x_2t - 2t + t^2)(1 - x_3t + t)\right.\\ &\left.\times(1 - x_3t - t)(1 - x_1t - t)\right]\phi_{123} - t^{10}(x_1x_2x_3)^2 = 0\end{aligned} \tag{4.3.6}$$

To see that, notice that $\phi_{123}(x_1, x_2, x_3, t)$ satisfies the umbral evolution equation

$$\phi_{123} = U_1\left(\frac{xt^2}{1-t-xt}\right) + U_2(\phi_{123}) \tag{4.3.7}$$

because U_1 takes a Ferrers diagram and begins the gate which can be extended using U_2.

One solves (4.3.6) for ϕ_{123} and notices that only the solution with the positive root has the correct Taylor expansion. Then one simply plugs it into (4.3.7) and finds that it is satisfied.

Now substitute $x_1 = x_2 = x_3 = 1$ into (4.3.6) to find that $\phi_{123}(1, 1, 1, t) = \phi(t)$ satisfies (4.3.5). □

4.3.3 Termination of the wicket

To describe wicketed Ferrers diagrams , we also need the umbral operator describing the end of the wicket. This is essentially the reverse process of adding the wicket. Now we go from three catalytic variables to one.

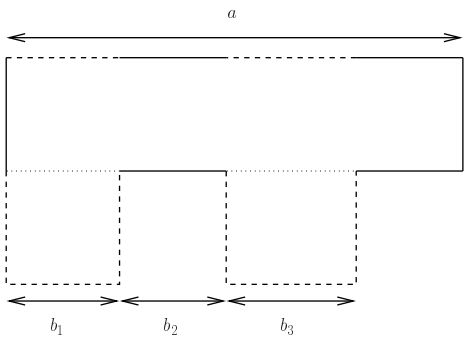


Figure 4.7: The umbral evolution operator for ending the wicket. The darkened lines contribute to the perimeter.

The action of U_3 is relatively straightforward.

$$\begin{aligned} x_1^{b_1}x_2^{b_2}x_3^{b_3} \mapsto & \sum_{a=b_1+b_2+b_3}^{\infty} x^a t^{1+a-b_1-b_3} \\ & = \frac{x^{b_1+b_2+b_3}t^{1+b_2}}{1-tx} \end{aligned} \tag{4.3.8}$$

Thus, U_3 acts on formal power series by

$$U_3(p(x_1,x_2,x_3)) = \frac{t}{1-tx}p(x,tx,x) \tag{4.3.9}$$

Since one needs to start with a gated Ferrers diagram to terminate the wicket, the initial condition for the evolution of Ferrers diagrams with one Ferrers wicket will be $U_3(\phi)$.

4.3.4 Termination of the diagram

Once the wicket has been sealed up, the umbral operator is again given by (4.2.5). This is because the evolution of the Ferrers diagram will continue as usual once the wicket has been established.

Theorem 4.3.2 *Let $\psi(t)$ be the half-perimeter generating function of wicketed Ferrers diagrams. Then ψ satisfies the following quadratic equation*

$$(2t-1)^8\psi^2 - t^6(2t-1)^4\psi + t^{14} = 0. \tag{4.3.10}$$

As far as we can tell, the sequence enumerated by ψ has also not been studied before and is present in the integer sequence database as A133107 [Slo07].

Proof of Theorem 4.3.2: Let $\psi_1(x,t)$ be the generating function of wicketed Ferrers diagrams where t counts the half-perimeter and x counts the width of the topmost segment. We claim that ψ_1 satisfies the equation below.

$$\begin{aligned} & \left[(xt+t-1)^6(xt^2-t^2+2t-1)(xt^2-t^2-xt-t+1)\right]\psi_1^2 \\ & - \left[t^6x^2(xt+t-1)^3(xt-t-1)(xt^2-t^2+2t-1)\right]\psi_1 - t^{14}x^5 = 0. \end{aligned} \tag{4.3.11}$$

To see this, notice that ψ_1 satisfies an umbral equation where the initial condition consists of Ferrers diagrams with a just-finished Ferrers wicket and the evolution is the usual Ferrers diagram evolution given by (4.2.5). That is,

$$\psi_1 = U_3(\phi_{123}) + U_0(\psi_1) \tag{4.3.12}$$

One solves the quadratic equation (4.3.11) for ψ_1 and takes the negative root, which is the only one that has the correct Taylor expansion. One plugs the solutions of ϕ_{123} and ψ_1 into (4.3.12) and uses the definition of the operators in (4.2.5),(4.3.9) to easily check that it is verified. Finally, one substitutes $x = 1$ into (4.3.11) to find that $\psi_1(1, t) = \psi(t)$ satisfies (4.3.10). □

4.4 Remarks

The generating function for wicketed Ferrers diagrams is very pretty. Solving for ψ in (4.3.10) yields

$$\psi(t) = t^6 \frac{1 - \sqrt{1 - 4t^2}}{2(1 - 2t)^4} \tag{4.4.1}$$

and after some factorizing gives

$$\psi(t) = \frac{1 - \sqrt{1 - 4t^2}}{2t^2} \left(\frac{t^2}{1 - 2t} \right)^4 . \tag{4.4.2}$$

which is a product of the half-perimeter generating function of staircase polygons (which gives the Catalan numbers) and *four* sets of Ferrers diagrams as shown in (4.2.11).

This nice factorization suggests an alternative proof using the technique of squeezing [Zei07a]. We give some details of the idea in Appendix C. It is not clear, however, that this proof will be considerably simpler than the umbral transfer matrix proof.

One might wonder whether all the machinery used here is superfluous and whether this generating function might be arrived at by simpler means. We do not believe it to be so. We present a couple of arguments in favour of this assertion.

First off, a direct bijection is not so trivial because the first two terms in the series expansion of ψ are 1 and 8 and in both of them the Catalan part is trivial (i.e. contributes 1). The series grows sufficiently fast so as to make experimentation difficult. Any natural bijection would be extremely interesting.

Secondly, the Gessel-Viennot determinant formula [GV85] is not applicable here because the beginning and end of the Ferrers wicket are not constrained to lie at the corners of the Ferrers diagram and in fact, are forbidden from being there.

References

[ABL88] E. D. Andjel, M. D. Bramson, and T. M. Liggett. Shocks in the asymmetric exclusion process. *Probab. Theory Related Fields*, 78(2):231–247, 1988.

[ALS08] Arvind Ayyer, Joel L. Lebowitz, and E. R. Speer. On the two species asymmetric exclusion process with semi-permeable boundaries. *preprint,* `arXiv: 0807.2423 [cond-mat]`, 2008.

[Ari06a] Chikashi Arita. Exact analysis of two-species totally asymmetric exclusion process with open boundary condition. *Journal of the Physical Society of Japan*, 75:065003, 2006.

[Ari06b] Chikashi Arita. Phase transitions in the two-species totally asymmetric exclusion process with open boundaries. *Journal of Statistical Mechanics: Theory and Experiment*, 2006(12):P12008, 2006.

[Ayy07] Arvind Ayyer. The half-perimeter generating function of gated and wicketed Ferrers diagrams. *J. Integer Seq.*, 10(10):Article 07.10.3, 11, 2007.

[AZ07a] Arvind Ayyer and Doron Zeilberger. The number of [old-time] basketball games with final score $n : n$ where the home team was never losing but also never ahead by more than w points. *Electron. J. Combin.*, 14(1):Research Paper 19, 8 pp. (electronic), 2007.

[AZ07b] Arvind Ayyer and Doron Zeilberger. Two dimensional directed lattice walks with boundaries. In *Tapas in Experimental Mathematics*, volume 457 of *Contemp. Math.*, pages 1–19. Amer. Math. Soc., Providence, RI, 2007.

[BDSG+07] L. Bertini, A. De Sole, D. Gabrielli, G. Jona-Lasinio, and C. Landim. Stochastic interacting particle systems out of equilibrium. *J. Stat. Mech. Theory Exp.*, pages P07014, 35 pp. (electronic), 2007.

[BDSG+08] L. Bertini, A. De Sole, D. Gabrielli, G. Jona-Lasinio, and C. Landim. Towards a nonequilibrium thermodynamics: a self-contained macroscopic description of driven diffusive systems. *preprint,* `arXiv: 0807.4457 [cond-mat]`, 2008.

[BE07] R A Blythe and M R Evans. Nonequilibrium steady states of matrix-product form: a solver's guide. *Journal of Physics A: Mathematical and Theoretical*, 40(46):R333–R441, 2007.

[Bir30] George D. Birkhoff. Formal theory of irregular linear difference equations. *Acta Math.*, 54(1):205–246, 1930.

[BM07] Mireille Bousquet-Mélou. Discrete excursions. *Sém. Lothar. Combin.*, 57:Art. B57d, 23, 2006/07.

[BORW05] R. Brak, A. L. Owczarek, A. Rechnitzer, and S. G. Whittington. A directed walk model of a long chain polymer in a slit with attractive walls. *J. Phys. A*, 38(20):4309–4325, 2005.

[BT33] George D. Birkhoff and W. J. Trjitzinsky. Analytic theory of singular difference equations. *Acta Math.*, 60(1):1–89, 1933.

[Car00] John Cardy. Linking numbers for self-avoiding loops and percolation: Application to the spin quantum hall transition. *Phys. Rev. Lett.*, 84(16):3507–3510, Apr 2000.

[Com64] L. Comtet. Calcul pratique des coefficients de Taylor d'une fonction algébrique. *Enseignement Math. (2)*, 10:267–270, 1964.

[DDM92] B. Derrida, E. Domany, and D. Mukamel. An exact solution of a one-dimensional asymmetric exclusion model with open boundaries. *J. Statist. Phys.*, 69(3-4):667–687, 1992.

[DE99] B. Derrida and M. R. Evans. Bethe ansatz solution for a defect particle in the asymmetric exclusion process. *J. Phys. A*, 32(26):4833–4850, 1999.

[DEHP93] B. Derrida, M. R. Evans, V. Hakim, and V. Pasquier. Exact solution of a 1D asymmetric exclusion model using a matrix formulation. *J. Phys. A*, 26(7):1493–1517, 1993.

[Der96] B. Derrida. Systems out of equilibrium: some exactly soluble models. In *STATPHYS 19 (Xiamen, 1995)*, pages 243–253. World Sci. Publ., River Edge, NJ, 1996.

[DHV92] B. Derrida, V. Hakim, and J. Vannimenus. Effect of disorder on two-dimensional wetting. *J. Statist. Phys.*, 66(5-6):1189–1213, 1992.

[DJLS93] B. Derrida, S. A. Janowsky, J. L. Lebowitz, and E. R. Speer. Exact solution of the totally asymmetric simple exclusion process: shock profiles. *J. Statist. Phys.*, 73(5-6):813–842, 1993.

[DLS03] B. Derrida, J. L. Lebowitz, and E. R. Speer. Exact large deviation functional of a stationary open driven diffusive system: the asymmetric exclusion process. *J. Statist. Phys.*, 110(3-6):775–810, 2003. Special issue in honor of Michael E. Fisher's 70th birthday (Piscataway, NJ, 2001).

[Duc00] Philippe Duchon. On the enumeration and generation of generalized Dyck words. *Discrete Math.*, 225(1-3):121–135, 2000. Formal power series and algebraic combinatorics (Toronto, ON, 1998).

[EH05] M R Evans and T Hanney. Nonequilibrium statistical mechanics of the zero-range process and related models. *Journal of Physics A: Mathematical and General*, 38(19):R195–R240, 2005.

[Fel71] William Feller. *An introduction to probability theory and its applications. Vol. II.* Second edition. John Wiley & Sons Inc., New York, 1971.

[FFK94] P. A. Ferrari, L. R. G. Fontes, and Y. Kohayakawa. Invariant measures for a two-species asymmetric process. *J. Statist. Phys.*, 76(5-6):1153–1177, 1994.

[FGW91] M E Fisher, A J Guttmann, and S G Whittington. Two-dimensional lattice vesicles and polygons. *Journal of Physics A: Mathematical and General*, 24(13):3095–3106, 1991.

[FM07] Pablo A. Ferrari and James B. Martin. Stationary distributions of multi-type totally asymmetric exclusion processes. *Ann. Probab.*, 35(3):807–832, 2007.

[GJ06] Anthony J. Guttmann and Iwan Jensen. The perimeter generating function of punctured staircase polygons. *J. Phys. A*, 39(15):3871–3882, 2006.

[GJWE00] Anthony J Guttmann, Iwan Jensen, Ling Heng Wong, and Ian G Enting. Punctured polygons and polyominoes on the square lattice. *Journal of Physics A: Mathematical and General*, 33(9):1735–1764, 2000.

[GV85] Ira Gessel and Gérard Viennot. Binomial determinants, paths, and hook length formulae. *Adv. in Math.*, 58(3):300–321, 1985.

[Hil56] Terrell L. Hill. *Statistical mechanics: Principles and selected applications.* The McGraw-Hill Series in Advanced Chemistry. McGraw-Hill Book Co., Inc., New York, 1956.

[KJS03] K Krebs, F H Jafarpour, and G M Schütz. Microscopic structure of travelling wave solutions in a class of stochastic interacting particle systems. *New Journal of Physics*, 5:145, 2003.

[KLS84] Sheldon Katz, Joel L. Lebowitz, and Herbert Spohn. Nonequilibrium steady states of stochastic lattice gas models of fast ionic conductors. *J. Statist. Phys.*, 34(3-4):497–537, 1984.

[Kru91] Joachim Krug. Boundary-induced phase transitions in driven diffusive systems. *Phys. Rev. Lett.*, 67(14):1882–1885, Sep 1991.

[KUH63] M. Kac, G. E. Uhlenbeck, and P. C. Hemmer. On the van der Waals theory of the vapor-liquid equilibrium. I. Discussion of a one-dimensional model. *J. Mathematical Phys.*, 4:216–228, 1963.

[Lig99] Thomas M. Liggett. *Stochastic interacting systems: contact, voter and exclusion processes*, volume 324 of *Grundlehren der Mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences]*. Springer-Verlag, Berlin, 1999.

[LSF87] Stanislas Leibler, Rajiv R. P. Singh, and Michael E. Fisher. Thermodynamic behavior of two-dimensional vesicles. *Phys. Rev. Lett.*, 59(18):1989–1992, Nov 1987.

[Mal96] K Mallick. Shocks in the asymmetry exclusion model with an impurity. *Journal of Physics A: Mathematical and General*, 29(17):5375–5386, 1996.

[MS93] Neal Madras and Gordon Slade. *The self-avoiding walk*. Probability and its Applications. Birkhäuser Boston Inc., Boston, MA, 1993.

[Per87] J. K. Percus. Exactly solvable models of classical many-body systems. In *Simple models of equilibrium and nonequilibrium phenomena*, volume 13 of *Stud. Statist. Mech.*, pages 1–158. North-Holland, Amsterdam, 1987.

[PS66a] Douglas Poland and Harold A. Scheraga. Occurrence of a phase transition in nucleic acid models. *The Journal of Chemical Physics*, 45(5):1464–1469, 1966.

[PS66b] Douglas Poland and Harold A. Scheraga. Phase transitions in one dimension and the helix—coil transition in polyamino acids. *The Journal of Chemical Physics*, 45(5):1456–1463, 1966.

[PWZ96] Marko Petkovšek, Herbert S. Wilf, and Doron Zeilberger. $A = B$. A K Peters Ltd., Wellesley, MA, 1996. With a foreword by Donald E. Knuth, With a separately available computer disk.

[RJG07] Christoph Richard, Iwan Jensen, and Anthony J. Guttmann. Area distribution and scaling function for punctured polygons, 2007.

[RPS03] A. Rákos, M. Paessens, and G. M. Schütz. Hysteresis in one-dimensional reaction-diffusion systems. *Phys. Rev. Lett.*, 91(23):238302, Dec 2003.

[SD93] G. Schütz and E. Domany. Phase transitions in an exactly soluble one-dimensional asymmetric exclusion model. *J. Statist. Phys.*, 72(1-2):277–296, 1993.

[Slo07] N. J. A. Sloane. The on-line encyclopedia of integer sequences, 2007. Published electronically at `http://www.research.att.com/~njas/sequences/`.

[Spe94] E. R. Speer. The two species totally asymmetric exclusion process. In *On Three Levels: The Micro-, Meso-, and Macroscopic Approaches in Physics*, volume 324 of *NATO ASI Series B: Physics*, pages 91–112. Plenum, New York, 1994.

[Spi70] Frank Spitzer. Interaction of Markov processes. *Advances in Math.*, 5:246–290 (1970), 1970.

[SZ94] Bruno Salvy and Paul Zimmerman. Gfun: a maple package for the manipulation of generating and holonomic functions in one variable. *ACM Trans. Math. Softw.*, 20(2):163–177, 1994.

[Tem56] H. N. V. Temperley. Combinatorial problems suggested by the statistical mechanics of domains and of rubber-like molecules. *Phys. Rev. (2)*, 103:1–16, 1956.

[vR00] E. J. Janse van Rensburg. *The statistical mechanics of interacting walks, polygons, animals and vesicles*, volume 18 of *Oxford Lecture Series in Mathematics and its Applications*. Oxford University Press, Oxford, 2000.

[vRW89] E J Janse van Rensburg and S G Whittington. Self-avoiding surfaces. *Journal of Physics A: Mathematical and General*, 22(22):4939–4958, 1989.

[WZ85] Jet Wimp and Doron Zeilberger. Resurrecting the asymptotics of linear recurrences. *J. Math. Anal. Appl.*, 111(1):162–176, 1985.

[WZ90a] Herbert S. Wilf and Doron Zeilberger. Rational functions certify combinatorial identities. *J. Amer. Math. Soc.*, 3(1):147–158, 1990.

[WZ90b] Herbert S. Wilf and Doron Zeilberger. Towards computerized proofs of identities. *Bull. Amer. Math. Soc. (N.S.)*, 23(1):77–83, 1990.

[Zei90] Doron Zeilberger. A holonomic systems approach to special functions identities. *J. Comput. Appl. Math.*, 32(3):321–368, 1990.

[Zei91] Doron Zeilberger. The method of creative telescoping. *J. Symbolic Comput.*, 11(3):195–204, 1991.

[Zei00] Doron Zeilberger. The umbral transfer-matrix method. I. Foundations. *J. Combin. Theory Ser. A*, 91(1-2):451–463, 2000. In memory of Gian-Carlo Rota.

[Zei07a] Doron Zeilberger. Alternate proof of theorem 4.3.2. private communication, 2007.

[Zei07b] Doron Zeilberger. The holonomic ansatz. I. Foundations and applications to lattice path counting. *Ann. Comb.*, 11(2):227–239, 2007.

Appendix A

Notes for Chapter 2

A.1 A particular representation

A representation of the algebra (2.3.2)–(2.3.3) which satisfies (2.3.6) may be obtained from [DEHP93] and [DJLS93]:

$$X_1 = \begin{pmatrix} 1 & 1 & 0 & 0 & . & . \\ 0 & 1 & 1 & 0 & & \\ 0 & 0 & 1 & 1 & & \\ 0 & 0 & 0 & 1 & . & \\ . & & & & . & . \\ . & & & & & . \end{pmatrix}, \qquad X_0 = \begin{pmatrix} 1 & 0 & 0 & 0 & . & . \\ 1 & 1 & 0 & 0 & & \\ 0 & 1 & 1 & 0 & & \\ 0 & 0 & 1 & 1 & & \\ . & & & & . & . \\ . & & & & & . \end{pmatrix}. \quad \text{(A.1.1)}$$

$$X_2 = X_1X_0 - X_0X_1 = [X_1, X_0] = \begin{pmatrix} 1 & 0 & 0 & 0 & . & . \\ 0 & 0 & 0 & 0 & & \\ 0 & 0 & 0 & 0 & & \\ 0 & 0 & 0 & 0 & & \\ . & & & & . & \\ . & & & & & . \end{pmatrix}, \qquad \text{(A.1.2)}$$

$$\langle W_\alpha| = \left(1, \left(\frac{1-\alpha}{\alpha}\right), \left(\frac{1-\alpha}{\alpha}\right)^2, \dots\right), \qquad |V_\beta\rangle = \begin{pmatrix} 1 \\ \frac{1-\beta}{\beta} \\ \left(\frac{1-\beta}{\beta}\right)^2 \\ \vdots \end{pmatrix}. \quad \text{(A.1.3)}$$

The exponential growth of the components of $\langle W_\alpha|$ and $|V_\beta\rangle$ for certain values of α and β in fact causes no concern here: because we always have $n > 0$, the matrix product needed to calculate the probability of any configuration τ (see (2.3.4)) will contain at least one factor X_2, and using (2.3.6) one can see that this implies that the corresponding matrix element is finite.

A.2 Asymptotics of the partition function

We summarize here the asymptotics of the partition function which are needed in Section 2.6. For the case with no second class particles [DEHP93] we need $Z^{\alpha,\beta}$ only when $\alpha = 1$ and/or $\beta = 1$:

$$Z^{\alpha,1}(j,0) = Z^{1,\alpha}(j,0) \sim \begin{cases} \dfrac{1-2\alpha}{(1-\alpha)^2}\left(\dfrac{1}{\alpha(1-\alpha)}\right)^j, & \text{if } \alpha < 1/2, \\ \dfrac{2}{\sqrt{\pi}}\dfrac{4^j}{j^{1/2}}, & \text{if } \alpha = 1/2, \\ \dfrac{\alpha^2}{\sqrt{\pi}(2\alpha-1)^2}\dfrac{4^{j+1}}{j^{3/2}}, & \text{if } \alpha > 1/2. \end{cases} \tag{A.2.1}$$

The generating function is [BE07]

$$\sum_{L=1}^{\infty} \lambda^L Z_{L,0}^{\alpha,\beta} = \left(\frac{2\alpha}{2\alpha - 1 + \sqrt{1-4\lambda}}\right)\left(\frac{2\beta}{2\beta - 1 + \sqrt{1-4\lambda}}\right). \tag{A.2.2}$$

For the model with second class particles [Ari06b]:

- In region I, $(\alpha_c < \alpha, \beta)$

$$Z^{\alpha,\beta}(L,n) = \frac{n\alpha\beta\sqrt{L^2-n^2}}{\sqrt{\pi L}((2\alpha-1)L+n)((2\beta-1)L+n)}\left(\frac{4L^2}{L^2-n^2}\right)^{L+1}\left(\frac{L-n}{L+n}\right)^n \tag{A.2.3}$$

- In region II, $(\alpha < \alpha_c, \alpha < \beta)$

$$Z^{\alpha,\beta}(L,n) = \frac{\beta(1-2\alpha)}{\alpha(\beta-\alpha)}\left(\frac{1}{\alpha(1-\alpha)}\right)^{L+1}\left(\frac{\alpha}{1-\alpha}\right)^n; \tag{A.2.4}$$

- On the boundary of regions I and II, $(\alpha_c = \alpha < \beta)$

$$Z^{\alpha,\beta}(L,n) = \frac{\beta n(L-n)}{2L((2\beta-1)L+n)}\left(\frac{4L^2}{L^2-n^2}\right)^{L+1}\left(\frac{L-n}{L+n}\right)^n \tag{A.2.5}$$

- On the boundary of regions II and III, ($\alpha = \beta < \alpha_c$)

$$Z^{\alpha,\beta}(L,n) = \frac{(1-2\alpha)((1-2\alpha)L-n)}{(1-\alpha)^2}\left(\frac{1}{\alpha(1-\alpha)}\right)^L \frac{\alpha 1}{(1-\alpha)} \tag{A.2.6}$$

- At the triple point, ($\alpha_c = \alpha = \beta$)

$$Z^{\alpha,\beta}(L,n) = \frac{n(L-n)}{2L(L+n)}\sqrt{\frac{L^2-n^2}{L\pi}}\left(\frac{4L^2}{L^2-n^2}\right)^{L+1}\left(\frac{L-n}{L+n}\right)^n. \tag{A.2.7}$$

Asymptotics in region III and on the I/III boundary are obtained from those of region II and the I/II boundary by exchange of α and β.

A.3 Finite volume corrections to density profiles

We consider here again the problem of finding asymptotic values of the density profiles, beginning with a discussion of the method of [Ari06b]. The partition function can be expressed as

$$Z^{\alpha,\beta}(L,n) = \frac{\alpha\beta}{\alpha-\beta}\left[R(L,n,\beta) - R(L,n,\alpha)\right], \tag{A.3.1}$$

where

$$R(L,n,\alpha) = \sum_{k=0}^{L-n} C_{L-n-k}^{L+n-1}\frac{1}{\alpha^{k+1}}, \tag{A.3.2}$$

$$\tag{A.3.3}$$

with C_n^m the Catalan triangle numbers (2.4.8). An asymptotic analysis of (A.3.2) then leads, through (A.3.1) and the formulas (2.4.13)–(2.4.14) for the densities, to the density asymptotics. In [Ari06b] the asymptotic density at position x was calculated as

$$\lim_{L\to\infty} \langle \eta_a(i_L)\rangle_{\mu^{\alpha,\beta}_{L,\lfloor \gamma L\rfloor}}, \tag{A.3.4}$$

with $i_L = \lfloor xL \rfloor$. As observed in Section 2.6, however, if x is the location of the fixed shock in regions II or III, and one considers limits as in (A.3.4) with $i_L = \lfloor xL\rfloor + c\sqrt{L}$, then the limiting density value depends on c. This c dependence may be calculated by

the methods of Section 2.6 (see for example (2.6.22)); here we sketch briefly an alternate and more direct method which extends the work of [Ari06b].

The key step is the computation of the asymptotics of $R(L, n, \alpha)$; it is convenient to introduce $\alpha_c = (L - n)/(2L)$ (see (2.2.2)). We must determine which terms in (A.3.2) dominate the sum. If we let $L \to \infty$ at fixed n and α there are three regimes: (i) $\alpha > \alpha_c$, for which the maximum of the summand is attained when k is of order L and the sum can be approximated by a Gaussian integral; (ii) $\alpha < \alpha_c$, in which the maximum is attained when k is of order $-L$ and the sum can be approximated by a geometric series; and (iii) $\alpha = \alpha_c$, for which the maximum occurs when k is of order 1 and the sum can be approximated by half of a Gaussian integral. However, there are intermediate regimes in which the sum is dominated by terms in which k is of order $\pm\sqrt{L}$, and it is these which generate the finite volume density corrections that we seek.

One needs an asymptotic estimate of $R(L, n, \alpha)$ which holds for all large L and n. Such an estimate is $R(L, n, \alpha) \sim \tilde{R}(L, n, \alpha)$, where

$$\tilde{R}(L,n,\alpha) = \begin{cases} (1-2\alpha)\left(\dfrac{1}{\alpha(1-\alpha)}\right)^{L+1}\left(\dfrac{\alpha}{1-\alpha}\right)^{n}\Phi\left(\dfrac{L(1-2\alpha)-n}{\sqrt{\alpha(L+n)}}\right), \\ \qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad \alpha \le \alpha_c, \\ \dfrac{2nL}{(n-L(1-2\alpha))\sqrt{\pi L(L^2-n^2)}}\left(\dfrac{4L^2}{L^2-n^2}\right)^{L} \\ \qquad \times\left(\dfrac{L-n}{L+n}\right)^{n}\Psi\left(\dfrac{L(1-2\alpha)-n}{\sqrt{\alpha(L+n)}}\right), \qquad \alpha > \alpha_c; \end{cases} \tag{A.3.5}$$

here Φ is as 1 in (2.6.18) and $\Psi(t) = \sqrt{2\pi}e^{t^2/2}|t|\Phi(t)$. The asymptotic estimate $R \sim \tilde{R}$ holds in the sense that for α and the ratio n/L uniformly bounded away from 0 and 1 the quantity $|R/\tilde{R} - 1|$ is small when L is large—more precisely, for any $\epsilon > 0$ there is a constant C_ϵ such that $|R/\tilde{R} - 1| \le C_\epsilon L^{-1/2-\epsilon}$. We remark that the two forms in (A.3.5) in fact agree for $\alpha_c < \alpha < \alpha_c + O(1/\sqrt{L})$.

From (A.3.1) and (A.3.5) one obtains similarly improved asymptotics for the partition function $Z^{\alpha,\beta}(L,n)$, and the full density asymptotics then follows from the exact formulas of [Ari06b] or Theorem 2.4.3.

Appendix B

A Quick Maple Tutorial for Chapter 3

Here we describe the basic procedure for using the program to determine the necessary information for your favorite polymer.

First off, download the package *POLYMER* from the webpage of Doron Zeilberger or by downloading the source from
`http://arXiv.org/cond-mat/0701674` . Start Maple and at the prompt, type

```
> read 'POLYMER' :
```

If you start Maple in a different directory, you have to specify the path where you saved the package. For example, if you saved it in `C:\Packages` or in `/tmp/Packages` (depending on the OS), type

```
> read 'C : \ \ Packages \ \POLYMER' :
> read '/tmp/Packages/POLYMER' :
```

To see the list of programs, type

```
> Help();
```

We now describe the main tools of the package. The basic syntax is as follows. Any point (x_1, y_1) is represented as $[x_1, y_1]$. The set of steps is represented within curly braces. For example, the steps shown in Figure 3.3 are depicted by $\{[0, 1], [1, 0]\}$.

B.1 Walks

The most basic program in the package is the one that computes the number of walks from any point (x_1, y_1) to any other point (x_2, y_2) using any set of steps and any width w.

B.1.1 Simple Walks

For example, to see the number of ways of getting from the origin to the point $(2, 2)$ using the steps above with the constraint given by $0 \leq x - y \leq 3$ is

$$> \quad \text{polymerBE}(\{[0,1],[1,0]\},[0,0],[2,2],3);$$

$$2$$

while the same walk with the stronger constraint $0 \leq x - y \leq 1$ is

$$> \quad \text{polymerBE}(\{[0,1],[1,0]\},[0,0],[2,2],1);$$

$$1$$

To see why that is true, look at Figure 3.3.

B.1.2 Walks with Boundary Interactions

We repeat the calculation for exactly the same situation in the cases where the width, $w = 1, 2, 3$.

$$> \quad \text{WEpolymerBE}(\{[0,1],[1,0]\},[0,0],[2,2],1,\text{t},\text{s});$$

$$t^2 s^2$$

This is because the only walk touches both walls twice. For $w = 2$,

$$> \quad \text{WEpolymerBE}(\{[0,1],[1,0]\},[0,0],[2,2],2,\text{t},\text{s});$$

$$t^2 + ts$$

which is because there are two walks now and one walk does not touch the wall on the far right at all. And for $w = 3$,

$$> \quad \text{WEpolymerBE}(\{[0,1],[1,0]\},[0,0],[2,2],3,\text{t},\text{s});$$

$$t^2 + t$$

which is because neither of the two walks touches the wall on the far right.

B.2 Generating Functions

The package can be used to compute this generating function for any finite width as well as the special case of the infinite width.

As a simple example, consider the same steps as before. Then, for $w = 1$, there is only one way of getting to the point (n, n), which is by the zigzag route extending the walk on the left of Figure 3.3. Therefore, the generating function is given by

$$\begin{aligned} \phi_1(z) &= 1 + z + z^2 + z^3 + \cdots \\ &= \frac{1}{1-z} \end{aligned} \tag{B.2.1}$$

To verify this, type

$$> \quad \text{rigorgf}(\{[0,1],[1,0]\},1,\text{z});$$

$$\frac{1}{1-z}$$

For a more nontrivial example, see what happens for $w = 3$.

$$> \quad \text{rigorgf}(\{[0,1],[1,0]\},3,\text{z});$$

$$\frac{1-2z}{1-3z+z^2}$$

To get the number of walks up to (n, n), one simply needs to look at the nth Taylor coefficient.

We can also calculate the generating functions of walks with variables t, s. As an example, we take the same steps as before with $w = 1$

$$> \quad \text{rigorgfWE}(\{[0,1],[1,0]\},3,\text{z},\text{t},\text{s});$$

$$\frac{1-z-sz}{1-z-sz-tz+stz^2}$$

For the case of infinite width, one can again calculate the generating function. The program returns the polynomial equation that it satisfies.

$$> \quad \text{RGF2D}(\{[0,1],[1,0]\},\text{z},\text{F});$$

$$\{1-F+zF^2\}$$

This means that $F(z)$ satisfies the equation $1 - F(z) + zF^2(z) = 0$. Since this is a quadratic equation in F, it can be solved easily.

$$F(z) = \frac{1 \pm \sqrt{1-4z}}{2z} \tag{B.2.2}$$

Since we want a formal power series and the taylor coefficients to be non-negative, we take the negative root. Taking the Taylor expansion gives

$$> \quad \text{taylor}((1-\text{sqrt}(1-4\text{z}))/(2\text{z}),\text{z}=0,10);$$

$$1+z+2z^2+5z^3+14z^4+42z^5+132z^6+429z^7+1430z^8+O(z^9)$$

These coefficients are precisely the *Catalan numbers* (A000108 of [Slo07]).

One can also calculate the weight enumerator for the same set of walks with infinite width, where t is the parameter whose coefficient counts the number of times the walk touches the diagonal.

$$> \quad \text{RGF2DWE}(\{[0,1],[1,0]\},\text{z},\text{F},\text{t});$$

$$\{1+(t-2)F+(t^2z+1-t)F^2\}$$

and plugging in $t = 1$ gives the unweighted generating function, as expected.

B.3 Empirical Guessing

That is, let $P(w, W)$ be an operator where W acts by shifts: $WQ_w(z) = Q_{w+1}(z)$. The degree of W in P is called the *order of the recurrence* and the degree of w in P is called the *degree of the recurrence*. Note that P implicitly depends on z.

Suppose we have a walk with steps $[0, 1], [1, 0]$. Let us try to find a recurrence of order 2 and degree 0 as [BORW] suggests.

$$> \quad \text{stepsrec}(\{[1,0],[0,1]\}, 0, 2, \text{z}, \text{w}, \text{W});$$

$$z - W + W^2$$

Similarly, we can find recurrences for the weight enumerators. The previous steps satisfy exactly the same recurrence for their weight enumerators! Consider the walk with steps $[0, 1], [1, 1], [1, 0]$ and order 4 and degree 0, we find

$$> \quad \text{stepsrecWE}(\{[1,0],[1,1],[0,1]\}, 0, 2, \text{z}, \text{w}, \text{W}, \text{t}, \text{s});$$

$$W^2 + (z-1)W + z$$

We remind the reader that these are essentially empirical results. One way to prove these is write down nonlinear recurrence relations for the generating functions $\phi_w(z)$ and prove them on a case-by-case basis.

B.4 Free Energy

The package can be used to calculate the free energy for any specific width, plot the free energy (using the weight enumerating generating function) at a specific width for ranges of t, s as well as plot the ordinary generating function in a range of widths.

Suppose we want to calculate the free energy for a specific width. As an example,

consider the same steps and $w = 3$.

$$> \quad \text{FE}(\{[0,1],[1,0]\},3);$$

$$\log\left(\frac{3}{2}+\frac{\sqrt{5}}{2}\right)$$

One could also, for example, plot the free energies for the same walk from widths of 1 to 10.

$$> \quad \text{plotFE}(\{[0,1],[1,0]\},1,10);$$

The asymptotic value is 1.386294361

The output is Figure 3.4.

For the weight enumerators, one can plot free energies for a fixed width and range of t and s parameters. Unfortunately, we cannot get asymptotic values here. For instance, with the same steps as before, we can look at the case when $w = 3$ and the range $t = 1, ..., 10, s = 1, ..., 10$. The

$$> \quad \text{plotFEWE}(\{[0,1],[1,0]\},10,1,12,1,12);$$

The output is Figure 3.5.

And lastly, one can plot the free energy as a function of the variable t for the infinite width case.

$$> \quad \text{plotinfFE}(\{[0,1],[1,0]\},1,10);$$

The output is Figure 3.6.

B.5 Force on Walls

Using essentially the same algorithm as for the free energy, one can plot the force to get an idea of the adsorption/desorption phase diagram.

For example, for the walk with steps $(1,0),(0,1)$ and $w=10$, we can plot the force in the range $t,s=1,\cdots,10$.

$$> \quad \text{ForceWE}(\{[0,1],[1,0]\},10,1,10,1,10);$$

The output is Figure 3.7.

Appendix C

Sketch of an Alternate Proof of Theorem 4.3.2

The idea is to consider the generating function of a pair of nonintersecting paths satisfying some constaints imposed by the structure of the wicketed Ferrers Diagrams. These are enumerated by the sum of their lengths, which is essentially the half-perimeter.

Note that one can add horizontal and vertical segments, both just after the first segment as well as just before the last segment to get a new legal pair of paths. This can be done both for the inner and the outer path. Doing it for the outer path properly gives a factor of $(t/(1-2t))^2$.

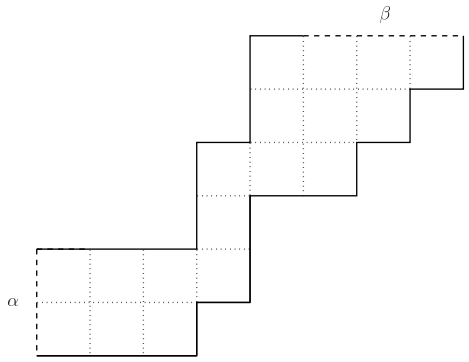


Figure C.1: A nibbled staircase polygon contributing $\alpha^2\beta^3$. Only the darkened lines contribute to the perimeter.

In an ideal world the same could be done for the inner path to give the same factor again. What would then be left is precisely the usual staircase polygons, which give the Catalan generating function. The problem is that various constraints come into play

for the inner path because the distance between both of them have to be atleast one both horizontally and vertically.

At this point one has reduced the problem to that of calculating the half-perimeter generating function of *Nibbled Staircase Polygons* with nibbling parameters α and β like that in Figure C.1, where the walk on top is the inner path and the one on the bottom is the outer path. The idea is that some part of the first vertical segment and the last horizontal segment of the higher path does not contribute to the half-perimeter. The generating function of these nibbled staircase polygons can be calculated again using the umbral transfer matrix method, but applied to simpler objects. Alternatively, it may be possible to adapt the elegant proof (originally due to Delest) of staircase polygons to count these nibbled objects.

Vita

Arvind Ayyer

2008 Ph. D. in Physics, Rutgers University

1998-2003 M. Sc. in Physics from Indian Institute of Technology, Kanpur, India.

2003-2006 Teaching assistant, Department of Physics, Rutgers University

2006-2008 Graduate assistant, Department of Physics, Rutgers University

Publications

1. Alok Sharan, Ramesh C. Sharma, S. N. Sandhya, Arvind Ayyer and K. K. Sharma, Modeling absorption in saturable absorbers, *Optics Communications*, **199** (2001), no. 1-4, 267–275.

2. Mahendra K. Verma, Arvind Ayyer, Amar V. Chandra, Olivier Debliquy and Shishir Kumar) Local Shell-to-Shell Energy Transfer via Nonlocal Interactions in Fluid Turbulence, *Pramana*, **65** (2005), 297–310.

3. Mahendra K. Verma, Arvind Ayyer and Amar V. Chandra, Energy Transfer and Locality in Magnetohydrodynamic Turbulence, *Phys. Plasmas*, **12** (2005), 082307 (7pp).

4. Arvind Ayyer and Doron Zeilberger, The Number of [Old-Time] Basketball games with Final Score n:n where the Home Team was never losing but also never ahead by more than w Points, *Electronic J. of Combinatorics* **14** (2007), no. 1, R19 (8pp).

5. Arvind Ayyer and Doron Zeilberger, Two Dimensional Directed Lattice Walks with Boundaries, *Tapas in Experimental Mathematics*, Volume 457 of Contemporary Mathematics, edited by Tewodros Amdeberhan and Victor Moll (2007), 1–19.

6. Arvind Ayyer and Mikko Stenlund, Exponential Decay of Correlations for Randomly Chosen Hyperbolic Toral Automorphisms, *Chaos* **17** (2007), 043116 (7pp).

7. Arvind Ayyer, The Half-Perimeter Generating Function of Gated and Wicketed Ferrers diagrams, *Journal of Integer Sequences* **10** (2007), no. 10, 07.10.3 (11pp).

8. Arvind Ayyer, Carlangelo Liverani and Mikko Stenlund, Quenched CLT for Random Toral Automorphisms, *submitted*.

9. Arvind Ayyer, Joel L. Lebowitz and Eugene R. Speer, On the Asymmetric Exclusion Process with Semi-Permeable Boundaries, *submitted*.