Combinatorics of Lattice Paths and Polygons

Nicholas Ross Beaton

Submitted in total fulfilment of the requirements of the degree of Doctor of
Philosophy

October 2012

Department of Mathematics and Statistics
The University of Melbourne

Abstract

We consider the enumeration of self-avoiding walks and polygons on regular lattices. Such objects are connected with many other problems in combinatorics, as well as in fields as diverse as physics and chemistry. We examine the general models of walks and polygons and methods we can use to study them; subclasses whose properties enable a rather deeper analysis; and extensions of these models which allow us to model physical phenomena like polymer collapse and adsorption.

While the general models of self-avoiding walks and polygons are certainly not considered to be 'solved', recently a great deal of progress has been made in developing new methods for studying these objects and proving rigorous results about their enumerative properties, particularly on the honeycomb lattice. We consider these recent results and show that in some cases they can be extended or generalised so as to enable further proofs, conjectures and estimates. In particular, we find that properties shared by all two-dimensional lattices allow us to develop new methods for estimating the growth constants and certain amplitudes for the square and triangular lattices.

The subclasses of self-avoiding walks and polygons that we consider are typically defined by imposing restrictions on the way in which a walk or polygon can be constructed. Ideally the restrictions should be as weak as possible, so as to result in a model that closely resembles the unrestricted case, while still enabling some manner of simple recursive construction. These recursions can sometimes lead to solutions for generating functions or other quantities of interest. Some of the models we consider display quite unusual asymptotic properties despite the relatively simple restrictions which lead to their construction.

We approach the modelling of polymer adsorption in several ways. Firstly, we adapt some of the new methods for studying self-avoiding walks on the honeycomb lattice to account for interactions with an impenetrable surface. In this way we are able to prove the exact value of the critical surface fugacity for adsorbing walks, confirming an existing conjecture. Then, we show that some key identities for the honeycomb lattice model lead to a new method for estimating the critical surface fugacities for adsorption models on the square and triangular lattices. Many of the estimates we obtain in this way are new; for the cases where previous estimates did already exist, our results are several orders of magnitude more precise. Finally, we define some new solvable models of polymer adsorption which generalise existing models, and find that some of these models exhibit interesting and unexpected critical behaviour.

Declaration

This is to certify that

    i. the thesis comprises only my own original work towards the PhD except where indicated in the Preface,

    ii. due acknowledgment has been made in the text to all other material used,

    iii. the thesis is fewer than 100 000 words in length, exclusive of tables, maps, bibliographies and appendices.

———————————————————————

Nicholas Ross Beaton

Preface

Much of the content of Section 2.2 was worked on in collaboration with Anthony Guttmann and Iwan Jensen, and has been published in [12]. The ideas arose from conversations with Guttmann. Jensen computed the series data and Guttmann and I calculated estimates and extrapolations.

The work on 3-sided prudent polygons (i.e. most of Subsections 3.2.3 and 3.2.4) was in collaboration with Philippe Flajolet and Anthony Guttmann, and has been published in [10, 11]. Flajolet contributed many of the ideas of the proof, such as performing a Mellin transform analysis, as well as providing a number of technical details and references to useful results. I calculated the original generating functions and computed series data, and was responsible for many of the technical details and calculations. Guttmann contributed series analysis and provided a number of useful references.

Some of the results of Section 3.3 appear in [9], which is primarily a review article. Those which appear here are entirely my own, except for some numerical results which are appropriately attributed in the text.

Subsections 4.1.1–4.1.3 are joint work with Mireille Bousquet-Mélou, Jan de Gier and Anthony Guttmann, and most of the results therein can be found in [8]. The original idea was Guttmann's, and he provided direction and useful references throughout. De Gier contributed to the ideas of Subsection 4.1.1 as well as the idea of Proposition 4.9. Bousquet-Mélou provided useful comments and many of the technical details in Subsections 4.1.2 and 4.1.3. The formulation of the results of Subsections 4.1.1 and 4.1.3 is primarily my own work, as was computation of a number of generating functions which assisted with our initial investigations.

An article presenting the results of Subsection 4.1.4 is currently in preparation.

Section 4.2 is joint work with Anthony Guttmann and Iwan Jensen, and the results have been published in [13]. The original idea arose from conversations with Guttmann. Jensen computed series data while Guttmann and I calculated series estimates and extrapolations.

The ideas and motivation behind Subsections 4.3.2 and 4.3.3 arose in conversation with Gary Iliev. The results presented here are my own. An article containing some of these results is currently being prepared by Iliev and myself.

Appendix A is mostly due to Hugo Duminil-Copin, with some contributions by Mireille Bousquet-Mélou. It is included here because of its importance in relation to the result of Subsec-

tion 4.1.3. It is largely a reproduction of material presented in [8].

Subsection B.1.2 is almost entirely due to Iwan Jensen, and is included here for completeness. It is largely a reproduction of material appearing in [12, 13].

Acknowledgments

First and foremost, my sincerest thanks to Tony Guttmann. From the moment I arrived in Melbourne, he has been an unwavering source of guidance. Not only has he offered helpful advice for just about every problem I've ever brought to him; he has provided a never-ending stream of new ideas, many of which grew into results in this thesis. On top of all this he has, against the odds, managed to get me to enjoy running, though I can but dream of ever approaching his level of fitness.

This thesis would not be possible without the financial support provided to me by a number of organisations, including

- the Australian Government (in the form of an Australian Postgraduate Award),

- the ARC Centre of Excellence for Mathematics and Statistics of Complex Systems (MASCOS, who provided me with a special studentship and funding for travel to numerous conferences and workshops),

- the University of Melbourne (in the form of a Melbourne Abroad Travelling Scholarship),

- the Mathematical Sciences Research Institute (who provided funding and assistance for travel to Berkeley), and

- the Department of Mathematics and Statistics at the University of Melbourne (who provided me with additional employment in the form of tutoring).

Additional thanks goes to all the hardworking and dedicated staff of the Department of Mathematics and Statistics, who not only provided me with a computer and a comfortable place to do maths but also offered assistance with incomprehensible paperwork and occasional server meltdowns.

I thank all the colleagues with whom I have had the pleasure of working, including Mireille Bousquet-Mélou, Iwan Jensen, Tim Garoni, Simone Rinaldi, Hugo Duminil-Copin and Richard Brak. Special thanks to Gary Iliev, for inviting me into his home on the other side of the world. Also to Jan de Gier, who not only amazes with brilliant ideas at wee hours of the morning, but can also swing a squash racquet like nobody's business. And, with a note of sadness, to Philippe

Flajolet, who was friendly, funny, and nothing short of a genius. The mathematics community has truly lost one of its greatest assets.

To my family, your unceasing love and support have carried me through the good times and the bad, and I would be hopelessly lost without you.

To the many friends I have made in Melbourne, be it through the department, at a pub trivia night, on the squash court, over a cup of coffee, while discussing a good (or perhaps not so good) book or while trying to finish a crossword, I thank you for making the last three years so enjoyable.

To Melbourne, you've been an awesome place to eat, drink, live and work, though I wish you'd lighten up a bit in winter.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

Combinatorics is an area of mathematics where remarkably simple problems often have surprisingly complicated solutions.[1] The problem of enumerating self-avoiding walks is one such; it is very easily stated, but seems extremely difficult to solve exactly.

Formally, a *self-avoiding walk* (SAW) ω on a graph $G = (V, E)$ is a sequence of vertices $(\omega_0, \omega_1, \ldots, \omega_n) \subseteq V$ such that $\{\omega_i, \omega_{i+1}\} \in E$ and $\omega_i \neq \omega_j$ for $i \neq j$. If ω contains $n + 1$ vertices we say it has *length* $|\omega| = n$. The graph G is usually taken to be an infinite periodic graph with a high degree of symmetry, i.e. a lattice. For example, if we embed the square lattice in the Cartesian plane with a vertex at each point with integer coordinates, then

$$\omega = ((0,0), (0,1), (1,1), (2,1), (2,2))$$

is a self-avoiding walk of length 4. (See Figure 1.1 for an example of a SAW on the square lattice.)

As suggested above, the central question considered by those who study SAWs is very simple:

For a given graph G, what is the number c_n of self-avoiding walks of length n on the vertices of G?

Clearly this question makes no sense for infinite graphs[2] – for example, our little SAW ω could be translated by any integer distance in the x or y directions and we would get another SAW of length 4. So perhaps a more sensible question might be:

For a given graph G, what is the number c_n of distinct self-avoiding walks (up to automorphisms of G) of length n on the vertices of G?

Now this has eliminated the possibility of translations, reflections and rotations on infinite lattices, and in our example c_4 will be a meaningful finite number.

---

[1]Perhaps the most famous such problem is that of the four-colour theorem, which states that the vertices of any planar graph can be properly coloured with at most four colours. The statement of the theorem can be very easily explained, but after being formally conjectured in 1852 it took 124 years for a correct proof to be produced, and even then the proof required hundreds of pages and extensive use of computer software.

[2]Excluding pathological graphs like the one without any edges, anyway.

In fact, when working on regular lattices, it is standard practice to allow reflections and rotations and forbid only translations. Since it is this type of graph that is the focus of this thesis, we finally phrase the question as it is addressed here:

For a given regular lattice L, what is the number c_n of distinct self-avoiding walks (up to translation) on the vertices of L?

Defining walks up to translation is of course equivalent to fixing the starting vertex, and so this point is usually taken to be the origin.

The difficulty of this problem becomes evident as soon as one attempts to find the first few terms of the sequence $\{c_n\}$. Consider for example the square lattice. The empty walk, containing only the origin, is defined to be a walk of length 0, so we have $c_0 = 1$. Then clearly $c_1 = 4, c_2 = 12, c_3 = 36$. The prevailing pattern suggests that the next term ought to be 108, but actually $c_4 = 100$, because certain walks of length 3 (e.g. $((0,0),(0,1),(1,1),(1,0))$) cannot be extended in every direction.

In essence this is why determing the sequence $\{c_n\}$ is so difficult – whether we are able to append a step to the end of a walk ω of length $n-1$, to form a walk of length n, depends on the entire structure of ω.[3] So the number c_n depends not just on the value of c_{n-1}, or even the entire sequence $\{c_0, c_1, \ldots, c_{n-1}\}$, but instead on the actual structures of each of the walks counted by c_{n-1}. This means there is no way to exploit a recursive relation on the c_n in order to find an exact solution.

This all may seem rather pessimistic; it appears that we've already proclaimed our main question unanswerable! Nothing, however, could be further from the truth. While there are no regular lattices in two or more dimensions for which we have an exact expression for c_n, there are still a number of results which have been proven about the behaviour of c_n, and many more conjectures. There are also enumerative techniques far more advanced than simply attempting to generate walks of length n by attaching steps to walks of length $n-1$, and remarkably long sequences are known for some lattices.

Going beyond the simple enumeration of SAWs of length n, there are many other questions we can ask about their behaviour:

- *As n grows large, what "size" or "shape" do SAWs of length n tend to assume? For example, how far is the endpoint likely to be from the origin?*

- *How likely is it that a walk will become "trapped" and be unable to step any further?*

- *How do the answers to these questions vary if we consider different lattices? Or higher dimensions?*

- *How do SAWs relate to other combinatorial objects? What about models in other areas, like physics and chemistry?*

---

[3] Self-avoiding walks are thus sometimes called *non-Markovian*.
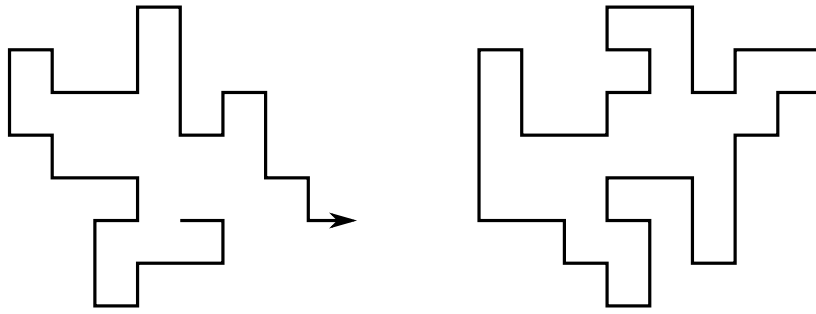
Figure 1.1: A self-avoiding walk of length 34 and an unrooted self-avoiding polygon of perimeter 44 and area 25 on the square lattice.

The goal of this thesis is to address some of these and other questions. We will do so using a number of different methods, and obtain results both rigorous, numerical and conjectural. Still, we are only scratching at the surface of all there is to know about self-avoiding walks.

Before moving onto a discussion of background and currently-known results, it seems prudent to explain the second part of the title of this thesis; that is, polygons. A *rooted self-avoiding polygon* (SAP) of perimeter $n \geq 4$ is a self-avoiding walk of length $n - 1$ whose endpoint is adjacent to its starting point. That is, a SAW $\omega = (\omega_0, \omega_1, \ldots, \omega_{n-1})$ is a SAP if ω_0 and ω_{n-1} are connected by an edge of the lattice. One usually imagines inserting the final edge to connect the first and last points, forming a closed loop (i.e. a polygon in the traditional sense of the word).

An *unrooted self-avoiding polygon* of perimeter n is a sequence of distinct lattice points

$$(\omega_0, \omega_1, \ldots, \omega_{n-1})$$

with ω_i adjacent to ω_{i+1} (and ω_0 adjacent to ω_{n-1}), defined up to cyclic permutations or order reversal of the ω_i (and translations, of course). Intuitively, an unrooted polygon is just a closed loop on the lattice with no vertex occurring more than once. (See Figure 1.1 for an example of an unrooted SAP on the square lattice.)

If p_n is defined as the number of rooted polygons of perimeter n and u_n the number of unrooted polygons of perimeter n, then it is easy to see that

$$p_n = 2n u_n. \tag{1.1}$$

(An unrooted polygon can be identified with a rooted one by selecting a vertex as the starting point and deleting one of the occupied edges adjacent to it. There are $2n$ ways to do this.)

3

Many of the questions mentioned above regarding SAWs also apply to SAPs – not only are we interested in their enumeration, but also their shape, size, and relationships with other objects and models. As will be discussed at length in Chapter 3, polygons can also be enumerated by *area*, rather than perimeter, and there are a number of differences in the methods used between the two approaches.

Self-avoiding walks are generally considered to have been introduced by Orr in the middle of the last century [93].[4] Orr, a polymer chemist, was interested in finding a mathematical model for long polymer molecules in solution which incorporated the *excluded volume effect* – the property that no pair of monomers (the individual units) of a molecule can occupy the same point in space. Prior to this, polymer models usually used random walks and completely disregarded the excluded volume effect. Other early related work, also in the field of polymer chemistry, was done by Flory [45].

Flory and Orr were not particularly interested in enumerating SAWs – they were concerned with the conformational properties of polymers, such as the average distance between endpoints and the density of monomers in solution. The earliest work to take into account the behaviour of c_n was by Hammersley and his collaborators [20, 60]. Hammersley noted that a SAW could be split at any vertex along its length into two shorter SAWs,[5] giving the very important relationship

$$c_{m+n} \leq c_m c_n, \qquad m, n \geq 0. \tag{1.2}$$

(Such an inequality implies that the sequence $\{c_n\}$ is *submultiplicative*.) If we take the logarithm of both sides of (1.2), we obtain

$$\log c_{m+n} \leq \log c_m + \log c_n,$$

and so the sequence $\{\log c_n\}$ is *subadditive*. Taking into account the finiteness of c_n, we are then able to apply Fekete's Lemma to obtain:

Lemma 1.1 (Hammersley). *The limit*

$$\lim_{n \to \infty} \frac{1}{n} \log c_n = \varkappa \tag{1.3}$$

exists and is equal to

$$\varkappa = \inf_{n \geq 1} \frac{1}{n} \log c_n. \tag{1.4}$$

---

[4]Though such a simple combinatorial object was likely considered by somebody well before then.

[5]Technically, if we define SAWs as having to start at the origin, then the second component is a SAW which has been translated so that its starting point coincides with the first component's endpoint.

The finiteness of c_n and (1.4) guarantee that $\varkappa < \infty$, but for general subadditive sequences it is possible that this limit could equal $-\infty$. Here however this cannot occur, since for example we clearly have $c_n \geq 3$ for all n on any regular lattice in two or more dimensions. Thus $\varkappa \in [0, \infty)$.

Corollary 1.2.

$$c_n = \exp(\varkappa n + o(n)). \tag{1.5}$$

Here, $o(f(n))$ denotes any function $g(n)$ satisfying $\lim_{n \to \infty} g(n)/f(n) = 0$, and $f(n) \sim g(n)$ if $\lim_{n \to \infty} f(n)/g(n) = 1$. Thus, (1.5) can be written in the alternative form

$$c_n = C(n)\mu^n \tag{1.6}$$

where $C(n) = e^{(o(n))}$ and $\mu = e^\varkappa$.

The constants $\varkappa$ and μ are known as the *connective constant* and *growth constant* respectively, and depend on the lattice under consideration.[6] Very little is known rigorously about the subexponential term $C(n)$, though there are many conjectures. In fact, it is generally believed that the asymptotic form of c_n is given by

$$c_n \sim A n^{\gamma - 1} \mu^n \tag{1.7}$$

with A and γ constant and known as the *amplitude* and *critical exponent* respectively. This form for c_n is supported by both numerical evidence [73] and by arguments from statistical physics [89]. In fact Nienhuis [89] was able to provide a conjecture for the exact value of γ in two dimensions, namely $\gamma = 43/32$. He also made a compelling argument that on the honeycomb lattice, the growth constant μ should be exactly $\sqrt{2 + \sqrt{2}}$. A rigorous proof of this fact went undiscovered for some 28 years, until it was finally announced in 2010 by Duminil-Copin and Smirnov [34]. (Their proof is of crucial importance to a number of results in this thesis, and is explored in detail in Chapter 2.)

In two, three and four dimensions the asymptotic form (1.7) is only conjectural, but in higher dimensions it has been proven [65, 64] that

$$c_n \sim A \mu^n,$$

i.e. that $\gamma = 1$. It is expected that if (1.7) is correct in all dimensions then γ depends only on the dimension and not the specific structure of the lattice. (This property is known in statistical physics as *universality*.) In three dimensions there is no conjecture for the exact value of γ, but numerical evidence and simulations [84] suggest $\gamma \approx 1.162$. Similarly, in four dimensions it is believed that $\gamma = 1$ with logarithmic corrections, so that $c_n \sim A \mu^n (\log n)^{1/4}$, but this has not been proven [85].

---

[6]Various articles in the literature instead refer to μ as the connective constant.

As regards the growth constant (which, unlike the critical exponent γ, has at least been proven to exist for all regular lattices), the only regular lattice[7] in two or more dimensions for which the exact value is known is the aforementioned honeycomb lattice, where it is equal to $\sqrt{2 + \sqrt{2}}$. For the other two-dimensional lattices, the best estimates for μ are [73, 75]

$$\mu_{\text{square}} \approx 2.63815853031(3)$$

and

$$\mu_{\text{triangular}} \approx 4.150797226(26),$$

based on analysis of the known values of the $\{c_n\}$ and $\{u_n\}$ sequences on those lattices.[8]

We mention here that, while expressions like (1.6) suggest that $\mu = \lim_{n \to \infty} c_{n+1}/c_n$, this result has not been proven in two, three or four dimensions[9] (though numerical evidence does support this conjecture). Kesten proved [78] that

$$\lim_{n \to \infty} \frac{c_{n+2}}{c_n} = \mu^2, \tag{1.8}$$

and O'Brien [90] has proven that $c_{n+1} > c_n$ for all $n \geq 0$.

The statement above, that current estimates of μ are based on known values of c_n and u_n, raises the question as to the relevance of the values of u_n, i.e. the number of self-avoiding *polygons* of perimeter n, to the value of the growth constant μ for self-avoiding *walks*. In fact the asymptotic behaviours of c_n and u_n are intimately connected.

Lemma 1.3 (Hammersley [61]). *For a given regular lattice, the limit*

$$\lim_{n \to \infty} \frac{1}{n} \log u_n \tag{1.9}$$

exists and is equal to $\varkappa$, the connective constant of self-avoiding walks on the lattice.

The proof (omitted for brevity) involves showing that any two polygons of perimeters m and n can be concatenated to form a unique polygon of perimeter $m + n$, and then bounding the number of polygons in terms of the number of walks.

---

[7]The exact value is known for some non-regular lattices, like the (3.12^2) lattice [54]. (Lattices like these are often called *quasi-regular*.)

[8]The value for the square lattice is indistinguishable from the smallest positive root of $581x^4 + 7x^2 - 13$, but thus far no one has produced an independent argument in support of this value.

[9]Kesten [78] does remark that this can be shown for the triangular lattice in two dimensions, and the same idea could be used for the face-centred cubic latticein three dimensions.

Just as with SAWs, this result implies that

$$u_n = U(n)\mu^n \tag{1.10}$$

with $U(n) = \exp(o(n))$. Also like the SAW case, it is generally believed that the asymptotic form of u_n is given by

$$u_n \sim B n^{\alpha-3} \mu^n \tag{1.11}$$

with B and α constant. The critical exponent α is, like γ, expected to depend only on the dimension of the lattice in question. In two dimensions there is strong evidence [89] that $\alpha = 1/2$, while in three dimensions numerical studies suggest [53] $\alpha \approx 0.23721$. In dimension four it is expected [53] that $\alpha = 0$ with a logarithmic correction term, while in five or more dimensions it is known that $\alpha = 0$.

As mentioned earlier, it is possible to enumerate SAPs in two dimensions by *area*, rather than perimeter.[10] If a_n denotes the number of SAPs with area n, then simple concatenation arguments like the one Hammersley uses for perimeter enumeration suffice to show

$$\lim_{n \to \infty} \frac{1}{n} \log a_n = \log \phi \tag{1.12}$$

exists and is equal to

$$\log \phi = \sup_{n \geq 1} \frac{1}{n} \log a_n.$$

Hence

$$a_n = A(n)\phi^n \tag{1.13}$$

with $A(n) = \exp(o(n))$. As with c_n and u_n, it is generally believed that

$$a_n \sim D n^\tau \phi^n \tag{1.14}$$

for constants D and τ. The exponent τ is thought to be -1, and current estimates [53] for ϕ are

$$\phi_{\text{square}} \approx 3.970944(2)$$
$$\phi_{\text{triangular}} \approx 2.9446596(3)$$
$$\phi_{\text{honeycomb}} \approx 5.16193016(3).$$

Asymptotic formulas like (1.6) and (1.10) are rigorous but rather vague, owing to the fact that we only know that $C(n)$ and $U(n)$ are subexponential. There also exist rigorous *bounds* on quantities like c_n which, though ultimately quite weak, are at least precise.

---

[10]In higher dimensions there exist quantities which resemble "area", but their definition is quite complicated.

Lemma 1.4 (Hammersley and Welsh [63]). *On a hypercubic lattice in $d \geq 2$ dimensions, for any constant $C > \pi \sqrt{2/3}$, there exists an $N_0(C)$ independent of d such that*

$$c_n \leq e^{C\sqrt{n}} \mu^{n+1} \qquad \text{for all } n \geq N_0. \tag{1.15}$$

The proof of this important result involves a number of technical details, but the basic structure is:

- Any SAW can be split at its bottom vertex[11] into two SAWs which begin at their bottom vertices.[12]

- A SAW which begins at its bottom vertex can be decomposed into a sequence of self-avoiding *bridges* – walks whose starting point has minimal x-coordinate and end point has (strictly) maximal x-coordinate.[13] The number b_n of bridges of length n has the same growth rate as SAWs; that is,

$$\lim_{n \to \infty} \frac{1}{n} \log b_n = \log \mu. \tag{1.16}$$

- The number of ways of concatenating a sequence of bridges to form a SAW of length n which starts at its bottom vertex is bounded above by the number $P_D(n)$ of strict partitions[14] of n. Hardy and Ramanujan [66] showed that

$$P_D(n) \sim \exp\left(\pi \sqrt{\frac{n}{3}}\right). \tag{1.17}$$

- The above can then be combined to give the lemma.

Since we clearly have that $p_n \leq c_n$, the bound of (1.15) applies equally well to p_n, and we can then use (1.1) to obtain a bound on u_n.

The following bound, due to Kesten (appearing in [85, Sec. 3.3]), is slightly stronger than that of Lemma 1.4 in three and four dimensions.

Lemma 1.5 (Kesten). *On a hypercubic lattice in $d \geq 2$ dimensions there exists a constant Q, depending on d, which satisfies*

$$c_n \leq \mu^n \exp(Q n^{2/(d+2)} \log n) \qquad \text{for } n \geq 2. \tag{1.18}$$

---

[11] If a walk $\omega = (\omega_0, \ldots, \omega_n)$ in d dimensions has vertices $\omega_i = (\omega_i(1), \ldots, \omega_i(d))$, then the bottom vertex is the point $\omega_B \in \omega$ satisfying

  (i) $\omega_B(1) \leq \omega_j(1)$ for all $0 \leq j \leq n$, and

  (ii) for $2 \leq m \leq d$ and $j \neq B$, $\omega_j(m) \leq \omega_B(m) \Rightarrow \exists m' < m$ such that $\omega_B(m') < \omega_j(m')$.

[12] The direction of one of the pieces will have to be reversed, of course.

[13] Technically the sequence alternates between bridges and reflected bridges.

[14] That is, the number of ways one can write $n = \sum_i \lambda_i$ with $\lambda_i \in \mathbb{Z}$ and $0 < \lambda_1 < \lambda_2 < \ldots$.

An important concept used extensively in this thesis is the *generating function* of a sequence. The (ordinary) generating function of SAWs on a given lattice is

$$Z(x) = \sum_{n=0}^{\infty} c_n x^n = \sum_{\omega} x^{|\omega|} \tag{1.19}$$

where the last sum is over all SAWs ω. We will often refer to the sequence $\{c_n\}$ as the *coefficients* of $Z(x)$. In a statistical physics context the variable x is known as an *activity* or *fugacity* and is normally taken to be real and non-negative.[15] When considering generating functions in combinatorics it is common to allow the variable x to be complex. We likewise define generating functions for the sequences p_n, u_n, a_n and so on.

The radius of convergence of $Z(x)$ is given by

$$\lim_{n \to \infty} c_n^{-1/n}$$

which is equal to μ^{-1} by Lemma 1.1. It also follows from (1.4) that

$$c_n \geq \mu^n, \tag{1.20}$$

and hence

$$Z(\mu^{-1}) = \infty. \tag{1.21}$$

(Of course, inequalities like (1.20) are not true for all sequences (polygons being one example), and so there are also cases where the generating functions are finite at μ^{-1}.)

The idea that generating functions can be used to determine growth constants is profoundly useful – for example, this is precisely how Duminil-Copin and Smirnov [34] prove that $\mu = \sqrt{2 + \sqrt{2}}$ on the honeycomb lattice. Likewise, in Chapters 3 and 4 we calculate a number of generating functions without directly determining the underlying sequences, and use these to determine the growth rates.

In fact, we can obtain a lot more information from the generating function than just the value of the connective constant or growth rate. The central tenet of *analytic combinatorics* [44] is that, if a generating function

$$F(x) = \sum_n f_n x^n$$

has radius of convergence ρ, then a complete asymptotic expansion of the coefficients f_n can be obtained by studying the behaviour of $F(x)$ at its *dominant singularities* – the singularities on the circle $|x| = \rho$. In particular, the contribution of a single isolated singularity is given in

---

[15] In statistical physics the function $Z(x)$ is sometimes known as the *susceptibility*, but we won't use this term.

Lemma 1.6 below [44, Thms. VI.1 & VI.4]. We first need a definition [44, Definition VI.1]: given two numbers ϕ, R with $R > 1$ and $0 < \phi < \pi/2$, the open domain $\Delta(\phi, R)$ is

$$\Delta(\phi, R) = \{z \mid |z| < R, \; z \neq 1, \; |\arg(z - 1)| > \phi\}.$$

A domain is a Δ-*domain* if it is a $\Delta(\phi, R)$ for some R and ϕ.

Lemma 1.6 (Flajolet and Sedgewick). *If*

$$F(x) = \sum_{n=0}^{\infty} f_n x^n$$

has a single dominant singularity at $x = x_0 \neq 0$ and can be continued to a domain of the form $x_0 \cdot \Delta_0$ for some Δ-domain Δ_0, where $x_0 \cdot \Delta_0$ is the image of Δ_0 by the map $z \mapsto x_0 z$, and

$$F(x) \underset{x \to x_0}{\sim} \frac{1}{(1 - x/x_0)^{\alpha}}, \qquad \alpha \in \mathbb{C} \backslash \mathbb{Z}_{\leq 0},$$

then

$$f_n \sim \frac{n^{\alpha - 1} x_0^{-n}}{\Gamma(\alpha)} \left(1 + \frac{\alpha(\alpha - 1)}{2n} + \frac{\alpha(\alpha - 1)(\alpha - 2)(3\alpha - 1)}{24n^2} + O(n^{-4}) \right). \tag{1.22}$$

In addition, [44, Thm. VI.2] extends this result to the cases when the behaviour of $F(x)$ around x_0 has a logarithmic term.

So for example, if the generating function $Z(x)$ displayed the behaviour

$$Z(x) \underset{x \to \mu^{-1}}{\sim} \frac{A'}{(1 - x\mu)^{43/32}} \tag{1.23}$$

then this could lead to

$$c_n \sim A n^{11/32} \mu^n,$$

just as conjectured (1.7) (with $A = A'/\Gamma(43/32)$). In fact, while Pringsheim's Theorem [44, Thm. IV.6] guarantees that $Z(x)$ must have a singularity at $x = \mu^{-1}$, it is believed [51, 58] that $Z(x)$ has a second singularity at $x = -\mu^{-1}$. In this case, as explained in [44, Thm. VI.5], the asymptotic form of c_n can be found by using the result of Lemma 1.6 to calculate the contribution of each singularity and then adding the contributions.

In Chapters 3 and 4 we will consider *multivariate* generating functions; that is, generating functions whose underlying sequences have more than one index. We do this for two reasons: firstly, because for some of the sequences f_n that we wish to compute (or at least, compute their generating functions), the only way to utilise a recursion on the coefficients is by introducing additional measurements. For example, to recursively generate walks which remain in a half-space, we might have to keep track of not only the length but also the distance between the endpoint of the walks and the boundary of the half-space.

Secondly, in Chapter 4 we will consider SAWs as a model of long chain polymers, and in these models it is often desirable to measure quantities besides the length of polymers. For

example, we consider polymers which interact with an impenetrable surface, and model this by studying walks in a half-space where we keep track of the number of times each walk visits the boundary of the half-space. By associating a fugacity with this count (i.e. by constructing the bivariate generating function) we can study walks which visit the surface a relatively large or small number of times.

Methods for relating the singularities of multivariate generating functions to their coefficients are covered in [44, Ch. IX], though for most of the problems considered in this thesis we are able to reduce the generating functions to univariate functions before attempting a singularity analysis.

As was mentioned at the start of this chapter, the enumeration of SAWs and SAPs is not the only problem considered by researchers in this field. Of equal importance to the *number* of SAWs or SAPs of a given length or perimeter is the *size* of these objects, and in particular the *average* size. For a SAW $\omega = (\omega_0, \omega_1, \ldots, \omega_n)$, there are three commonly-used metrics used to describe "size":

- The *squared end-to-end distance* is

$$R_e^2(\omega) = |\omega_n|^2. \tag{1.24}$$

- The *squared radius of gyration* is

$$R_g^2(\omega) = \frac{1}{2(n+1)^2} \sum_{i,j=0}^{n} |\omega_i - \omega_j|^2. \tag{1.25}$$

- The *squared distance of vertices from the endpoints* is

$$R_m^2(\omega) = \frac{1}{2(n+1)} \sum_{i=0}^{n} |\omega_i|^2 + |\omega_i - \omega_n|^2. \tag{1.26}$$

Here, $|\omega_i|$ is the usual Euclidean distance.

For unrooted SAPs, only one of the above three is defined:

- The *squared radius of gyration* of a SAP $\omega = (\omega_0, \omega_1, \ldots, \omega_{n-1})$ is

$$R_p^2(\omega) = \frac{1}{2n^2} \sum_{i,j=0}^{n} |\omega_i - \omega_j|^2. \tag{1.27}$$

With these quantities defined we are now most interested in their mean values with respect to the uniform measure;[16] for example, the *mean squared end-to-end distance* of SAWs of length n is

$$\langle R_e^2 \rangle_n = \frac{1}{c_n} \sum_\omega R_e^2(\omega), \qquad (1.28)$$

where the sum is over all SAWs of length n. We similarly define $\langle R_g^2 \rangle_n$, $\langle R_m^2 \rangle_n$ and $\langle R_p^2 \rangle_n$.

While very little has been proven about these quantities, it is widely believed that

$$\langle R_e^2 \rangle_n \sim E n^{2\nu} \qquad (1.29)$$

for constants E and ν, and likewise for $\langle R_g^2 \rangle_n$, $\langle R_m^2 \rangle_n$ and $\langle R_p^2 \rangle_n$, where the amplitude E varies between the four quantities but the exponent ν is the same. In fact, like γ and α, the exponent ν (if it exists) is expected to be the same for all lattices of a given dimension. In two dimensions there is substantial evidence [89, 83] to suggest that $\nu = 3/4$. In three dimensions it is estimated [84] that $\nu \approx 0.5876$, while in four dimensions it is believed [53] that $\nu = 1/2$ with logarithmic corrections, that is, $\langle R_e^2 \rangle_n \sim E n (\log n)^{1/4}$. In five or more dimensions it has been proven [65, 64] that $\nu = 1/2$.

As has been alluded to throughout this chapter, SAWs and SAPs on regular lattices have a strong connection to models in mathematical physics, and in particular statistical mechanics. While (as the title suggests) we are mostly interested in the *combinatorics* of SAWs and SAPs, such as their enumerative properties and generating functions, parts of this thesis (particularly Chapter 4) are motivated by the applications of SAWs to physics. Thus we dedicate part of this chapter to defining some key concepts in statistical mechanics and their connection with SAWs and SAPs.

Broadly speaking, statistical mechanics is the study of systems comprised of many individual components, and addresses how *local* interactions between components can produce large scale, *global* effects. A classical example is the model of ferromagnetism – each individual molecule in a bar of iron has a microscopic magnetic field, and in favourable conditions these magnetic fields influence those of their neighbours and cause them to become aligned. The collective alignment of a large fraction of the magnetic fields inside the bar produces a macroscopic field across the whole bar.

The first model that relates SAWs and SAPs to statistical mechanics is the *n-vector model*, introduced by Stanley in 1968 [107]. One considers a lattice of N sites (vertices), and to each site

---

[16]Of course there exist other probability measures on SAWs besides the uniform measure – one example is to consider SAWs as *kinetic growth walks* [86].

assigns an n-dimensional vector, called a *spin*, of magnitude $\sqrt{n}$. The *Hamiltonian*, or energy, of a particular configuration of spins is

$$\mathcal{H} = -\sum_{<i,j>} \mathbf{s}_i \cdot \mathbf{s}_j - H\sum_i s_i(1), \tag{1.30}$$

where $\mathbf{s}_i$ is the spin at site i and $s_i(1)$ is the first component of $\mathbf{s}_i$. The first sum is taken over all pairs of sites (i,j) which are adjacent on the lattice. The factor H represents an external magnetic field, assumed (without loss of generality) to be in the same direction as the first component of the $\mathbf{s}_i$. The $n = 1$ case is known as the *Ising model*, and is the simplest model of ferromagnetism. Other well-known models correspond to $n = 2, 3$ and the limit $n \to \infty$.

The *partition function* of the system of size N is

$$Z_N(T,H) = \sum_{\substack{\text{all} \\ \text{configurations}}} \exp(-\mathcal{H}/k_B T) \tag{1.31}$$

where k_B is Boltzmann's constant and T is the absolute temperature,[17] and the *free energy* is then

$$\varkappa(T,H) = \lim_{N\to\infty} \frac{-1}{N} k_B T \log Z_N(T,H). \tag{1.32}$$

(Taking the lattice size $N \to \infty$ is known as the *thermodynamic limit*.) Of relevance here is the quantity known as the *magnetic susceptibility*,

$$\chi(T) = -\left.\frac{\partial^2 \varkappa}{\partial H^2}\right|_{H=0}. \tag{1.33}$$

The susceptibility can be expanded as a power series, and the terms in the summation can be interpreted as weights assigned to graphs on the underlying lattice [85, 68]. De Gennes [25] observed that, in the (rather peculiar) limit $n \to 0$, the only such graphs with non-zero weight are non-intersecting paths from a site i to another site k. Thus the magnetic susceptibility of the n-vector model, in the limit $n \to 0$, is exactly the generating function of self-avoiding walks. This allows for some techniques from statistical mechanics, such as the renormalisation group and conformal field theory, to be applied to SAWs and SAPs.

The second set of models in statistical physics which are related to SAWs and SAPs are models of *polymers* and *vesicles*. Generally speaking, a polymer is a large molecule comprised of many repeated pieces (monomers). A vesicle is a bubble of fluid, typically found in biological cells (but they can also be artificial), which can be involved in a number of different processes within the cell. As was mentioned at the start of this chapter, SAWs were conceived by Flory and Orr [45, 93] as a model of polymers which accounted for the excluded volume principle.[18] Likewise, SAPs and related objects in higher dimensions (e.g. polycubes) are considered as models of vesicles.

---

[17] For $n \geq 2$ there are infinitely many possible values for the spin, and the sum (1.31) becomes an integral.

[18] Related objects like *self-avoiding trails* have also been used as models for polymers, but will not be discussed here.

Two phenomena which are frequently modelled with SAWs are polymer *adsorption* and *collapse*. In each of these cases, the system of size N is a SAW of length N (possibly in some kind of restricted geometry, like a half-space). The partition function is then

$$Z_N(\alpha) = \sum_\omega e^{\alpha c(\omega)} \qquad (1.34)$$

where the sum is over all SAWs ω of length N, and $c(\omega)$ is the number of *interactions* which occur in ω. In the case of adsorption, an interaction takes place when the walk ω visits the adsorbing surface; for example, $c(\omega)$ might count the number of edges in the surface which are occupied by ω. In the case of collapse, interactions will occur between sufficiently close (non-consecutive) monomers in ω; for example, a *nearest-neighbour* interaction occurs when two vertices in ω are adjacent on the lattice but not connected by an edge in ω. The variable (or *fugacity*) α is sometimes written as

$$\alpha = \frac{-\epsilon}{k_B T},$$

where ϵ is the energy associated with a single interaction, k_B is Boltzmann's constant and T is absolute temperature.

Here, the free energy is given by

$$\widehat{\varkappa}(\alpha) = \lim_{N \to \infty} \frac{1}{N} \log Z_N(\alpha). \qquad (1.35)$$

It can be shown [62] that the free energy is continuous but has a point of non-analyticity at $\alpha = \alpha_c$. In adsorption models, one can compute the *density of monomers in the surface* (in the limit of infinitely long polymers),

$$\delta(\alpha) = \frac{\partial \widehat{\varkappa}(\alpha)}{\partial \alpha}, \qquad (1.36)$$

and it follows that this density is 0 for $\alpha < \alpha_c$ and is > 0 for $\alpha > \alpha_c$. That is, the critical point α_c delimits the *desorbed* and *adsorbed* phases of the polymers. (See the introduction to Chapter 4 for further discussion regarding this density function.)

The behaviour of the free energy at the critical point α_c is thought to be characterised by a *crossover exponent* ϕ, satisfying

$$\widehat{\varkappa}(\alpha) - \widehat{\varkappa}(\alpha_c) \sim c(\alpha - \alpha_c)^{1/\phi} \qquad \text{as } \alpha \to \alpha_c^+ \qquad (1.37)$$

for some constant c.

Similarly, in the case of collapse the polymers undergo a collapse transition at α_c; for $\alpha < \alpha_c$, they tend to be stretched out and span a large region of space, while for $\alpha > \alpha_c$ the polymers are tightly wound and occupy a small region of space. (More precisely, we can consider metrics like the mean squared end-to-end distance or the squared radius of gyration, and observe that these averages display different behaviours in the $\alpha < \alpha_c$ and $\alpha > \alpha_c$ regimes.)

In the case of vesicles, researchers are primarily interested in the interplay between surface area and volume [120]. In two dimensions these reduce to perimeter and area, and so there are

two types of partition functions, *constant perimeter* and *constant area*. If $p_m(n)$ is the number of (unrooted) polygons with perimeter m and area n, then the partition functions are

$$P_m(q) = \sum_n p_m(n) q^n \tag{1.38}$$

and

$$A_n(x) = \sum_m p_m(n) x^m. \tag{1.39}$$

In the constant perimeter case, the critical point $q_c = 1$ delimits the collapsed phase, when fluid tends to flow out of the vesicle and the area shrinks, and the swollen phase, where fluid flows in and the area increases. As with the polymer models, we can then compute the free energy for either of these two cases. For example,

$$\varkappa_P(q) = \lim_{m \to \infty} \frac{1}{m} \log P_m(q). \tag{1.40}$$

From here, one can consider the behaviour of the free energy near the critical point $q_c = 1$ or the average area of perimeter m polygons as $m \to \infty$. Likewise, similar computations can be performed for the constant area case. See [4, 39] for further details.

We finally note that from partition functions like (1.34) and (1.38), we can define multivariate generating functions (or in the language of statistical mechanics, *grand canonical partition functions*) by associating a weight with the system size N and then summing over all N. For example, in the case of adsorbing polymers, we obtain

$$Z(x, \alpha) = \sum_{N=0}^{\infty} Z_N(\alpha) x^N. \tag{1.41}$$

For fixed α, the radius of convergence $\rho(\alpha)$ of $Z(x, \alpha)$ is given by

$$\rho(\alpha) = \lim_{n \to \infty} Z_n(\alpha)^{-1/n} = \exp(-\widehat{\varkappa}(\alpha)) \tag{1.42}$$

where $\widehat{\varkappa}(\alpha)$ is the free energy as per (1.35). So if we can compute the generating function and determine its dominant singularity as a function of α, then we get the free energy without having to first calculate the individual partition functions Z_N. In Chapter 4 we will make use of this property when we consider some solvable models of polymer adsorption.

This thesis is comprised of three main parts. In Chapter 2 we consider general self-avoiding walks on the three two-dimensional regular lattices. In particular, we focus on the recent

proof by Duminil-Copin and Smirnov [34] that the growth constant of the honeycomb lattice is $\sqrt{2 + \sqrt{2}}$. Unfortunately, it seems that their methods cannot be adapted to produce similar proofs for the other two-dimensional lattices. However, we argue that because all two-dimensional SAW models are expected to have the same conformal field theory in the appropriate scaling limit, we can expect some of the identities derived by Duminil-Copin and Smirnov to be "approximately" true for all two-dimensional lattices. Moreover, in the limit of the lattice size, the approximations should become exact identities.

Because these identities should hold only at the critical point of the SAW generating function (i.e. the reciprocal of the growth constant of the lattice), this leads to a powerful new method for computing numerical estimates of the growth constants of the square and triangular lattices. We use existing algorithms (namely the *finite lattice method*) to generate series for SAWs in finite-width strips of those lattices, and then use these series to compute a sequence of values $x^*(T)$ which we would expect, in the limit as $T \to \infty$, to converge to $x_c = \mu^{-1}$ for each lattice. We extrapolate these sequences and find that they do indeed appear to converge to the critical points of the square and triangular lattices, and in this way we compute estimates for the growth constants whose precision is only slightly less (1-2 digits) than that of the best currently-known estimates.

In Chapter 3 we turn our attention to subclasses of SAWs and SAPs for which we are able to calculate the exact generating functions or obtain recursive relations which allow fast computation of series coefficients. We first review some previously-studied models, including *directed* and *prudent* walks, before focusing on prudent polygons. The enumeration of these objects by *perimeter* has previously been considered by Schwerdtfeger [105], but our interest is in their enumeration by *area*.

A large section of the chapter is devoted to a particular subclass of prudent polygons, namely *3-sided prudent polygons*. We find that, while their area generating function is quite easily obtained, determining the singular behaviour and the asymptotic form of the coefficients requires some very careful analysis and a number of techniques from analytic combinatorics. While we believe that this fact alone makes the model worthy of considerable interest, the result itself is also very intriguing: the asymptotic form of the coefficients does not resemble that of any other lattice model we are aware of. Not only is the critical exponent *transcendental* where we would normally expect a small rational number, but the "amplitude" is not a constant, instead being a *periodic function* of $\log n$.

We then define and study some more general subclasses of square lattice SAWs and SAPs, namely *perimeter* and *quasi-prudent*. For some of these we are able to exactly solve the generating functions, while in other cases we are forced to resort to numerical results based on computer-generated series. Finally, we introduce some new solvable models on the triangular and honeycomb lattices, settings which seem to be rarely considered in the wider literature.

The third main part, Chapter 4, deals with applications of SAWs to models of polymer adsorption. We return to the ideas of Duminil-Copin and Smirnov, and show that some of their

key identities can be generalised to account for weights associated with visits to an impenetrable surface. Certain terms in our identities vanish at a particular value of the surface fugacity, and this value turns out to be precisely the critical surface fugacity as conjectured by Batchelor and Yung [7]. This allows us to formulate a proof of the critical surface fugacity, subject to the additional requirement that a certain generating function vanishes in the limit of the lattice size. The proof of that fact is presented in Appendix A. We also consider a different orientation of the surface for the honeycomb lattice, and show that in that case we can also obtain an identity involving the surface fugacity. The critical fugacity for that model was conjectured by Batchelor, Bennett-Wood and Owczarek [6]; we provide a proof of this value, subject to the same requirement as the original orientation.

We then take the ideas used in Chapter 2, where we computed estimates for the growth constants of the square and triangular lattices, and consider how they can be adapted for the purpose of estimating the critical surface fugacities of a number of adsorption models. We use the same arguments regarding the scaling limit of two-dimensional SAW models to infer that our identities should be "approximately" true for all finite-width lattice strips, and should become exact identities in the limit of the lattice size. The technique is then similar to the one used in Chapter 2: we generate series for finite-width strips, and use these series to compute a sequence of values which we expect to converge to the critical surface fugacity of the model in question. For all the models we consider (edge- and vertex-weightings on the square and triangular lattices and an additional vertex-weighted model on the honeycomb lattice) we find that the sequences do indeed converge, and comparison with existing estimates for the square lattice corroborates our expectation that the limits should be the critical fugacities. We thus obtain new estimates for the critical surface fugacities; to our knowledge, these are the only estimates for the triangular and honeycomb lattice models, while for the square lattice our estimates are several orders of magnitude more precise than existing values.

Finally, we consider solvable models of polymer adsorption. After briefly reviewing existing models based on directed walks, we introduce some new models based on *prudent* walks. These have the pleasing property that they do not have a directedness constraint; that is, they are able to take steps in all directions on the lattice. We find recursion relations satisfied by the series coefficients of these walks, and then determine their generating functions by solving functional equations which encode those recursions. A careful singularity analysis allows us to compute the free energy of one of these models and make a strong conjecture for the other. We find that the adsorption transitions of these models are rather unlike those of the simpler directed walks, and we attempt to find an intuitive explanation for these differences.

In Appendix A is the proof that the generating function of self-avoiding bridges on the honeycomb lattice which span a strip of width T, evaluated at the SAW critical point $x_c = 1/\sqrt{2 + \sqrt{2}}$, vanishes in the limit $T \to \infty$. This proof is largely due to Hugo Duminil-Copin, and is presented here due to its relevance to results in Chapter 4.

Appendix B contains a brief review of some of the methods used for generating and analysing the series discussed in this thesis.

Chapter 2

General SAWs and SAPs

The new methods and results presented in this chapter are largely inspired by the proof of Duminil-Copin and Smirnov [34], announced in 2010, that the growth constant of SAWs on the honeycomb lattice is $\sqrt{2+\sqrt{2}}$. Since our methods are so closely related to those used by Duminil-Copin and Smirnov, we reproduce their proof in Section 2.1 and provide some related identities for the $O(n)$ *loop model*, a generalisation of the self-avoiding walk model.

In Section 2.2 we consider how the result of Duminil-Copin and Smirnov can be applied to SAW models on other lattices – namely, the square and triangular lattices. While it seems that their methods cannot be used to compute the precise values of the growth constants of those lattices, it is possible to adapt them to obtain numerical estimates whose accuracy rivals those of any other currently-known method.

We will further extend the Duminil-Copin and Smirnov method in Chapter 4, when we examine the problem of polymer adsorption.

2.1 Exact results

The n-vector model, introduced by Stanley in 1968 [107] is described by the Hamiltonian

$$\mathcal{H}(d,n) = -J \sum_{\langle i,j \rangle} \mathbf{s_i} \cdot \mathbf{s_j},$$

where d denotes the dimensionality of the lattice, and $\mathbf{s_i}$ is an n-dimensional unit vector. When $n = 1$ this Hamiltonian describes the Ising model, and when $n = 2$ it describes the classical XY model. Two other interesting limits, which leave a lot to be desired from a pure mathematical perspective, are the limit $n \to 0$, in which case one recovers the self-avoiding walk model, as first pointed out by de Gennes [25], and the limit $n = -2$, corresponding to random walks, or more generally a free-field Gaussian model, as shown by Balian and Toulouse [3]. Of particular importance here is the fact that the n-vector model on the honeycomb lattice has been shown [29] to be equivalent to a loop model with a weight n attached to closed loops.

In 1982 Nienhuis [89] showed that, for $n \in [-2, 2]$, the model on the honeycomb lattice could be mapped onto a solid-on-solid model, from which he was able to derive the critical points and critical exponents, subject to some plausible assumptions. These results agreed with the known exponents and critical point for the Ising model, and predicted exact values for those models corresponding to other values of the spin dimensionality n. Smirnov [106] has recently derived an identity for the general honeycomb $O(n)$ model with $n \in [-2, 2]$, presented in Lemma 2.1. This identity provides an alternative way of predicting the value of the critical point $x_c(n) = 1/\sqrt{2 + \sqrt{2 - n}}$ as conjectured by Nienhuis for $n \in [-2, 2]$.

We define a *mid-edge* of the honeycomb lattice to be the point on an edge halfway between its two adjacent vertices. Let H be the set of mid-edges of the honeycomb lattice. We define a *domain* $\Omega \subset H$ to be a connected collection of mid-edges with the property that if two mid-edges $p, q \in H$ are adjacent to a vertex v, then the third mid-edge adjacent to v must also be in H. The set of vertices adjacent to the mid-edges of Ω is denoted $V(\Omega)$. Those mid-edges of Ω which are adjacent to only one vertex in $V(\Omega)$ form $\partial\Omega$. A *step* in Ω is a vertex $v \in V(\Omega)$ together with two of the three mid-edges adjacent to v. Two steps in Ω are *adjacent* if they contain different vertices but share a mid-edge.

A *configuration* γ in Ω is a set of steps in Ω which either

(i) is empty;

(ii) contains a single step;

(iii) has the property that every step is adjacent to two other steps in γ; or

(iv) has the property that every step is adjacent to two other steps in γ except for two special steps, which are each adjacent to only one other step in γ.

It is clear then that a configuration is comprised of a (possibly empty) collection of mutually disjoint closed loops together with a (possibly empty) self-avoiding walk which starts and ends on mid-edges of Ω. We denote by $|\gamma|$ the number of vertices in γ and by $c(\gamma)$ the number of closed loops. For configurations with a non-empty SAW component, we can label one of the endpoints a of the SAW as the starting point and the other endpoint z as the end, and we then denote by $W(\gamma : a \to z)$ the *winding angle* – the net angle (in radians) the walk turns through as it runs from a to z (i.e. each left step adds $\pi/3$ to the winding angle and each right step subtracts $\pi/3$). If the SAW component of γ is empty we just assign it a winding angle of 0.

We then define the following so-called *parafermionic observable*: for $a \in \partial\Omega$ and $z \in \Omega$,

$$F(\Omega, a, z; x, n, \sigma) := F(z) = \sum_{\gamma \subset \Omega : a \to z} e^{-i\sigma W(\gamma : a \to z)} x^{|\gamma|} n^{c(\gamma)},$$

where the sum is over all configurations whose self-avoiding walk component starts at a and ends at z. See Figure 2.1 for an example of a configuration and its contribution to $F(z)$.

Smirnov [106] proves the following:
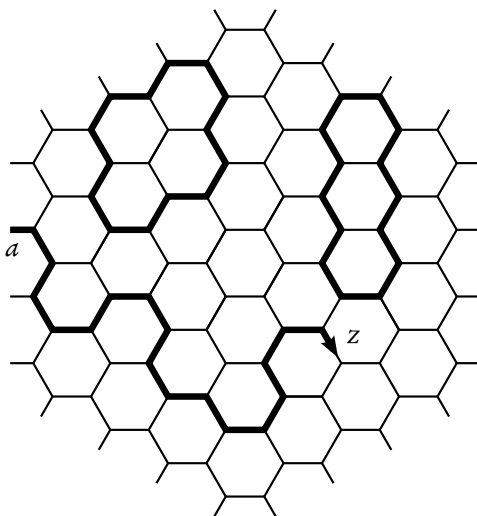
Figure 2.1: A configuration (the dark lines) on a domain of the honeycomb lattice. The contribution of this particular configuration to the observable $F(z)$ would be $e^{\pi i \sigma/3} x^{45} n^2$.

Lemma 2.1 (Smirnov). *For $n \in [-2, 2]$, set $n = 2\cos\theta$ with $\theta \in [0, \pi]$. Then for*

$$\sigma = \frac{\pi - 3\theta}{4\pi}, \qquad x^{-1} = 2\cos\left(\frac{\pi + \theta}{4}\right) = \sqrt{2 - \sqrt{2 - n}}, \qquad or \qquad (2.1)$$

$$\sigma = \frac{\pi + 3\theta}{4\pi}, \qquad x^{-1} = 2\cos\left(\frac{\pi - \theta}{4}\right) = \sqrt{2 + \sqrt{2 - n}}, \qquad (2.2)$$

the observable F satisfies the following relation for every vertex $v \in V(\Omega)$:

$$(p - v)F(p) + (q - v)F(q) + (r - v)F(r) = 0, \qquad (2.3)$$

where p, q, r are the three mid-edges adjacent to v, and the multipliers $(p - v)$, etc. are calculated by considering p, q, r and v as points in the complex plane.

The first pair of values (2.1) corresponds to the larger of the two critical values of the step weight x and thus describes the so-called dense regime, as configurations with many closed loops are favoured. The second pair of values (2.2) corresponds to the line of critical points separating the dense and dilute phases [89].

Proof. The lemma follows from considering contributions around the vertex v (see Figure 2.2). There are two possible cases. In the following, we define $\lambda = e^{-i\sigma\pi/3}$ (the weight accrued by a walk for each left turn) and $j = e^{2i\pi/3}$ (the value of $p - v$ when mid-edge p is to the north-west of its adjacent vertex v[1]).

---

[1]More specifically, j is the value of $p - v$ when p is to the north-west of its adjacent vertex v and the lattice is oriented as in Figure 2.1; that is, when the lattice contains horizontal edges. If we rotate the lattice by angle θ then that would simply result in the identity (2.3) being multiplied by $e^{i\theta}$.
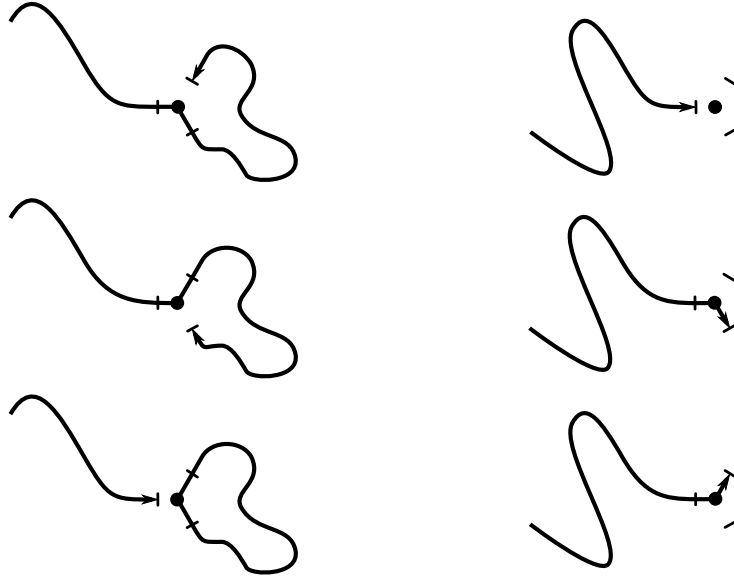
Figure 2.2: The two ways of grouping the configurations which end at mid-edges p, q, r adjacent to vertex v. On the left, configurations which visit all three mid-edges; on the right, configurations which visit one or two of the mid-edges.

1. In the first case (the left-hand side of Figure 2.2), all mid-edges p, q, r are visited by a configuration and hence two of the three edges incident on v must be occupied. There are three ways for this to occur: two with the self-avoiding walk visiting all three sites, and one with a closed loop running through v. If we fix the rest of the configuration not incident on v, the three contributions add up to zero if

$$\bar{j}\lambda^4 + j\bar{\lambda}^4 + n = 0.$$

This equation determines the possible values of the parameter σ. There are two solutions: if $\bar{j}\lambda^4 = -e^{i\theta}$ we get $\sigma = \frac{\pi - 3\theta}{4\pi}$, while if $\bar{j}\lambda^4 = -e^{-i\theta}$ we get $\sigma = \frac{\pi + 3\theta}{4\pi}$.

2. In the second case (the right-hand side of Figure 2.2) only one or two mid-edges are occupied in the configuration, and the condition that all contributions add up to zero becomes

$$1 + xj\bar{\lambda} + x\bar{j}\lambda = 0,$$

which leads to

$$x^{-1} = 2\cos\left(\frac{\pi}{3}(\sigma - 1)\right).$$

The two possible values give rise to the corresponding two values for x. ∎

Duminil-Copin and Smirnov [34] use Lemma 2.1 to prove the growth constant of self-avoiding walks ($n = 0$ in the dilute regime) is equal to $x_c^{-1} = 2\cos(\pi/8) = \sqrt{2 + \sqrt{2}}$. They establish an identity which relates generating functions of SAWs ending on the boundary of a
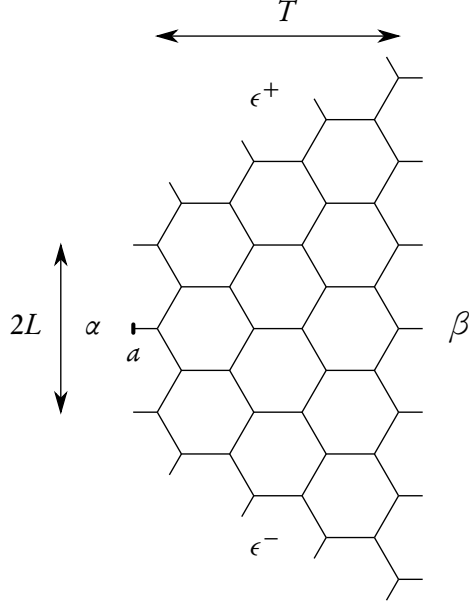
Figure 2.3: The special domain $\Omega = S_{3,1}$. The starting point of the SAW component of a configuration is fixed to be the middle of the left boundary, marked as a, and the mid-edges in $\partial\Omega$ are grouped into four sets $\alpha, \beta, \epsilon^+$ and ϵ^-.

special domain $\Omega = S_{T,L}$, as shown in Figure 2.3. In Lemma 2.2 we present the generalisation of this identity to arbitrary $n \in [-2, 2]$. The proof is identical to that of [34].

Given the domain $\Omega = S_{T,L}$, we define the generating functions

$$A_{T,L}(x) := \sum_{\gamma \subset S_{T,L}: a \to \alpha \setminus \{a\}} x^{|\gamma|} n^{c(\gamma)},$$

$$B_{T,L}(x) := \sum_{\gamma \subset S_{T,L}: a \to \beta} x^{|\gamma|} n^{c(\gamma)},$$

$$E_{T,L}(x) := \sum_{\gamma \subset S_{T,L}: a \to \epsilon^+ \cup \epsilon^-} x^{|\gamma|} n^{c(\gamma)},$$

where the sums are over all configurations whose SAW component runs from a to the α, β or ϵ^+, ϵ^- boundaries respectively. Furthermore define the special generating function

$$C_{T,L}(x, y) := \sum_{\gamma \subset S_{T,L}: a \to a} x^{|\gamma|} n^{c(\gamma)}$$

which sums over configurations comprising *only* closed loops inside $S_{T,L}$; that is, configurations whose self-avoiding walk component is the empty walk $a \to a$.

Lemma 2.2. *Let $n = 2\cos\theta$ with $\theta \in [0, \pi]$ and define*

$$x_c = \frac{1}{2} \cos\left(\frac{\pi \pm \theta}{4}\right)^{-1} = \left(2 \mp \sqrt{2 - n}\right)^{-1/2}.$$

Then

$$C_{T,L}(x_c) = \cos\left(\frac{3(\pi \pm \theta)}{4}\right) A_{T,L}(x_c) + \cos\left(\frac{\pi \pm \theta}{2}\right) E_{T,L}(x_c) + B_{T,L}(x_c). \tag{2.4}$$

23

Proof. Let p_v, q_v, r_v be the mid-edges adjacent to a vertex v. We set $x = x_c$ and compute the sum

$$\sum_{v \in V(S_{T,L})} (p_v - v)F(p_v) + (q_v - v)F(q_v) + (r_v - v)F(r_v) \qquad (2.5)$$

in two ways. Firstly, (2.3) holds for all $v \in V(S_{T,L})$, so (2.5) adds to 0. On the other hand, any mid-edge in $S_{T,L} \backslash \partial S_{T,L}$ will contribute to two terms in the sum, and these two terms will be precisely the negatives of one another, and hence will cancel. So the only mid-edges which do not cancel in this way are those in $\partial S_{T,L}$; their contributions are

$$-\sum_{z \in \alpha} F(z) + j \sum_{z \in \epsilon^+} F(z) + \bar{j} \sum_{z \in \epsilon^-} F(z) + \sum_{z \in \beta} F(z). \qquad (2.6)$$

Now SAWs which start at a and end at $z \in \alpha$ can be divided into three groups: those ending above a, below a, and the empty walk. By symmetry the generating functions of the first two groups are equal, and their winding angles will be π and $-\pi$ respectively. So we have

$$\begin{aligned}
\sum_{z \in \alpha} F(z) &= C_{T,L}(x_c) + \frac{\lambda^3 + \bar{\lambda}^3}{2} A_{T,L}(x_c) \\
&= C_{T,L}(x_c) + \cos\left(\frac{\pi \mp 3\theta}{4}\right) A_{T,L}(x_c)
\end{aligned} \qquad (2.7)$$

Similarly, by symmetry the generating functions of walks ending in ϵ^+ and ϵ^- are equal with winding angles $2\pi/3$ and $-2\pi/3$ respectively, and we obtain

$$\begin{aligned}
j \sum_{z \in \epsilon^+} F(z) + \bar{j} \sum_{z \in \epsilon^-} F(z) &= \frac{j\lambda^2 + \bar{j}\bar{\lambda}^2}{2} E_{T,L}(x_c) \\
&= \cos\left(\frac{\pi \pm \theta}{2}\right) E_{T,L}(x_c).
\end{aligned} \qquad (2.8)$$

Since walks ending in β have winding angle 0, we also have

$$\sum_{z \in \beta} F(z) = B_{T,L}(x_c). \qquad (2.9)$$

So equating (2.6) with 0 and substituting (2.7), (2.8) and (2.9), we obtain

$$0 = -C_{T,L}(x_c) - \cos\left(\frac{\pi \mp 3\theta}{4}\right) A_{T,L}(x_c) + \cos\left(\frac{\pi \pm \theta}{2}\right) E_{T,L}(x_c) + B_{T,L}(x_c),$$

and the lemma follows. ∎

Lemma 2.2 is further generalised in Chapter 4, where we introduce another weight y associated with occupied vertices in the β boundary. For now, however, we are only interested in self-avoiding walks, so we take (2.4) in the dilute regime and set $n = 0$:

$$1 = \cos\left(\frac{3\pi}{8}\right) A_{T,L}(x_c) + \cos\left(\frac{\pi}{4}\right) E_{T,L}(x_c) + B_{T,L}(x_c) \qquad (2.10)$$

with $x_c = \sqrt{2+\sqrt{2}}^{-1}$.

Duminil-Copin and Smirnov [34] then use (2.10) in their proof.

Theorem 2.3 (Duminil-Copin & Smirnov). *The growth constant of self-avoiding walks on the honeycomb lattice is*

$$\mu = x_c^{-1} = \sqrt{2+\sqrt{2}}.$$

All terms in (2.10) are non-negative, and $A_{T,L}(x_c)$ and $B_{T,L}(x_c)$ are non-decreasing with L (as L increases, more walks are counted by these generating functions). It follows that $E_{T,L}(x_c)$ is non-increasing with L, and so all the terms in (2.10) have limits as $L \to \infty$. Taking this limit, we obtain

$$1 = \cos\left(\frac{3\pi}{8}\right) A_T(x_c) + \cos\left(\frac{\pi}{4}\right) E_T(x_c) + B_T(x_c). \tag{2.11}$$

The generating functions $A_T(x)$ and $B_T(x)$ now count walks in a *strip* – $A_T(x)$ counts walks which start and end on the same side of the strip, while $B_T(x)$ counts walks which start and end on opposite sides of the strip. It is not entirely clear what the object $E_T(x)$ is (the walks it "counts" are *infinitely long*), but as we will see, it turns out $E_T(x)$ is irrelevant to our calculations.

Lemma 2.4.

$$x_c \leq \mu^{-1}.$$

Proof. Define $x_c(T)$ to be the radius of convergence of $A_T(x)$. By Lemma 6.1 of [115], we have $x_c(T) > x_c(T+1)$ (that result is for a hypercubic lattice, but the proof for the honeycomb lattice is identical; in Section 4.1 we will demonstrate a generalisation of this result in which we associate a fugacity y with visits to vertices in the β boundary), and then by Theorem 6.2 of [115],

$$\lim_{T \to \infty} x_c(T) = \mu^{-1}.$$

In particular, $x_c(T) > \mu^{-1}$ for all T. Since (2.11) implies that $A_T(x)$ is a convergent power series for $x \leq x_c$ and for all T, it follows that $x_c \leq x_c(T)$, which gives the lemma. ∎

Corollary 2.5.

$$E_T(x) \equiv \lim_{L \to \infty} E_{T,L}(x) = 0 \qquad \text{for } x \leq x_c.$$

Proof. If $Z_T(x)$ is the generating function for *all* walks in a strip of width T starting at the point a, then $Z_T(x)$ has the same radius of convergence as $A_T(x)$. (The equivalent result for the square lattice is Corollary 4.7 of [115]. We again prove a generalisation of this result for the honeycomb lattice in Section 4.1.) Then by Lemma 2.4, $Z_T(x) < \infty$ for all $x \leq x_c$ and for all T. In particular,

$$\sum_{L \geq 0} E_{T,L}(x_c) \leq Z_T(x_c) < \infty,$$

so $E_{T,L}(x_c) \to 0$ as $L \to \infty$. ∎

So (2.11) can be reduced to

$$1 = \cos\left(\frac{3\pi}{8}\right) A_T(x_c) + B_T(x_c). \tag{2.12}$$

Proof of Theorem 2.3. It remains to be shown that $x_c \geq \mu^{-1}$. Since μ^{-1} is the radius of convergence of $Z(x)$ (the generating function of all half-plane walks; this follows from arguments in e.g. [118]), it suffices to show

$$Z(x_c) = \infty.$$

First, note that any walk counted by $A_T(x)$ but not $A_{T-1}(x)$ must span the entire strip of width T. Such a walk can then be cut at the first visit to the β boundary, and by adding half-edges to the two pieces we obtain a unique pair of walks counted by $B_T(x)$. So it follows that

$$A_T(x) - A_{T-1}(x) \leq x B_T(x)^2. \tag{2.13}$$

If we take $x = x_c$ in (2.13) then we can substitute (2.12), and we obtain

$$x_c \cos\left(\frac{3\pi}{8}\right) B_T(x_c)^2 + B_T(x_c) \geq B_{T-1}(x_c).$$

It then follows by induction that

$$B_T(x_c) \geq \min\left[B_1(x_c), \sec\left(\frac{3\pi}{8}\right) x_c^{-1}\right]/T, \tag{2.14}$$

and so

$$Z(x_c) \geq \sum_{T \geq 0} B_T(x_c) = \infty. \qquad \blacksquare$$

2.2 Numerical estimates

Unfortunately the proof of Duminil-Copin and Smirnov identifying the growth constant of the honeycomb lattice cannot be repeated for the square and triangular lattices, as there is no known appropriate parafermionic observable satisfying an identity like (2.3). As shown by Ikhlef and Cardy [69], the dilute $O(n)$ model on the square lattice does have a parafermionic observable, and this can be used to identify its critical point. In the $n \to 0$ limit, however, that model is not the usual SAW model.[2]

However, it was pointed out by Cardy [22] that in the appropriate *scaling limit* all two-dimensional self-avoiding walk models, at their respective critical points, have the same *conformal field theory* [28, 69, 106]. It follows that we may expect an "identity" like (2.12) to hold for these other lattices in the limit $T \to \infty$. This raises two important questions:

1. How do these "identities" differ from (2.12)? For example, do the constants change, or can we identify correction terms which vanish as $T \to \infty$?

---

[2]It is a model where walks are allowed to visit *vertices* more than once but may not repeat *edges*.

2. Can these "identities" be used to obtain new results, such as accurate estimates for the growth constants of these lattices?

We attempt to answer the first question by calculating data for the generating functions $A_T(x)$ and $B_T(x)$ in strips, for $T \leq 15$ on the square lattice and for $T \leq 11$ on the triangular lattice (we also compute data for $T \leq 10$ on the honeycomb lattice). Using the best available estimates for the critical points of these models, namely [77, 72, 75]

$$x_c(\text{sq}) \approx 0.37905227776$$
$$x_c(\text{tr}) \approx 0.2409175745 \tag{2.15}$$

(with uncertainty in the last digit of each), we search for solutions to "identities" of the form

$$1 = c_\alpha(T)A_T(x_c) + c_\beta(T)B_T(x_c)$$
$$1 = c_\alpha(T)A_{T+1}(x_c) + c_\beta(T)B_{T+1}(x_c) \tag{2.16}$$

and examine the behaviour of $c_\alpha(T)$ and $c_\beta(T)$ as T increases.[3]

More details appear below in Subsections 2.2.2 and 2.2.3, but in summary we find a weak T dependence in both coefficients $c_\alpha(T)$ and $c_\beta(T)$. More significantly however, we conjecture that

$$\lim_{T \to \infty} c_\alpha(T)/c_\beta(T) = \cos(3\pi/8), \tag{2.17}$$

just as in the honeycomb lattice case, based on agreement to more than 5 significant digits for both the square and triangular lattices. In hindsight, this is perhaps not too surprising, as the constants multiplying the two generating functions arise from the winding angle of contributing graphs, and these are independent of lattice for the two generating functions considered, being $\pm\pi$ radians for $A_T(x)$ and 0 for $B_T(x)$.

This behaviour suggests an affirmative answer to our second question, by implying that by seeking solutions $x^*(T)$ to

$$\cos(3\pi/8)A_T(x) + B_T(x) = \cos(3\pi/8)A_{T+1}(x) + B_{T+1}(x),$$

we would expect to find $x^*(T) \to x_c$ as $T \to \infty$. Indeed, in this way we estimate

$$x_c(\text{sq}) = 0.3790522775 \pm 0.0000000005$$

and

$$x_c(\text{tr}) = 0.240917572 \pm 0.000000005.$$

These estimates are comparable to those in (2.15) obtained by other methods.

---

[3]Where no confusion should arise we will use x_c in this section to denote the critical point of any of the three regular two-dimensional lattices; in cases of ambiguity, we will specify the lattice by $x_c(\text{sq})$, etc.

Finally, we study numerically the behaviour of the two generating functions $A_T(x)$ and $B_T(x)$ as $T \to \infty$. We compute estimates of $a_0 = A(x_c)$, on the square and triangular lattices. We also find that

$$B_T(x_c) \sim \text{const.}/T^\alpha,$$

where $\alpha \approx 0.25$, which matches a prediction of Duminil-Copin and Smirnov [34]. Similarly, we can investigate how $\tilde{A}_T(x_c) = A_T(x_c) - a_0$ behaves as T tends to infinity. If $B_T(x_c)$ does indeed decay like $T^{-1/4}$ then we should expect the same behaviour in $\tilde{A}_T(x_c)$, and this is observed numerically.

We note here that the estimates for critical points, amplitudes, etc. for the square and triangular lattices computed in this section are *biased* estimates – that is, their validity is reliant upon unproven assumptions about the limiting behaviour of $A_T(x_c)$ and $B_T(x_c)$. (Namely, our estimates depend on the validity of statements like (2.16) and (2.17).) Of course, since the precision obtained here is no greater than that of existing (unbiased) estimates, this poses no great problem at this stage.

2.2.1 Honeycomb lattice

Duminil-Copin and Smirnov proved that on the honeycomb lattice, the unique solution of

$$\cos(3\pi/8)A_T(x) + B_T(x) = 1, \tag{2.18}$$

for any $T \geq 0$, occurs at $x = x_c = 1/\sqrt{2 + \sqrt{2}}$. (Their identity also involved an $E_T(x)$ term, but by Corollary 2.5 this term can be ignored.) It follows then that we could work backwards: given only the simple rational generating functions $A_0(x)$ and $B_0(x)$, we could identify the exact value of x_c simply by seeking the solution of

$$\cos(3\pi/8)A_0(x) + B_0(x) = 1.$$

If we did not already know x_c (but knew it was the value which satisfies this identity), this would be a particularly simple way to find it.

In a further demonstration of this invariant, we show in Figure 2.4 a plot of $\cos(3\pi/8)A_T(x) + B_T(x)$ for $T \in [1 \dots 10]$, where it can be seen that the curves intersect at $(x_c, 1)$, in accordance with the identity (2.12).

Let us assume that we didn't even know the Duminil-Copin and Smirnov identity (2.12), but rather just conjectured that some linear combination of $A_T(x_c)$ and $B_T(x_c)$ was invariant. We write this invariant as $\lambda A_T(x_c) + B_T(x_c)$. Then by seeking the solutions, for $x > 0$, $\lambda > 0$ of the equations

$$\lambda A_0(x) + B_0(x) - \lambda A_1(x) - B_1(x) = 0,$$

and

$$\lambda A_2(x) + B_2(x) - \lambda A_1(x) - B_1(x) = 0,$$
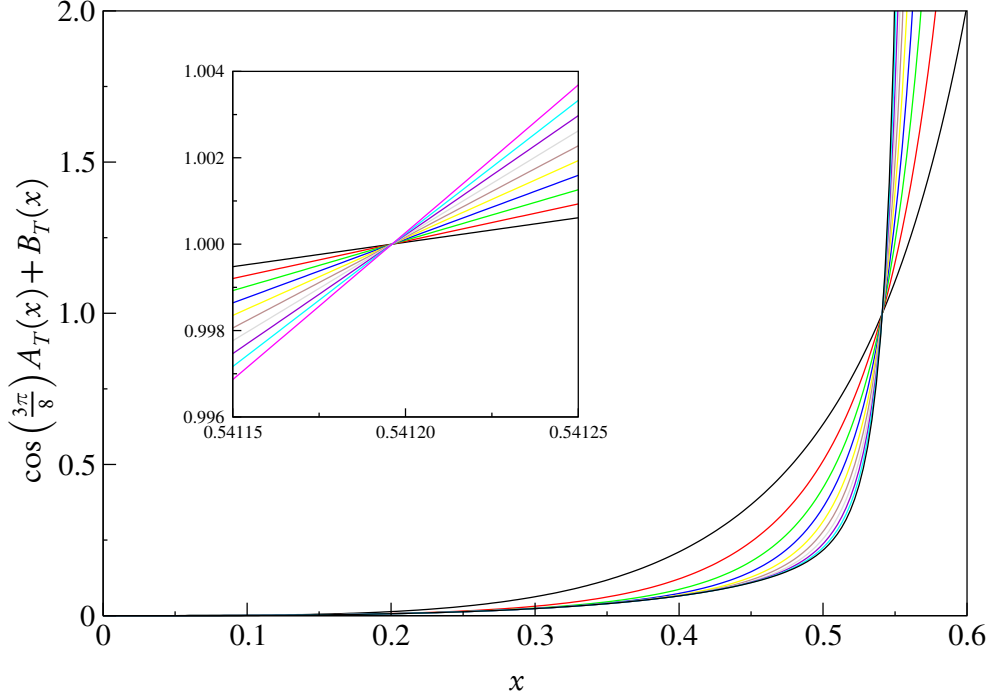
Figure 2.4: Plot of $\cos(3\pi/8)A_T(x) + B_T(x)$ for $T \in [1..10]$, for the honeycomb lattice, showing intersection of all curves at $(x_c, 1)$. The insert shows a close-up of the region of intersection.

we could discover both the invariant and the exact value of the critical point from the exact solutions for strips of small width. As we show below, this suggests a way to *approximate* x_c for other lattices, by similar means.

Next we consider the behaviour of the generating functions $A_T(x)$ and $B_T(x)$ in the limit $T \to \infty$. Denote $\lim_{T \to \infty} B_T(x)$ by $B(x)$, with a similar definition of $A(x)$. We wish to understand exactly how $B_T(x_c)$ behaves as $T \to \infty$, and in particular if we can verify the conjectured behaviour [34] $B_T(x_c) \sim c/T^{1/4}$. (Appendix A contains a proof that $B_T(x_c) \to 0$ as $T \to \infty$, though we still do not have a rigorous proof of the exponent.)

If $B_T(x_c) \sim \text{const.}/T^\alpha$, a log-log plot of $B_T(x_c)$ against T should be linear with slope $-\alpha$ as $T \to \infty$. We only have data for width $T \leq 10$, so the gradient is still changing slightly with T in that plot. To accommodate this, we extrapolate estimates of the local gradient of the log-log plot. We define the local gradient as

$$\text{gradB}(T) = \log\left(\frac{B_T(x_c)}{B_{T-1}(x_c)}\right) / \log\left(\frac{T}{T-1}\right),$$

(this is of course just the gradient of the line connecting two consecutive points in the log-log plot) and plot $\text{gradB}(T)$ against $1/T^{0.85}$, where the exponent 0.85 was chosen empirically to make the plot linear. The last few points of the plot are shown in Figure 2.5, and it is manifestly clear that the locus extrapolates to a value of $\alpha \approx 1/4$.

From (2.12) it follows that if $B_T(x_c) \sim c/T^{1/4}$, then $\tilde{A}_T(x_c) = A_T(x_c) - a_0$ also decays as $T^{-1/4}$, and this was observed numerically by a similar plot to Figure 2.5.

29
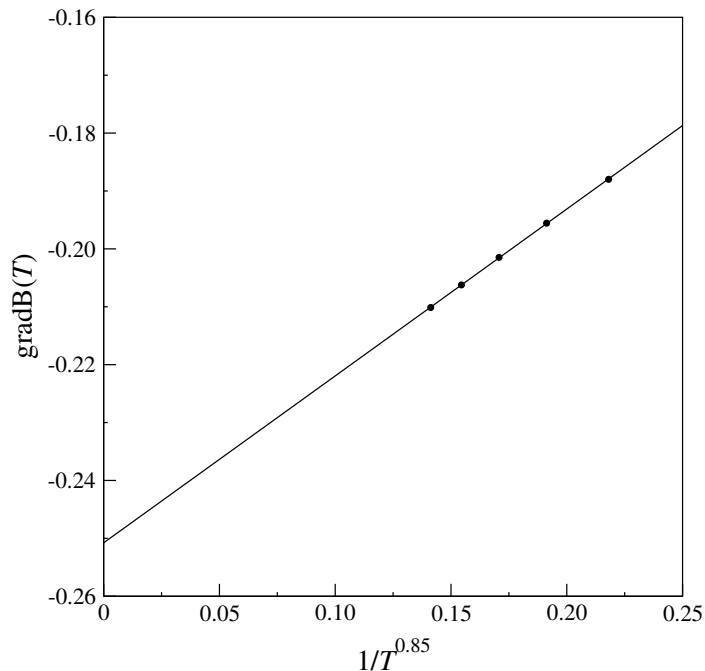
Figure 2.5: Plot of the last few points of the local gradient of $B_T(x_c)$ and T against $1/T^{0.85}$. The straight line is a linear fit to the data in the plot.

2.2.2 Square lattice

As discussed above, there is no parafermionic operator that applies to the SAW model on the square lattice or the triangular lattice, so we can't identify the critical point for SAWs on these lattices as did Duminil-Copin and Smirnov for honeycomb SAWs. We might, however, expect that in the limit $L \to \infty$ (so that we are again considering SAWs in a strip) there should be a similar relationship between the two generating functions $A_T(x)$ and $B_T(x)$, perhaps with some T-dependence that vanishes as $T \to \infty$.

That is to say, while the relationship

$$1 = \cos(3\pi/8)A_T(x_c) + B_T(x_c),$$

which is an identity for honeycomb lattice SAWs for finite width T, cannot be expected to hold for the square and triangular lattices, we might expect that solutions to

$$
\begin{aligned}
1 &= c_\alpha(T)A_T(x_c) + c_\beta(T)B_T(x_c) \\
1 &= c_\alpha(T)A_{T+1}(x_c) + c_\beta(T)B_{T+1}(x_c)
\end{aligned}
\tag{2.19}
$$

would display only a very weak T-dependence, and would quickly converge to constant limits.

We have computed data for the square lattice generating functions in strips of width T, that is, $A_T(x)$ and $B_T(x)$, for $T \leq 15$, and used our best estimate $1/x_c = 2.63815853031$ [77, 72] to tabulate $A_T(x_c)$ and $B_T(x_c)$, shown in Table 2.1. In [34] these generating functions for the

Table 2.1: Estimates of $A_T(x_\mathrm{c})$ and $B_T(x_\mathrm{c})$ for the square lattice.

T	$A_T(x_\mathrm{c})$	$B_T(x_\mathrm{c})$
1	0.684928096008073	0.760082094484555
2	0.825972541624066	0.707257323612670
3	0.927565166390104	0.668934606497192
4	1.006072923950508	0.639202723889591
5	1.069537792384553	0.615108345881821
6	1.122482001562161	0.594974760428940
7	1.167689112421950	0.577763265643123
8	1.206987841982332	0.562788338725227
9	1.241640411741764	0.549575210877016
10	1.272552495675558	0.537782341996967
11	1.300394615482380	0.527156358502502
12	1.325676196007041	0.517504450137522
13	1.348792763213512	0.508676719252903
14	1.370057142972426	0.500554481834765
15	1.389720731591218	0.493042273647721

honeycomb lattice were defined to include an extra half-step at the beginning of the walk and at the end of the walk. This introduces an extra factor of x (or, as appropriate, x_c) and we have used this definition of the generating functions $A_T(x)$ and $B_T(x)$ for the square lattice data.

We then fitted successive pairs of values $(A_T(x_\mathrm{c}), B_T(x_\mathrm{c}))$ and $(A_{T+1}(x_\mathrm{c}), B_{T+1}(x_\mathrm{c}))$ for $T = 1 \ldots 14$ to (2.19) and solved the associated linear equations for $c_\alpha(T)$ and $c_\beta(T)$, using our best estimate of x_c. In Figures 2.6 and 2.7 we show plots of values of $c_\alpha(T)$ against $1/T^{1.15}$ and $c_\beta(T)$ against $1/T^{0.85}$.

We have no basis for assuming that this is the correct form we should choose to extrapolate these plots; rather, the T-dependence was chosen experimentally to give a linear plot. Extrapolated to $T = \infty$, we find $c_\alpha \approx 0.3734$ and $c_\beta \approx 0.9756$. To obtain more precise estimates, we extrapolated these sequences using the Bulirsch-Stoer algorithm [21]. This algorithm requires a parameter w which can be thought of as a correction-to-scaling exponent. For the purpose of the current exercise, we have set this parameter to 1, corresponding to a T^{-2} correction term; this was chosen to match the observed behaviour of the generating functions (see below for further details). Our implementation of the algorithm is precisely as described by Monroe [88], and we retained 40 digit precision throughout. We also applied a range of standard extrapolation algorithms to the sequences $\{c_\alpha(T)\}$ and $\{c_\beta(T)\}$. These were Levin's u-transform, Brezinskii's θ algorithm, Neville tables, Wynn's ϵ algorithm and the Barber-Hamer algorithm. Descriptions
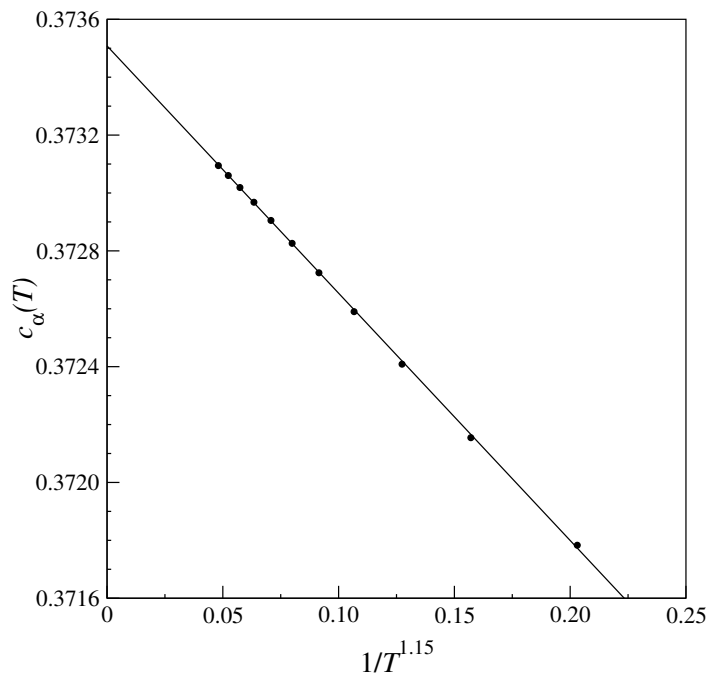
Figure 2.6: Plot of $c_\alpha(T)$ against $1/T^{1.15}$ for square lattice A walks. The straight line is a linear fit to the last seven data-points in the plot.

of these algorithms, and codes for their implementation, can be found in [52]. These gave results totally consistent with, but less precise than, those from the Bulirsch-Stoer algorithm.

In this way we estimated $c_\alpha = 0.373362 \pm 0.000001$ and $c_\beta = 0.975644 \pm 0.000002$. Thus the ratio $c_\alpha/c_\beta = 0.382683(2)$. For the honeycomb lattice the corresponding ratio is $\cos(3\pi/8) = 0.3826834\ldots$, which is close to, and probably equal to, the square lattice value. We shall see in the next section that this apparent agreement also holds for the triangular lattice.

Assuming that $c_\alpha/c_\beta = \cos(3\pi/8)$ for the square lattice, we calculated elements of the sequence $\cos(3\pi/8)A_T(x_c) + B_T(x_c)$ and extrapolated these using the same method as described above. We found the limit of the sequence to be 1.024966 ± 0.000001, compared to a value of exactly 1 for the honeycomb lattice. Using this estimate of the limit, we plotted

$$\log(\cos(3\pi/8)A_T(x_c) + B_T(x_c) - 1.024966)$$

against $\log T$. The plot displayed slight curvature, so we plotted the local gradient,

$$\log\left(\frac{\cos(3\pi/8)A_T(x_c) + B_T(x_c) - 1.024966}{\cos(3\pi/8)A_{T-1}(x_c) + B_{T-1}(x_c) - 1.024966}\right) \Big/ \log\left(\frac{T}{T-1}\right)$$

against $1/T$. This extrapolated to a value in the range $(1.9, 2.1)$, so we took the central value and concluded that $1.024966 - \frac{c_1}{T^2} \approx \cos(3\pi/8)A_T(x_c) + B_T(x_c)$ is the asymptotic behaviour. Finally, extrapolating estimates of the constant c_1, we estimate $c_1 \approx 0.14 \pm 0.02$. So our final result is

$$1.024966 - \frac{0.14}{T^2} \approx \cos(3\pi/8)A_T(x_c) + B_T(x_c),$$

32
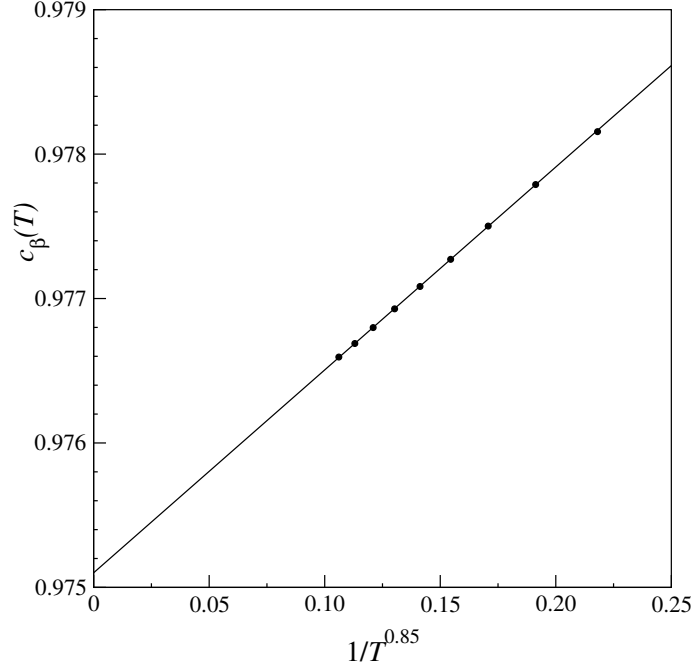
Figure 2.7: Plot of $c_\beta(T)$ against $1/T^{0.85}$ for square lattice B walks. The straight line is a linear fit to the last seven data-points in the plot.

which is an accurate mnemonic for square lattice strips.

It has been proven (see Appendix A) that on the honeycomb lattice, $B_T(x_c) \to 0$ as $T \to \infty$. This strongly suggests that the same behaviour should be observed on the square and triangular lattices. If true, this implies $1 \approx 0.3733621 A(x_c)$, giving a prediction for the critical amplitude $A(x_c) \approx 2.678365$. Current series estimates (unpublished[4]) are 2.66 ± 0.03, some 4 orders of magnitude less accurate than this new estimate.

For the honeycomb lattice the intersection point of $\cos(3\pi/8)A_T(x) + B_T(x)$ for any two distinct values of T uniquely determines x_c. For the square and triangular lattices, we instead looked at the intersection point of $\cos(3\pi/8)A_T(x) + B_T(x)$ and $\cos(3\pi/8)A_{T+1}(x) + B_{T+1}(x)$. Call this intersection point $x^*(T)$. Then one might expect $\lim_{T \to \infty} x^*(T) = x_c$. We calculated these points (see Table 2.2), and extrapolated the sequence $\{x^*(T)\}$ using the same Bulirsch-Stoer method described above, and in this way we estimated

$$x_c = 0.3790522775 \pm 0.0000000005.$$

This estimate can be compared to the best series estimates, based on analysis of very long polygon series, of $x_c = 0.37905227776$ [77, 72], with uncertainty in the last digit. Thus this method is seen to be a powerful new method for estimating critical points, giving very good accuracy, though it doesn't rival the most powerful methods based on series analysis of polygon series.

---

[4]Thanks to Tony Guttmann and Iwan Jensen for supplying this.

Table 2.2: Estimates of $x^*(T)$ and $\cos(3\pi/8)A_T(x^*(T)) + B_T(x^*(T))$ for the square lattice.

T	$x^*(T)$	$\cos(3\pi/8)A_T(x^*(T)) + B_T(x^*(T))$
1	0.3788492524430118	1.020519674456135
2	0.3789563888425186	1.022008494651236
3	0.3789993272206387	1.022854223759819
4	0.3790200211228027	1.023383853320655
5	0.3790312102673680	1.023737120932237
6	0.3790377790369609	1.023984250161982
7	0.3790418841341524	1.024163781522133
8	0.3790445790762211	1.024298263768565
9	0.3790464201463068	1.024401585674300
10	0.3790477199337992	1.024482674335293
11	0.3790486632778957	1.024547477105570
12	0.3790493642305414	1.024600079027025
13	0.3790498957528298	1.024643360731480
14	0.3790503059898279	1.024679400527091

However it does give comparable accuracy to methods based on series analysis of SAWs (rather than SAPs).

T-dependence of the generating functions $A_T(x_c)$ and $B_T(x_c)$.

As for the honeycomb lattice, it is expected that $B_T(x_c) \sim \text{const.}/T^{1/4}$. In Figure 2.8 we have plotted estimates of the exponent

$$\text{gradB}(T) = \log\left(\frac{B_T(x_c)}{B_{T-1}(x_c)}\right) \bigg/ \log\left(\frac{T}{T-1}\right)$$

against $1/T^{0.85}$. This local gradient should approach $-1/4$ and from the plot is seen to do so.

As for the honeycomb lattice, from (2.12) it follows that if $B_T(x_c) \sim \text{const.}/T^{1/4}$, then $\tilde{A}_T(x_c) = A_T(x_c) - a_0$ also decays as $T^{-1/4}$. This was observed numerically by a similar plot to that described in the preceding paragraph.

Alternative estimate of the critical point

For the square lattice, we expect the simultaneous solution of the pair of equations

$$\lambda A_{T-1}(x) + B_{T-1}(x) - \lambda A_T(x) - B_T(x) = 0,$$

and

$$\lambda A_T(x) + B_T(x) - \lambda A_{T+1}(x) - B_{T+1}(x) = 0,$$

Figure 2.8: Plot of gradB(T) against $1/T^{0.85}$ for square lattice B walks. The straight line is a linear fit to the last seven data-points in the plot. Linear extrapolation to -0.25 is well supported.

to give a sequence of estimates $x^{\dagger}(T)$ that should converge to the critical point x_c. Similarly, the parameter λ should converge to the ratio $c_{\alpha}/c_{\beta} = \cos(3\pi/8)$. The merit of this method of estimating the critical point is that it makes no assumption about the value of λ, while the other method assumes that $\lambda = \cos(3\pi/8)$.

We solved these equations by seeking the solution of

$$(A_{T-1}(x) - A_T(x))(B_{T+1}(x) - B_T(x)) = (A_T(x) - A_{T+1}(x))(B_T(x) - B_{T-1}(x)),$$

which we call $x^{\dagger}(T)$, and then found $\lambda(T)$ by back substitution. The results are shown in Table 2.3.

We plotted (not shown) the estimates of $x^{\dagger}(T)$, against various powers of $1/T$, and found a linear plot if we plotted against $1/T^2$. We extrapolated the estimates $x^{\dagger}(T)$ for steadily increasing T values using the same Bulirsch-Stoer extrapolation method described above. Rapid convergence was observed, and we estimate $x_c = 0.37905228 \pm 0.00000001$. This is consistent with the limit found from our previous method described above, though not quite as precise.

We have similarly extrapolated the estimates of $\lambda(T)$, and find $\lambda \approx 0.38268$, compared to the expected value $\cos(3\pi/8) = 0.382682$.

Table 2.3: Estimates of $x^\dagger(T)$ and $\lambda(T)$ for the square lattice.

T	$x^\dagger(T)$	$\lambda(T)$
2	0.3792132510996564	0.3680865016160631
3	0.3791354275802486	0.3724957043646600
4	0.3791014587212902	0.3750327779790171
5	0.3790837649841775	0.3766921607963391
6	0.3790736177161640	0.3778473665510655
7	0.3790673896037665	0.3786868392606266
8	0.3790633602596354	0.3793176229093515
9	0.3790606406190476	0.3798046222881576
10	0.3790587398862656	0.3801891415440993
11	0.3790573721782793	0.3804985249279112
12	0.3790563633515162	0.3807514859621669
13	0.3790556032334455	0.3809611990869139
14	0.37905501982268	0.3811371681858631

2.2.3 Triangular lattice

We have also generated data for the triangular lattice in strips of widths up to and including 11. Using the best estimate [75] of the critical point, $x_c = 0.2409175745$, we show, in Table 2.4, the values of $A_T(x_c)$ and $B_T(x_c)$ for each strip width. For the triangular lattice there are two edges incident upon the origin in a strip geometry, and this complicates matters. (See Figure 2.9 for an illustration.) To simplify things, we start and finish our SAWs *on* the boundary in the case of the triangular lattice, in order to avoid the complications that arise when including an incident edge. So the extra factor of x_c included in the definition of these amplitudes for the square and honeycomb lattice data is not present in the triangular lattice data.

As in the analysis of the square lattice data, we fitted successive pairs of values $(A_T(x_c), B_T(x_c))$ and $(A_{T+1}(x_c), B_{T+1}(x_c))$ for $T = 1\dots10$ to

$$1 = c_\alpha(T)A_T(x_c) + c_\beta(T)B_T(x_c)$$
$$1 = c_\alpha(T)A_{T+1}(x_c) + c_\beta(T)B_{T+1}(x_c),$$

and solved the associated linear equations for $c_\alpha(T)$ and $c_\beta(T)$.

To obtain precise estimates, we again applied the Bulirsch-Stoer extrapolation algorithm to the sequences $\{c_\alpha(T)\}$ and $\{c_\beta(T)\}$. Combining the results from these different algorithms, we estimate $c_\alpha = 0.2012028(3)$ and $c_\beta = 0.525770(3)$. Thus the ratio $c_\alpha/c_\beta = 0.382682(3)$. For the honeycomb lattice the corresponding ratio is $\cos(3\pi/8) = 0.3826834\dots$, which (as we also saw for the square lattice) is close to, and probably equal to, the triangular lattice value.
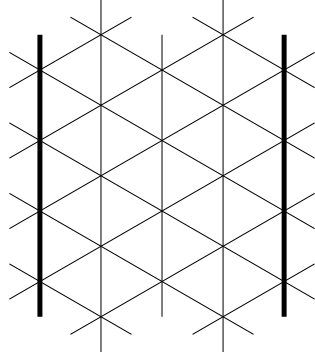
Figure 2.9: A section of a strip of the triangular lattice. Note that vertices on the boundaries of the strip are incident to two different external mid-edges. For this reason, for the triangular lattice calculations we consider walks to start and end at *vertices* rather that *mid-edges*.

Table 2.4: Estimates of $A_T(x_c)$ and $B_T(x_c)$ for the triangular lattice.

T	$A_T(x_c)$	$B_T(x_c)$
1	1.139480549210468	1.457161363236105
2	1.435344242350752	1.348134252648887
3	1.641756149326264	1.270897362392145
4	1.798515045521241	1.211810836367619
5	1.923848231267622	1.164374555192450
6	2.027608945103857	1.125001488941636
7	2.115709764900265	1.091512525007183
8	2.191966367371986	1.062490013670246
9	2.258977760090717	1.036962918106255
10	2.318589791981952	1.014238779515961
11	2.372157936598986	0.993807536013206

Table 2.5: Estimates of $x^*(T)$ and $\cos(3\pi/8)A_T(x^*(T)) + B_T(x^*(T))$ for the triangular lattice.

T	$x^*(T)$	$\cos(3\pi/8)A_T(x^*(T)) + B_T(x^*(T))$
1	0.2406280859767665	1.887740698623511
2	0.2407992581986084	1.893455041773400
3	0.2408577458655671	1.896277560401485
4	0.2408831660425418	1.897888855534072
5	0.2408959926648450	1.898898663967309
6	0.2409031617790197	1.899574374861755
7	0.2409074803284284	1.900049163002255
8	0.2409102353990879	1.900395685815981
9	0.2409120748141937	1.900656411828138
10	0.2409133491646394	1.900857546720268

Assuming the ratio $c_\alpha/c_\beta = \cos(3\pi/8)$ for the triangular lattice too, we extrapolated $\cos(3\pi/8)A_T(x_c) + B_T(x_c)$ for increasing values of T, using our standard suite of extrapolation algorithms and the Bulirsch-Stoer algorithm. We estimated the limit to be 1.901979 ± 0.000001. We then repeated the analysis described above for the square lattice data *mutatis mutandis*, and found

$$1.901979 - \frac{0.1}{T^2} \approx \cos(3\pi/8)A_T(x_c) + B_T(x_c).$$

As remarked above, it is expected that $\lim_{T\to\infty} B_T(x_c) = B(x_c) = 0$. Thus in the limit of infinite strip width we expect $1.901979 \approx \cos(3\pi/8)A(x_c)$, a prediction for the critical amplitude $A(x_c) \approx 4.970111$.

As for the square lattice case, we estimated the critical point x_c by extrapolating the intersection point of $\cos(3\pi/8)A_T(x) + B_T(x)$ and $\cos(3\pi/8)A_{T+1}(x) + B_{T+1}(x)$, called $x^*(T)$ (see Table 2.5). One expects $\lim_{T\to\infty} x^*(T) = x_c$.

In this way we estimated

$$x_c = 0.240917572 \pm 0.000000005$$

for the triangular lattice. This can be compared to the best series estimate, based on analysis of very long polygon series $x_c = 0.2409175745$ [75] with uncertainty in the last quoted digit.

T-dependence of the generating functions $A_T(x_c)$ and $B_T(x_c)$.

As for the honeycomb and square lattices, we expect $B_T(x_c) \sim \text{const.}/T^{1/4}$. We plotted estimates of the exponent

$$\text{gradB}(T) = \log\left(\frac{B_T(x_c)}{B_{T-1}(x_c)}\right) / \log\left(\frac{T}{T-1}\right)$$

against $1/T^{0.85}$, which should approach $-1/4$, and were seen to do so. The figure was visually indistinguishable from the corresponding Figure 2.8 for the square lattice, so is not shown.

Similarly, it follows from (2.12) that $\tilde{A}_T(x_c) = A_T(x_c) - a_0 \approx A_T(x_c) - 4.97011$ decays as $1/T^{1/4}$ as T tends to infinity. As we did for the square lattice case, we also confirmed this numerically.

Alternative estimate of the critical point

In the previous section, we showed that, for the square lattice data, the simultaneous solution of the pair of equations

$$\lambda A_{T-1}(x) + B_{T-1}(x) - \lambda A_T(x) - B_T(x) = 0,$$

and

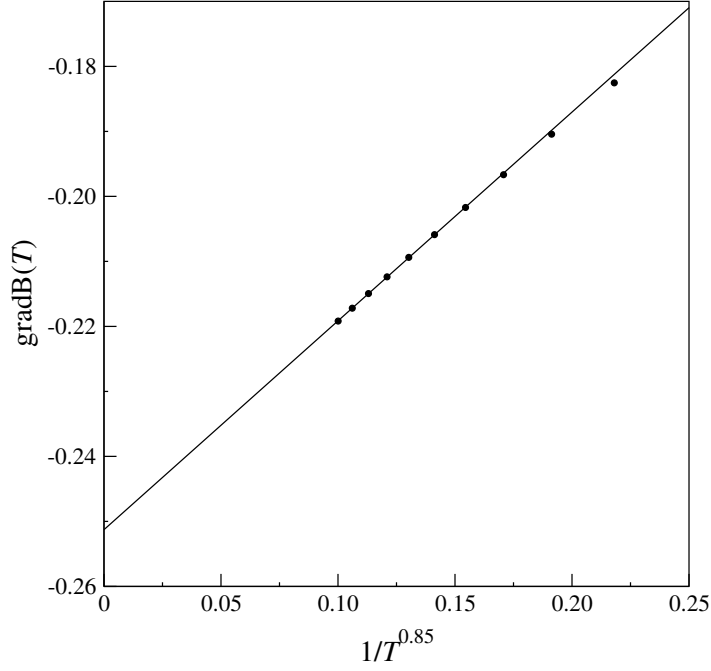$$\lambda A_T(x) + B_T(x) - \lambda A_{T+1}(x) - B_{T+1}(x) = 0,$$

gives a sequence of estimates $x^\dagger(T)$ that is expected to converge to the critical point x_c. Similarly, the parameter λ converges to the ratio $c_\alpha/c_\beta = \cos(3\pi/8)$, where the equality is conjectural. We solved these equations using the triangular lattice data, by seeking the solution of

$$(A_{T-1}(x) - A_T(x))(B_{T+1}(x) - B_T(x)) = (A_T(x) - A_{T+1}(x))(B_T(x) - B_{T-1}(x)),$$

called $x^\dagger(T)$, and then found $\lambda(T)$ by back substitution. The results are shown in Table 2.6.

We plotted (not shown) the estimates of $x^\dagger(T)$, against various powers of $1/T$, and found a linear plot if we plotted against $1/T^2$. We analysed the sequences in precisely the same way as for the corresponding square lattice data, using the Bulirsch-Stoer algorithm. For the critical point we estimate $x_c = 0.240917575 \pm 0.000000005$. This is of comparable precision to the other estimate given above, but slightly less precise than the best series estimate [75] of $x_c = 0.2409175745$, with uncertainty in the last digit.

Thus this method is again seen to be a powerful one for estimating critical points, giving very good accuracy. We have similarly extrapolated the estimates of $\lambda(T)$, and find $\lambda \approx 0.38268$, compared to the expected value $\cos(3\pi/8) = 0.382682$, exactly as for the square lattice.

Table 2.6: Estimates of $x^\dagger(T)$ and $\lambda(T)$ for the triangular lattice.

T	$x^\dagger(T)$	$\lambda(T)$
2	0.241168440165255	0.356318356471223
3	0.241030169141752	0.366143831978748
4	0.240977832351101	0.371147184391665
5	0.240953612190006	0.374091365359823
6	0.240940839933527	0.375992279128027
7	0.240933460889859	0.377300697288292
8	0.240928899289076	0.378244599187661
9	0.240925927855959	0.378950553554884
10	0.240923909640445	0.379493901730187

Chapter 3

Solvable subclasses

It has been clearly illustrated throughout Chapters 1 and 2 that, while many facts are known about the general models of self-avoiding walks and polygons, they are still a long way from being regarded as "solved". In order to say that such a combinatorial model has been solved, one would ideally like to have an explicit formula for c_n, the number of SAWs of length n. Failing this, an explicit expression for the generating function

$$Z(x) = \sum_{n \geq 0} c_n x^n$$

would suffice. Lacking even this, it would still be a huge step forward for one to prove the exact asymptotic form of c_n; for example, to prove that

$$c_n \sim An^{\alpha-1}\mu^n$$

as $n \to \infty$, with the constants A, α and μ exactly known. But in two and three dimensions, only the *existence* of the growth constant μ has been proved, and its value $\mu = \sqrt{2 + \sqrt{2}}$ for the honeycomb lattice [34].

   While methods like those described in Chapter 2 offer hope that progress can be made towards "solving" general SAW and SAP models, many researchers have turned their focus towards simpler models for which solutions can be obtained. These simpler models are usually subclasses of SAWs or SAPs which are defined by placing restrictions on the way the objects can be constructed. Typically, the hope is that the solutions of these models can aid in the understanding of the general, more complex models. Beyond this goal, the consideration of such models has served as a powerful incentive to develop new counting methods based on generating functions [19, 26, 44, 109], including transfer matrix methods and what is known as the "kernel method", which will be discussed later in this chapter. Furthermore, these objects can be modified so as to be useful in modelling physical objects like polymers and vesicles, and the solutions of these models can display interesting critical behaviour. (In Chapter 4 a number of these models are considered in the context of polymer adsorption.)

In this chapter we consider a number of self-avoiding walk and polygon subclasses which are either solvable (in at least one of the senses described above) or are such that the sequence $\{a_n\}$, enumerating the objects of size n, can be generated recursively in polynomial time. (This is in contrast to general SAWs and SAPs, for which the fastest algorithms still take exponential time – see Appendix B for further details.)

In Section 3.1 we study some of the simplest such objects, namely *directed walks*. For the most part these models were solved decades ago, and their solutions are relatively simple. This simplicity lends itself well to extending these models for use in polymer models, like adsorption and collapse, and there is a large body of work in which these models are studied in depth.

In Section 3.2 we consider a newer and more general class of objects: *prudent* walks and polygons. Several of these models have been solved in the last few years by a number of researchers; we present some new results which, in some cases, have surprising and unusual properties. Some of the methods we develop for analysing these solutions use concepts which we believe have never been applied to this type of problem. In Chapter 4 we also adapt some prudent walk models to the polymer adsorption problem, and observe that they display interesting critical phenomena.

Further generalisations of these models are studied in Section 3.3, including *perimeter* and *quasi-prudent* walks and polygons. We present new solutions for several classes of these objects. However, despite sharing a number of features with prudent walks and polygons, these models are for the most part not exactly solvable (by currently-known methods), and so we are often restricted to writing down functional equations and recursive algorithms.

In Sections 3.4 and 3.5 we consider the ways in which some of the models described above can be adapted to lattices other than the square lattice, namely the triangular and honeycomb lattices. We define and solve several new subclasses of SAWs on these lattices, including what we believe to be the most numerous (in terms of the growth constant) solved models. We again find a number of cases where we are unable to derive exact solutions.

Classifying generating functions

Before commencing our study of solvable models, we will briefly review some commonly-used terminology regarding the complexity of generating functions. (See, for example, [44, Appendix B.4] for further details.)

The simplest generating functions considered in this thesis are *rational* functions; these, of course, can be written as the ratio of two polynomials in the power series variable. The next most general class of generating functions is *algebraic*; these are solutions to polynomial equations whose coefficients are polynomials in the series variable. That is, a function $f(x)$ is algebraic if there exists a two-variable polynomial $P(x, y)$ such that $P(x, f(x)) = 0$.

The natural generalisation of algebraic functions is *D-finite*;[1] these are solutions to linear

---

[1]Also known as *holonomic* functions.

ODEs whose coefficients are polynomials in the series variable. That is, a function $f(x)$ is D-finite if it is the solution to an equation of the form

$$c_n(x)\frac{d^n}{dx^n}f(x) + c_{n-1}(x)\frac{d^{n-1}}{dx^{n-1}}f(x) + \cdots + c_0(x)f(x) = 0, \qquad (3.1)$$

where the coefficients $c_i(x)$ are polynomials in $\mathbb{C}[x]$. It follows from (3.1) that the coefficients f_n of $f(x)$ will satisfy a recurrence of the form

$$\widehat{c}_k(n)f_{n+k} + \widehat{c}_{k-1}(n)f_{n+k-1} + \cdots + \widehat{c}_0(n)f_n = 0, \qquad n \geq n_0$$

for some $n_0, k \geq 0$, where the coefficients $\widehat{c}_i(n)$ are polynomials.

It is not terribly difficult (again, see [44] for further details) to verify that algebraic functions are in fact D-finite, so that we have the hierarchy

$$rational \subset algebraic \subset \text{D-finite}.$$

It also follows from (3.1) that a D-finite function can only have a finite number of singularities in the complex plane.

The above definitions can also be generalised to multivariate generating functions.[2]

3.1 Directed walks

In this section we define a self-avoiding walk to be *directed* if there is a direction on the lattice in which it never steps. (For example, a SAW on the square lattice which never takes a west step.) More specifically, if S is a subset of the directions on a lattice then a walk is *S-directed* if it only steps in directions in S. (So again, a SAW on the square lattice which never takes a west step would be {north, south, east}-directed, or NSE-directed for short.)

Keeping in line with terminology used by other authors, we say a walk is *fully directed* if, for every direction X that it steps in, it never steps in the opposite direction $-X$. (So a NE-directed walk is fully directed, but a NSE-directed walk which steps both north and south is not.) A directed walk on the square lattice is generally referred to as a *partially directed* walk (PDW), and we will also use this terminology. See Figure 3.1 for examples of fully directed and partially directed walks.

It is easy to see that fully directed walks on the square lattice are trivially self-avoiding; their solution is likewise very easily found:

Lemma 3.1. *The number f_n of fully directed walks on the square lattice of length n is*

$$f_n = \begin{cases} 1 & n = 0 \\ 4(2^n - 1) & n \geq 1. \end{cases}$$

---

[2]Rechnitzer [100] has proven that the *anisotropic generating function* of self-avoiding polygons – that is, the bivariate generating function with one variable x corresponding to horizontal edges and another variable y corresponding to vertical edges – is not D-finite.
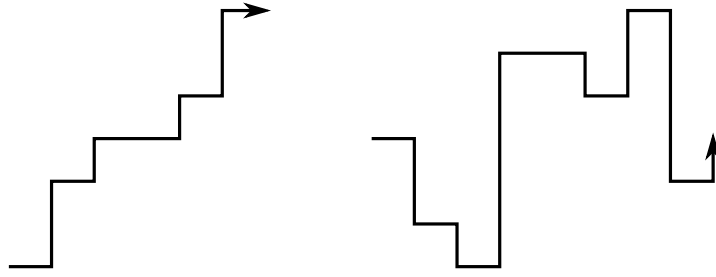
Figure 3.1: A fully directed and partially directed walk on the square lattice. The former is NE-directed and the latter is NSE-directed.

Proof. The number of NE-directed walks of length n is 2^n, since any ordered sequence of north and east steps forms a unique walk. A fully directed walk can be NE-, SE-, SW- or NW-directed, and so we multiply by 4, but need to account for walks which step in only one direction (which have been counted twice), and so finally subtract 4. This argument breaks down for the empty walk so it is treated separately. ∎

Corollary 3.2. *The mean squared end-to-end distance of fully directed walks of length $n \geq 1$ on the square lattice is*

$$\frac{n(2^n + n(2^n - 2))}{2(2^n - 1)} \sim \frac{n^2}{2} \qquad \textit{as } n \to \infty.$$

Proof. The number of NE-directed walks with k east steps is $\binom{n}{k}$; such a walk has squared end-to-end distance $k^2 + (n-k)^2$. Any fully directed walk can be uniquely obtained by rotating a NE-directed walk (excluding the E-directed walk) through $0, \pi/2, \pi$ or $3\pi/2$ radians; thus the mean squared end-to-end distance is

$$\sum_{k=0}^{n-1} 4(k^2 + (n-k)^2)\frac{\binom{n}{k}}{4(2^n - 1)} = \frac{n(2^n + n(2^n - 2))}{2(2^n - 1)}. \qquad ∎$$

It will be seen throughout the remainder of this chapter that results of the precision of Lemma 3.1 and Corollary 3.2 are frequently unobtainable; instead, we will be forced to settle for the generating functions of subclasses of SAWs and SAPs and the asymptotic form of their coefficients.

It will also be evident that the result of Corollary 3.2 is typical of solvable subclasses of SAWs – the mean squared end-to-end distances of almost all the models considered in this chapter are known or expected to be $O(n^2)$. This is disappointing, as for general SAWs this metric is widely believed to behave as $O(n^{3/2})$.[3]

---

[3]There exist solvable models whose mean squared end-to-end distance is not $O(n^2)$, namely *spiral* walks [15, 59] – see Subsection 3.3.3.

Lemma 3.3. *The generating function of NES-directed walks on the square lattice is*

$$Q(t) = \sum_{n \geq 0} q_n t^n = \frac{1+t}{1-2t-t^2}, \tag{3.2}$$

and so

$$q_n = \frac{(2+\sqrt{2})(1+\sqrt{2})^n - (2-\sqrt{2})(1-\sqrt{2})^n}{2\sqrt{2}} \sim \frac{2+\sqrt{2}}{2\sqrt{2}}(1+\sqrt{2})^n. \tag{3.3}$$

Proof. We use this simple example to illustrate the usefulness of generating functions when solving walk models. NES-directed walks can be divided into two groups: those which have taken an east step, and those which have not. Those with no east step are either the empty walk or contain only north or only south steps. Their generating function is thus

$$1 + \frac{2t}{1-t}. \tag{3.4}$$

Walks which do contain an east step must have a last such step; they are thus uniquely determined by what came before the last east step (which could be any NES-directed walk) and by what came after the east step (which could be nothing, north steps or south steps). Thus the generating function of these walks is

$$Q(t) \cdot t \left(1 + \frac{2t}{1-t}\right). \tag{3.5}$$

These two cases cover all NES-directed walks, thus by adding (3.4) and (3.5) we obtain $Q(t)$:

$$Q(t) = 1 + \frac{2t}{1-t} + t\left(1 + \frac{2t}{1-t}\right)Q(t). \tag{3.6}$$

Solving (3.6) gives $Q(t)$, and then the coefficients q_n can be obtained with partial fractions. ∎

Lemma 3.4. *The mean squared width (i.e. the distance in the x-direction between first and last points) of NES-directed walks of length n is asymptotically*

$$\frac{n^2}{4}.$$

Thus the mean squared end-to-end distance of NES-directed walks is $O(n^2)$.

Proof. We again use generating functions, but this time insert a second variable u which tracks the number of east steps taken by a walk. Then if $Q(t; u)$ is the new generating function,

$$Q(t; u) = \sum_{n,k \geq 0} q_{n,k} t^n u^k,$$

with $q_{n,k}$ being the number of n-step NES-directed walks with k east steps. Note that $Q(t, 1) = Q(t)$ as defined in Lemma 3.3. The same methodology as for Lemma 3.3 yields

$$Q(t; u) = 1 + \frac{2t}{1-t} + tu\left(1 + \frac{2t}{1-t}\right)Q(t; u), \tag{3.7}$$

45

and we thus get

$$Q(t;u) = \frac{1+t}{1-t-tu-t^2u}. \tag{3.8}$$

Now

$$\tilde{Q}(t) = \sum_{n,k \geq 0} q_{n,k} k^2 t^n = \frac{\partial}{\partial u}\left(u\frac{\partial Q(t;u)}{\partial u}\right)\Bigg|_{u=1}$$
$$= \frac{t(1+t)^2(1+t^2)}{(1-2t-t^2)^3} \tag{3.9}$$

is the generating function of NES-directed walks with each walk weighted by the square of its width. Thus the mean squared width of walks of length n is given by

$$\frac{[t^n]\tilde{Q}(t)}{[t^n]Q(t)}.$$

The asymptotic form of $[t^n]Q(t)$ is given in Lemma 3.3, and the asymptotic form of $[t^n]\tilde{Q}(t)$ can likewise be calculated with the basic techniques of analytic combinatorics applied to (3.9). After simplification the ratio is as it appears in the lemma.

The mean squared end-to-end distance is bounded below by the mean squared width and bounded above by n^2, and thus must be $O(n^2)$ (and is in fact $\Theta(n^2)$). $\blacksquare$

3.2 Prudent walks and polygons on the square lattice

First introduced by Préa [96], *prudent walks* are a subset of SAWs which never take a step towards any lattice vertex they have already visited. Apart from being intuitive and easily understood, this definition allows for prudent walks to be recursively generated in a systematic (and fast) manner, and these recursive methods have enabled researchers to compute a number of generating functions [31, 17].

The reason that we are able to recursively generate prudent walks quickly is that the endpoint of a prudent walk always lies on the boundary of the walk's *bounding box* – the smallest rectangle on the lattice which contains the entire walk.[4] So it is always possible to extend a prudent walk in two or three different directions, and it is easy to track exactly how many valid extensions are possible after each step.

The bounding box property also allows for a pleasing sub-classification of prudent walks:

- A *1-sided prudent walk* must, after every step, end on the east side of its current bounding box.

- A *2-sided prudent walk* must, after every step, end on the east or north side of its current bounding box. Equivalently, a 2-sided prudent walk is a prudent walk where a west step cannot be followed by a south step or vice versa.

---

[4]Vertices and edges of the walk may lie *on* the box, just not *outside* it.

Figure 3.2: Prudent walks on the square lattice: (a) 2-sided, (b) 3-sided and (c) 4-sided (unrestricted).

- A *3-sided prudent walk* must, after every step, end on the east, north or west side of its current bounding box, and in addition may not step from the east side to the west side of its box (or vice versa) when the box has width one, unless the endpoint is also on the north side. Equivalently, 3-sided prudent walks are prudent walks where a south step along the east side (and not also the west side) cannot be followed by a west step, and a south step along the west side (and not also the east side) cannot be followed by an east step.

- A *4-sided* or *unrestricted prudent walk* may end on any side of its bounding box.

See Figure 3.2 for examples of 2-, 3- and 4-sided prudent walks.

The decreasing levels of restriction applied to 1-, 2-, 3- and 4-sided prudent walks correspond to increasing levels of complexity of their solutions. Indeed, to date an exact solution to 4-sided prudent walks has eluded us completely. (As we will see later, we do have a functional equation for the generating function.) In the next subsection we will briefly discuss the known solutions and results for prudent walks.

The connection between prudent walks and *prudent polygons* is precisely the same as for SAWs and (rooted) SAPs – a prudent polygon is a prudent walk which ends at a vertex adjacent to its starting point. We define prudent polygons to be 1-sided, 2-sided, etc. by just applying the restrictions to the underlying prudent walks.

3.2.1 Prudent walks

1-sided prudent walks

Clearly 1-sided prudent walks are unable to take any west steps. Any self-avoiding walk comprised of north, east and south steps will be prudent and will always end on the east side of its box, so we have that 1-sided prudent walks are exactly NES-directed walks. Hence, their generating function and asymptotic behaviour is given in Lemma 3.3.

2-sided prudent walks

Duchi [31] obtained the solution to 2-sided prudent walks (see also [17]). The recursion she uses to generate walks is essentially an extension of the method used in the proof of Lemma 3.3. Instead of considering the last *east* step taken by a walk, she considers the last *inflating* step – the last step which resulted in either the east or north sides of the bounding box being moved. (In general, an inflating step for a k-sided prudent walk is a step which moves one of the k sides of interest.) The basic schema is:

- By symmetry (reflection in the line $y = x$), we can just count walks which end on the east side of their bounding box and double at the end. (Though we must account for those which end at the north-east corner of their box, which would be counted twice.)

- Walks ending on the east side of the box which take no inflating steps must be empty or just a sequence of south steps.

- Walks whose last inflating step was east can be decomposed into a walk which ended on the east side of its box, followed by the inflating east step, followed by either an unbounded number of south steps or a bounded number of north steps (the north steps cannot go past the top of the box).

- Walks whose last inflating step was north can be decomposed into a walk which ended on the north side of its box, followed by the inflating north step, followed by exactly enough east steps to take the endpoint to the north-east corner of the box.

So if we define $r_{n,k}^{(2)}$ as the number of 2-sided prudent walks of length n which end on the east side of their box with distance k between the endpoint and the north-east corner of the box (see Figure 3.3 for an illustration), we can consider the bivariate generating function

$$R^{(2)}(t; u) = \sum_{n,k \geq 0} r_{n,k}^{(2)} t^n u^k.$$

We will refer to u as a *catalytic variable*. From an enumerative point of view, we don't actually care about the measurement that u is tracking. We need u only because the recursion depends on knowing the distance between the endpoint of a walk and the north-east corner of its bounding box. In the end, after we solve $R^{(2)}(t; u)$, we will simply set $u = 1$; this is equivalent to just summing over all the values of k for each n.

Duchi's recursion can be encoded in a functional equation in $R^{(2)}(t; u)$:

Lemma 3.5 (Duchi [31], Bousquet-Mélou [17])**.** *The generating function $R^{(2)}(t; u)$ satisfies the functional equation*

$$R^{(2)}(t; u) = \frac{1}{1 - tu} + t R^{(2)}(t; t) + \frac{t^2 u}{1 - tu} R^{(2)}(t; u) + \frac{tu}{u - t} R^{(2)}(t; u) - \frac{t^2}{u - t} R^{(2)}(t; t). \quad (3.10)$$

Figure 3.3: A 2-sided prudent walk with the distance k measured by the catalytic variable u indicated, as defined by Duchi and Bousquet-Mélou.

Upon rearrangement (3.10) becomes

$$\left(1 - \frac{tu(1-t^2)}{(1-tu)(u-t)}\right) R^{(2)}(t;u) = \frac{1}{1-tu} + \frac{t(u-2t)}{u-t} R^{(2)}(t;t). \tag{3.11}$$

This equation can now be solved via the "kernel method", a technique for solving multivariate functional equations whose discovery is generally credited to Knuth [80]. (See also [5, 99].) In essence, the idea is to find a value for one of the variables (in our case, u) which will cancel the coefficient of one or more of the unknown terms in the equation. The coefficient being cancelled is called the *kernel*. Here the kernel will be

$$K(t;u) = 1 - \frac{tu(1-t^2)}{(1-tu)(u-t)}. \tag{3.12}$$

There are two solutions to $K(t;u) = 0$ in the variable u:

$$U_{\pm} = \frac{1 - t + t^2 + t^3 \pm \sqrt{(1-t^4)(1-2t-t^2)}}{2t}. \tag{3.13}$$

Only one of these roots is a power series in t, namely U_-. Substituting $u = U_-$ into (3.11) yields

$$R^{(2)}(t;t) = \frac{U_- - t}{t(1-tU_-)(2t-U_-)} \tag{3.14}$$

$$= 1 + 2t + 5t^2 + 12t^3 + 29t^4 + O(t^5).$$

This finally enables us to solve for the overall generating function $R^{(2)}(t;u)$:

Lemma 3.6 (Duchi [31], Bousquet-Mélou [17]). *The generating function for 2-sided prudent walks which end on the right side of their bounding box is*

$$R^{(2)}(t;u) = K(t;u)^{-1}\left(\frac{1}{1-tu} + \frac{t(u-2t)}{u-t} \cdot \frac{U_- - t}{t(1-tU_-)(2t-U_-)}\right) \tag{3.15}$$

where U_- is as defined in (3.13). The length generating function for 2-sided prudent walks is then

$$P^{(2)}(t) = 2R^{(2)}(t;1) - R^{(2)}(t;0) = \frac{1}{1 - 2t - 2t^2 + 2t^3} \left(1 + t - t^3 + t(1-t)\sqrt{\frac{1 - t^4}{1 - 2t - t^2}} \right)$$

$$= 1 + 4t + 10t^2 + 26t^3 + 66t^4 + O(t^5).$$

(3.16)

It is clear by inspection that the generating function $P^{(2)}(t)$ is algebraic. From this generating function, Bousquet-Mélou [17] has extracted the asymptotic behaviour of $p_n^{(2)}$, the number of 2-sided prudent walks of length n.

Corollary 3.7 (Bousquet-Mélou [17])**.** *The generating function of 2-sided prudent walks has a unique dominant singularity at $\rho \approx 0.403$, a root of $1 - 2\rho - 2\rho^2 + 2\rho^3$. The singularity is a simple pole, and so the number $p_n^{(2)}$ of 2-sided prudent SAWs of length n satisfies*

$$p_n^{(2)} \sim \varkappa \tau^n$$

with

$$\tau = \rho^{-1} \approx 2.48 \text{ and } \varkappa = \frac{\rho(3\rho - 1)}{(3\rho + 1)(5\rho - 2)} \approx 2.51.$$

Moreover, the mean sum $\langle S_n \rangle$ of the coordinates of the endpoints of a 2-sided prudent walk of length n is asymptotically $\langle S_n \rangle \sim Mn$, where $M = (\rho + 1)/(3\rho + 1) \approx 0.63$.

The asymptotic behaviour of $p_n^{(2)}$ is easily extracted from the generating function (3.16). Bousquet-Mélou obtains the result for the sum of the endpoint coordinates by adding a new variable to the generating function which tracks this measurement, similarly to the method demonstrated in Lemma 3.4.

3-sided prudent walks

Bousquet-Mélou [17] solves the generating function of 3-sided prudent walks in a similar manner to the 2-sided case. She defines two generating functions – one counts walks which end on the east side of their box, and the other counts those which end on the north side of their box. Each generating function makes use of two catalytic variables. Just as in the 2-sided case, walks are categorised according to the direction of their last inflating step, which could be east, north or west, and then recursively generated. These recursions lead to a pair of functional equations (encoding the method for generating walks which end on the east and north sides of their boxes respectively), which, after some manipulation, can be solved by a variation of the kernel method sometimes known as the *iterated kernel method*. (See also [116]. Later in this chapter and in Chapter 4 we will use the iterated kernel method to solve some related models, and will give a more detailed explanation.)

Lemma 3.8 (Bousquet-Mélou [17]). *The generating function of 3-sided prudent walks is given by*

$$P^{(3)}(t) = \sum_{n \geq 0} p_n^{(3)} t^n = \frac{1}{1 - 2t - t^2} \left(2t^2 q T(t) + \frac{(1+t)(2 - t - t^2 q)}{1 - tq} \right) - \frac{1}{1-t}, \qquad (3.17)$$

where

$$T(t) = \sum_{k \geq 0} (-1)^k \frac{\prod_{i=0}^{k-1} \left(\frac{t}{1-tq} - U(q^{i+1}) \right)}{\prod_{i=0}^{k} \left(\frac{tq}{q-t} - U(q^i) \right)} \left(1 + \frac{U(q^k) - t}{t(1 - tU(q^k))} + \frac{U(q^{k+1}) - t}{t(1 - tU(q^{k+1}))} \right),$$

$$U(w) = U(t; w) = \frac{1 - tw + t^2 + t^3 w - \sqrt{(1-t^2)(1 + t - tw + t^2 w)(1 - t - tw - t^2 w)}}{2t},$$

and $q = U(1) = U_-$ as per (3.13).

Similarly to Corollary 3.7, Bousquet-Mélou is able to apply the techniques of analytic combinatorics to determine the asymptotic behaviour of $p_n^{(3)}$ and extract information about the limiting shape of 3-sided prudent walks.

Corollary 3.9 (Bousquet-Mélou [17]). *The generating function of 3-sided prudent walks has a unique dominant singularity at $\rho \approx 0.403$, the same singularity as for 2-sided prudent walks. The singularity is again a simple pole. Thus the number of 3-sided prudent walks of length n satisfies*

$$p_n^{(3)} \sim \varkappa \tau^n$$

with $\tau = \rho^{-1} \approx 2.48$ and $\varkappa$ a positive constant.

The mean width $\langle W_n \rangle$ (i.e. width of the bounding box) of 3-sided prudent walks of length n is asymptotically $\langle W_n \rangle \sim M n$, where $M = (1 + \rho)/(2(1 + 3\rho)) \approx 0.31$.

Bousquet-Mélou also shows that the generating function $P^{(3)}(t)$ has an infinite number of singularities in the complex plane, and thus is not D-finite.

4-sided prudent walks

The progression we have just observed, from 1-sided prudent walks to 2- and then 3-sided, seems to suggest we should be able to follow the same methodology to obtain the solution for 4-sided prudent walks. But while Duchi [31] and Bousquet-Mélou [17] were able to obtain functional equations for this class of walks, nobody has yet been able to solve these equations and find the generating function.

We present here the functional equation as it is written by Bousquet-Mélou. By symmetry, we only need to count walks ending on the north side of their box. The last inflating step for these walks could be north, east or west.

Figure 3.4: A 4-sided prudent walk with the distances i, j, k measured by the catalytic variables u, v, w indicated, as defined by Bousquet-Mélou.

Lemma 3.10 (Bousquet-Mélou [17]). *Define the generating function*

$$T(t; u, v, w) = \sum_{n, i, j, k \geq 0} t_{n, i, j, k} t^n u^i v^j w^k,$$

where $t_{n, i, j, k}$ is the number of 4-sided prudent walks of length n which end on the north side of their bounding box, with distance i (resp. j) from the endpoint to the north-east (resp. north-west) corner of the box and with bounding box of height k. (See Figure 3.4 for an illustration.) Then $T(t; u, v, w)$ satisfies

$$\left(1 - \frac{t u v w(1 - t^2)}{(u - t v)(v - t u)}\right) T(t; u, v, w) = 1 + t u T(t; w, t w, u) + t v T(t; w, t w, v)$$

$$- \frac{t^2 v w}{u - t v} T(t; v, t v, w) - \frac{t^2 u w}{v - t u} T(t; u, t u, w). \tag{3.18}$$

The generating function for 4-sided prudent walks is then

$$P^{(4)}(t) = \sum_{n \geq 0} p_n^{(4)} t^n = 1 + 4 T(t; 1, 1, 1) - 4 T(t; 0, 1, 1). \tag{3.19}$$

The problem with solving this equation seems to be that it involves *three* catalytic variables. The mechanisms used for solving the 2- and 3-sided cases – the kernel method and its variant the iterated kernel method – can be used to solve functional equations involving generating functions with more than two catalytic variables, but usually only when there is a high degree of symmetry between the variables.[5] Here, it seems, things are simply too complicated. (This problem will come up again later, when we consider prudent polygons and some other related models.)

Dethridge and Guttmann [27] produced a long series (over 400 terms) of values for $p_n^{(4)}$ by iterating (3.18). They analysed this series and predicted that the dominant singularity is again a

---

[5]Bousquet-Mélou [18] has even used the kernel method to solve generating functions with an *arbitrary* number of catalytic variables, when considering the problem of permutations avoiding long ascending subsequences.
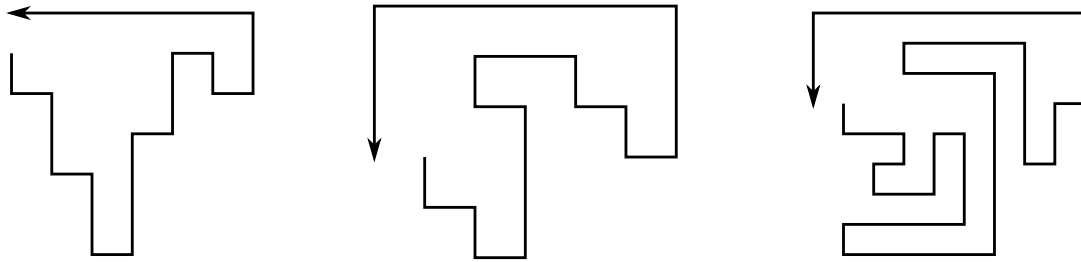
Figure 3.5: Examples of 2-sided (in this case, a bargraph), 3-sided and 4-sided prudent polygons.

simple pole at $\rho \approx 0.403$, and so conjecture that the asymptotic form of $p_n^{(4)}$ is

$$p_n^{(4)} \sim \varkappa \tau^n$$

with $\tau = \rho^{-1} \approx 2.48$ and $\varkappa \approx 16.12$. They also predicted that the generating function $P^{(4)}(t)$ is not D-finite.

3.2.2 Prudent polygons by perimeter

From prudent walks we now move on to prudent polygons. Recall that a k-sided prudent polygon is simply a k-sided prudent walk which ends at a point on the lattice adjacent to the starting point (i.e. the origin).[6] See Figure 3.5 for examples of 2-, 3- and 4-sided prudent polygons. In this subsection we summarise the work, primarily by Schwerdtfeger [105], on enumeration of prudent polygons by *perimeter*. In the next subsection we will consider enumeration by *area*, and find that 3-sided prudent polygons in particular display some interesting and unexpected behaviour. Since SAPs on the square lattice always have an even perimeter, we will follow Schwerdtfeger's lead and count them by *half-perimeter*.

1-sided prudent polygons

The only vertex that a 1-sided prudent polygon can end at is $(1,0)$. Thus, for any $n \geq 2$, there are only two 1-sided prudent polygons with half-perimeter n: the one which takes $n-1$ north steps followed by an east step and then $n-1$ south steps, and its reflection in the x-axis. The half-perimeter generating function of such objects is the rather uninteresting function

$$PP^{(1)}(t) = \frac{2t^2}{1-t}. \tag{3.20}$$

2-sided prudent polygons

A 2-sided prudent polygon can end at $(1,0)$ or $(0,1)$, and by reflective symmetry in the line $y = x$ the numbers of polygons ending at each vertex are equal. Thus it suffices to count the polygons ending at $(0,1)$ and multiply by two at the end. Such polygons can be divided into two classes:

---

[6]Of course we exclude a single step from being a polygon.

- If a polygon with half-perimeter n starts with a west step, it is forced to take $n-1$ west steps, a north step, then $n-1$ east steps. (We call these *degenerate* polygons.)

- If a polygon starts with a south or east step, it will continue as a NES-directed walk which stays *strictly below* the line $y = 1$, until it steps up to this line and must then step west to the end point.

The degenerate polygons have generating function $t^2/(1-t)$. Polygons in the second class are also known as *bargraphs*, and their generating function has long been known [97]. As demonstrated by Schwerdtfeger, they can be constructed recursively by adding columns to the rightmost side of existing bargraphs. A single catalytic variable is required – it keeps track of the height of the rightmost column of a polygon. The resulting functional equation can be solved by the kernel method, analogously to the technique used in Lemma 3.5.

Lemma 3.11 (Schwerdtfeger [105]). *The half-perimeter generating function of 2-sided prudent polygons is equal to*

$$PP^{(2)}(t) = \frac{1}{t}\left(\frac{1-3t+t^2+3t^3}{1-t} - \sqrt{(1-t)(1-3t-t^2-t^3)} \right). \tag{3.21}$$

The dominant singularity of $PP^{(2)}(t)$ is a square-root singularity at $t = \sigma \approx 0.296$, a root of $1-3\sigma-\sigma^2-\sigma^3 = 0$. Thus the number of 2-sided prudent polygons of half-perimeter n is asymptotically

$$pp_n^{(2)} \sim \varkappa \lambda^n n^{-3/2}, \tag{3.22}$$

where $\lambda = \sigma^{-1} \approx 3.38$ and $\varkappa \approx 0.855$.

3-sided prudent polygons

A 3-sided prudent polygon can end at $(-1,0),(0,1)$ or $(1,0)$. By symmetry (reflection in the y-axis), we only need to count those ending at $(-1,0)$ and those ending at $(0,1)$ in a counterclockwise direction. (By this we mean if we inserted an extra step to join the first and last vertices of the walk, the polygon will have been traversed in a counter-clockwise direction.) The latter are just the bargraphs discussed in the solution to 2-sided polygons. The polygons ending at $(-1,0)$ in a clockwise direction are degenerate, and consist only of a sequence of south steps, then a west step, then north steps.

The polygons ending at $(-1,0)$ in a counter-clockwise direction are the interesting class. Here, Schwerdtfeger first constructs all polygons which do not step above the line $y = 1$: these are bargraphs whose leftmost column always has height one. Then given a polygon we can construct a new one in two ways: by either adding a row to the top (whose left edge aligns with the left side of the existing polygon) which is shorter than the width of the current polygon, or by adding a row longer than the width of the exisiting polygon and possibly appending a

bargraph to the right edge of this row. This construction requires two catalytic variables, and so the resulting functional equation is solved in a similar way to 3-sided prudent walks.

Lemma 3.12 (Schwerdtfeger [105])**.** *The half-perimeter generating function of 3-sided prudent polygons ending at* $(-1, 0)$ *in a counter-clockwise direction is*

$$R(t) = \sum_{k \geq 0} L(t^k Q^{2k}) \prod_{j=0}^{k-1} K(t^j Q^{2j}), \tag{3.23}$$

where

$$Q \equiv Q(t) = \frac{1 + t^2 - \sqrt{(1-t)(1-3t-t^2-t^3)}}{2t},$$

$$L(w) \equiv L(t; w) = \frac{(1 + t^2 - (1 - 2t + 2t^2 + t^4)Q)(B(t; Qw) + t)w}{1 - t(1+t)Q - (t(1 - t - t^3)Q + t^2)(B(t; Qw) + t)w},$$

$$K(w) \equiv K(t; w) = \frac{(1-t)Q - 1 - ((1 - t + t^2)Q - 1)(B(t; Qw) + t)w}{1 - t(1+t)Q - (t(1 - t - t^3)Q + t^2)(B(t; Qw) + t)w},$$

and $B(t; u)$ *is the generating function of bargraphs with* t *conjugate to half-perimeter and* u *conjugate to width:*

$$B(t; u) = \frac{1 - t - tu(1+t) - \sqrt{t^2 u^2 (1-t)^2 - 2tu(1-t^2) + (1-t)^2}}{2tu}.$$

The half-perimeter generating function of 3-sided prudent polygons is then

$$PP^{(3)}(t) = 2\left(\frac{t^2}{1-t} + B(t; 1) + R(t)\right). \tag{3.24}$$

The generating function $PP^{(3)}(t)$ *is not D-finite. The dominant singularity of this generating function is a square root singularity at* $t = \eta \approx 0.244$, *a root of* $4 - 17\eta + 4\eta^2 - 6\eta^3 - \eta^5 = 0$. *So the number of 3-sided prudent polygons with half-perimeter n satisfies*

$$pp_n^{(3)} \sim \varkappa \omega^n n^{-3/2} \tag{3.25}$$

with $\omega = \eta^{-1} \approx 4.10$ *and* $\varkappa$ *a positive constant.*

4-sided prudent polygons

In a preprint of [105], Schwerdtfeger wrote down functional equations for 4-sided prudent polygons. Like the case of 4-sided prudent walks, the generating functions use three catalytic variables, and so at this stage no solution is known for this class. Here we modify Schwerdtfeger's definitions to make the constructions slightly more aesthetically pleasing, though ultimately the equations are no easier to solve.

Unrestricted prudent polygons can end at any of the four vertices adjacent to the origin, and can do so in either a clockwise or counter-clockwise direction. By symmetry then it suffices to count only those polygons ending at $(-1, 0)$ in a counter-clockwise direction. We partition these polygons into three subclasses $\mathcal{X}$, $\mathcal{Y}$ and $\mathcal{Z}$:

- Polygons in $\mathscr{X}$ are those for which removing the top row does not change the width of the polygon, result in two or more disconnected pieces, or remove the vertex $(-1,0)$ from the polygon. The generating function is

$$X(t;u,v,w) = \sum_{n,i,j,k \geq 0} x_{n,i,j,k} t^n u^i v^j w^k,$$

where $x_{n,i,j,k}$ is the number of polygons in $\mathscr{X}$ with half-perimeter n, top row of width i, total width $i+j$, and height k.

- Polygons in $\mathscr{Y}$ are the unit square plus those polygons not in $\mathscr{X}$ for which removing the rightmost column (which will necessarily be a single connected piece) does not change the height of the polygon nor result in two or more disconnected pieces. The generating function is

$$Y(t;u,v,w) = \sum_{n,i,j,k \geq 0} y_{n,i,j,k} t^n u^i v^j w^k,$$

where $y_{n,i,j,k}$ is the number of polygons in $\mathscr{Y}$ with half-perimeter n, rightmost column of height i, total height $i+j$, and width k.

- Polygons in $\mathscr{Z}$ are those not in $\mathscr{X}$ or $\mathscr{Y}$. The generating function is

$$Z(t;u,v,w) = \sum_{n,i,j,k \geq 0} z_{n,i,j,k} t^n u^i v^j w^k,$$

where $z_{n,i,j,k}$ is the number of polygons in $\mathscr{Z}$ with half-perimeter n, bottom row of width i, total width $i+j+1$, and height k.

Then the recursive construction works as follows (more details will be given in the proof of Proposition 3.18, when we consider the enumeration of 4-sided prudent polygons by area):

- A polygon in $\mathscr{X}$ is constructed by taking any polygon of width x and adding a row of width $\leq x$ to the top, with the leftmost edges aligned.

- A polygon in $\mathscr{Y}$ is constructed by:

  - Taking a polygon in $\mathscr{Y}$ or $\mathscr{Z}$ of height x and adding a new column of height $\leq x$ to the right, with the top edges aligned; or

  - Taking any polygon of width x and adding a row to the top of width $x+1$, with leftmost edges aligned.

- A polygon in $\mathscr{Z}$ is constructed by:

  - Taking a polygon in $\mathscr{Z}$ of width $x+1$ and adding a row of width $\leq x$ to the bottom, with the rightmost edges aligned; or

– Taking a polygon in $\mathcal{Y}$ or $\mathcal{Z}$ of height x and adding new column of height $x+1$ to the right, with the top edges aligned.

These constructions can then be encoded as funtional equations in X, Y and Z.

Lemma 3.13 (Schwerdtfeger [105]). *The generating functions $X(t; u, v, w), Y(t; u, v, w)$ and $Z(t; u, v, w)$ satisfy*

$$X(t; u, v, w) = \frac{tuw}{u-v}[X(t; u, v, w) - X(t; v, v, w)] + \frac{t^2 uw}{v-tu}[X(t; u, v, w) - X(t; u, tu, w)]$$
$$+ \frac{tuw}{u-v}[Y(t; w, w, u) - Y(t; w, w, v)] + \frac{tuw}{u-v}[uZ(t; u, u, w) - vZ(t; v, v, w)],$$

$$(3.26)$$

$$Y(t; u, v, w) = t^2 uw + \frac{tuw}{u-v}[Y(t; u, v, w) - Y(t; v, v, w)]$$
$$+ \frac{t^2 uw}{v-tu}[Y(t; u, v, w) - Y(t; u, tu, w)] + \frac{tuw^2}{u-v}[Z(t; w, w, u) - Z(t; w, w, v)] \qquad (3.27)$$
$$+ t^2 uwX(t; w, tw, v) + t^2 uwY(t; v, v, w) + t^2 uw^2 Z(t; w, w, v),$$

$$Z(t; u, v, w) = \frac{tuw}{u-v}[Z(t; u, v, w) - Z(t; v, v, w)] + \frac{t^2 uw}{v-tu}[Z(t; u, v, w) - Z(t; u, tu, w)]$$
$$+ \frac{t^2 uw}{v}Y(t; w, tw, v) + t^2 uwZ(t; v, v, w).$$

$$(3.28)$$

Note that in the end we can obtain the overall generating function by taking

$$PP^{(4)}(t) = 8[X(t; 1, 1, 1) + Y(t; 1, 1, 1) + Z(t; 1, 1, 1)]. \qquad (3.29)$$

Garoni et al. [47] have generated some 500 terms in the sequence $pp_n^{(4)}$, and based on a number of different numerical analyses, estimate that

$$pp_n^{(4)} \sim \varkappa \chi^n n^{\alpha-3}$$

where $\chi \approx 4.415$ and $\alpha \approx -1.5$. While these estimates are very rough, two facts do seem clear: firstly, the exponential growth rate changes from 2- to 3- to 4-sided prudent polygons, unlike prudent walks where it appears to stay at the same value. Secondly, the dominant singularity for 4-sided prudent polygons is almost certainly not a square root, unlike the 2- and 3-sided cases. It would be nice to have an intuitive argument as to why this is the case.

3.2.3 Prudent polygons by area: Generating functions

Our study of prudent walks and polygons on the square lattice finally turns to the enumeration of polygons by area. In this subsection, we construct functional equations for the four subclasses of prudent polygons, and in the cases of 1-, 2- and 3-sided polygons we solve these functional equations and write down their generating functions. For 4-sided prudent polygons we are unable to solve the functional equations.

In the next subsection we will focus on the generating function for 3-sided prudent polygons. It has a form which is reminiscent of, but ultimately quite different from, the generating functions of 3-sided prudent walks and 3-sided prudent polygons by perimeter. The analysis of its singularity structure is quite involved, and the asymptotic form of its coefficients is quite unlike any other solvable model considered in this thesis, or indeed any other polygon model known to the author.

The constructions we use here are essentially the same as Schwerdtfeger's [105]; the resulting functional equations and their solutions, however, turn out to be quite different. We will denote the area generating function for k-sided prudent polygons by

$$PA^{(k)}(q) = \sum_{n \geq 1} pa_n^{(k)} q^n,$$

where $pa_n^{(k)}$ is the number of k-sided prudent polygons of area n.

1-sided prudent polygons

As was explained in the last subsection, a 1-sided prudent polygon is just a single column of cells above or below the x-axis. There are then exactly two polygons of area n for each $n \geq 1$, so

$$PA^{(1)}(q) = \frac{2q}{1-q}. \tag{3.30}$$

2-sided prudent polygons

The non-trivial 2-sided prudent polygons can be constructed from bargraphs. Let $B(q) = \sum_{n \geq 1} b_n q^n$ be the area generating function for these objects. The area generating function for bargraphs, $B(q)$, is

$$B(q) = \frac{q}{1-2q}$$

and so $b_n = 2^{n-1}$ for $n \geq 1$.

Proposition 3.14. *The area generating function for 2-sided prudent polygons is*

$$PA^{(2)}(q) = \frac{2q}{1-2q} + \frac{2q}{1-q},$$

and so the number of such polygons is $pa_n^{(2)} = 2^n + 2$ for $n \geq 1$.

Proof. A 2-sided prudent polygon must end at either $(0, 1)$ or $(1, 0)$. Reflection in the line $y = x$ will not invalidate the 2-sided property, so it is sufficient to enumerate those polygons ending at $(1, 0)$ and then multiply the result by two.

The underlying 2-sided prudent walk cannot step above the line $y = 1$, nor to any point (x, y) where $x, y < 0$. So any polygon beginning with a west step must be a single row of cells to the left of the y-axis. The generating functions for these polygons is then $\frac{q}{1-q}$.

A polygon starting with a south or east step must remain on the east side of its box until it reaches the line $y = 1$, at which point it has no choice but to take west steps back to the y-axis. It is thus a bargraph, with generating function $B(q) = \frac{q}{1-2q}$.

Adding these two possibilities together and doubling gives the result. ∎

3-sided prudent polygons

When constructing 3-sided prudent polygons, we will use a single catalytic variable which measures width. To do so we will need to measure bargraphs by width. Let

$$B(q; u) = \sum_{n,i \geq 1} b_{n,i} q^n u^i$$

be the area-width generating function for bargraphs (so $b_{n,i}$ is the number of bargraphs with area n and width i).

The area-width generating function for bargraphs, $B(q; u)$, satisfies the equation

$$B(q; u) = \frac{qu}{1-q} + \frac{qu}{1-q} B(q; u), \tag{3.31}$$

which is obtained by successively adding columns. Accordingly, by solving the functional equation, we obtain

$$B(q; u) = \frac{qu}{1 - q - qu}$$

and so $b_{n,i} = \binom{n-1}{i-1}$ for $n, i \geq 1$. (Clearly, $b_{n,i}$ counts compositions of n into i summands.)

Let

$$W(q; u) = \sum_{n,i \geq 1} w_{n,i} q^n u^i$$

be the area-width generating function for 3-sided prudent polygons which end at $(-1, 0)$ in a counter-clockwise direction. As mentioned in the previous subsection, this is the most complex type of 3-sided prudent polygon; everything else is either a reflection of this or can be constructed from something simpler.

Lemma 3.15. *The area-width generating function for 3-sided prudent polygons ending at $(-1, 0)$ in a counter-clockwise direction, $W(q; u)$, satisfies the functional equation*

$$W(q; u) = qu(1 + B(q; u)) + \frac{q}{1-q}(W(q; u) - W(q; qu)) + qu(1 + B(q; u))W(q; qu). \tag{3.32}$$
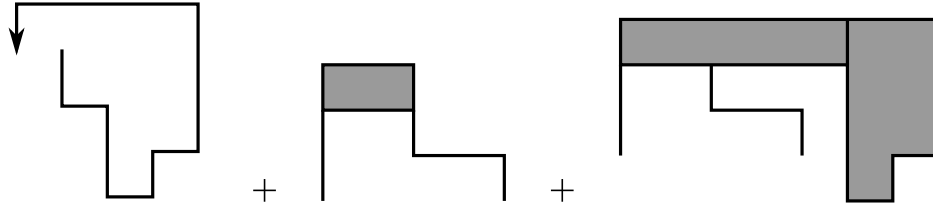
59

Figure 3.6: The decomposition used to construct 3-sided prudent polygons.

Proof. The underlying prudent walk cannot step to any point (x, y) with $x, y < 0$, nor to any point with $x < -1$. It must approach the final vertex $(-1, 0)$ from above. So the only time the endpoint can be on the west side of the box and not the north or south is when the walk is stepping south along the line $x = -1$. So prior to reaching the line $x = -1$, the walk must in fact be 2-sided. Note that the north-west corner of the box must be a part of of the polygon.

If the walk stays on or below the line $y = 1$, then (as has been seen in Proposition 3.14), it either reaches the point $(0, 1)$ with a single north step, or by forming a bargraph. This must then be followed by a west step to $(-1, 1)$, then a south step. This will form either a single square or a bargraph with a single square attached to the north-west corner, giving the first term on the right-hand side of (3.32). (These objects are represented by the left-hand side of Figure 3.6.)

Since the north-west corner of the box of any of these polygons is part of the polygon, it is valid to add a row of cells to the top of an existing polygon (so that the west sides line up). This can be done to any polygon. If the new row is not longer than the width of the existing polygon we obtain the term

$$\sum_{n \geq 1} \sum_{i \geq 1} w_{n,i} q^n u^i \cdot \sum_{k=1}^{i} q^k = q \sum_{n \geq 1} \sum_{i \geq 1} w_{n,i} q^n u^i \cdot \frac{1 - q^i}{1 - q},$$

giving the second term in the right-hand side of (3.32). (These objects are represented by the polygon in the centre of Figure 3.6.)

Note. For the remainder of this subsection, we will omit unwieldy double or triple sums like the one above, and instead give recursive relations only in terms of the generating functions.

Instead, the new row may be longer than the width of the existing polygon. In this case, as the walk steps east along this new row, it will reach a point at which there are no occupied vertices south of its position, and it will hence be able to step south in a prudent fashion. It must then remain on the east side of the box until reaching the north side, at which point it steps west to $x = -1$ and then south to the endpoint. This effectively means we have added a row of length equal to the width $+1$, and then (possibly) an arbitrary bargraph. So we obtain

$$q u W(q; qu)(1 + B(q; u))$$

which gives the final term in the right-hand side of (3.32). See the right side of Figure 3.6 for an illustration. ∎

Lemma 3.16. *The area-width generating function for 3-sided prudent polygons ending at $(-1, 0)$ in a counter-clockwise direction is*

$$W(q; u) = \sum_{m=0}^{\infty} F(q; q^m u) \prod_{k=0}^{m-1} G(q; q^k u),$$

where

$$F(q; u) = \frac{qu(1-q)^2}{(1-2q)(1-q-qu)}, \qquad G(q; u) = \frac{-q(1-q-u+qu-q^2u)}{(1-2q)(1-q-qu)}.$$

Proof. Substituting $B(q; u) = \frac{qu}{1-q-qu}$ into (3.32) and rearranging gives

$$W(q; u) = F(q; u) + G(q; u) W(q; qu). \tag{3.33}$$

Substituting $u \mapsto uq$ gives

$$W(q; qu) = F(q; qu) + G(q; qu) W(q; q^2 u) \tag{3.34}$$

and combining these yields

$$W(q; u) = F(q; u) + F(q; qu) G(q; u) + G(q; u) G(q; qu) W(q; q^2 u). \tag{3.35}$$

Repeating for $u \mapsto q^2 u, q^3 u, \ldots, q^M u$ will give

$$W(q; u) = \sum_{m=0}^{M} F(q; q^m u) \prod_{k=0}^{m-1} G(q; q^k u) + \prod_{m=0}^{M} G(q; q^m u) W(q; q^{M+1} u). \tag{3.36}$$

We now seek to take $M \to \infty$. To obtain the result stated in the lemma, it is necessary to show that

$$\sum_{m=0}^{M} F(q; q^m u) \prod_{k=0}^{m-1} G(q; q^k u)$$

converges, and

$$\prod_{m=0}^{M} G(q; q^m u) W(q; q^{M+1} u) \to 0$$

as $M \to \infty$ (both considered as power series in q and u).

Both F and G are bivariate power series in q and u. We have that

$$F(q; u) = qu + q^2(u + u^2) + q^3(2u + 2u^2 + u^3) + O(q^4)$$
$$G(q; u) = q(-1 + u) + q^2(-2 + u + u^2) + q^3(-4 + 2u + 2u^2 + u^3) + O(q^4)$$

It follows that $F(q; q^m u) = O(q^{m+1})$ and $G(q; q^k u) = O(q)$ for all $m, k \geq 0$. So then

$$F(q; q^m u) \prod_{k=0}^{m-1} G(q; q^k u) = O(q^{2m+1})$$

61

So considered as a power series in q and u, the first term in the right-hand side of (3.36) does converge to a fixed power series as $M \to \infty$.

By the same argument, we see that

$$\prod_{m=0}^{M} G(q;q^m u) \to 0$$

as $M \to \infty$. So it suffices to show that $W(q;q^{M+1}u)$ converges to a fixed power series. But now every term in the series $W(q;u)$ has at least one factor of u (since every polygon has positive width), so it immediately follows that $W(q;q^{M+1}u) \to 0$ as $M \to \infty$.

So both terms in (3.36) behave as required as $M \to \infty$, and the result follows. ∎

Theorem 3.17. *The area generating function for 3-sided prudent polygons is*

$$PA^{(3)}(q) = \frac{-2q^3(1-q)^2}{(1-2q)^2} \sum_{m=1}^{\infty} \frac{(-1)^m q^{2m}}{(1-2q)^m (1-q-q^{m+1})} \prod_{k=1}^{m-1} \frac{1-q-q^k+q^{k+1}-q^{k+2}}{1-q-q^{k+1}}$$
$$+ \frac{2q(3-10q+9q^2-q^3)}{(1-2q)^2(1-q)}$$
$$= 6q + 10q^2 + 20q^3 + 42q^4 + 92q^5 + 204q^6 + 454q^7 + 1010q^8 + \cdots$$

Proof. A 3-sided prudent polygon must end at $(-1,0),(0,1)$ or $(1,0)$, in either a clockwise or counter-clockwise direction. Setting $u = 1$ in $W(q,u)$ gives the area generating function

$$W(q;1) = \frac{-q^3(1-q)^2}{(1-2q)^2} \sum_{m=1}^{\infty} \frac{(-1)^m q^{2m}}{(1-2q)^m (1-q-q^{m+1})} \prod_{k=1}^{m-1} \frac{1-q-q^k+q^{k+1}-q^{k+2}}{1-q-q^{k+1}}$$
$$+ \frac{q(1-q)^2}{(1-2q)^2}. \quad (3.37)$$

A clockwise polygon ending at $(-1,0)$ can only be a single column, which has generating function

$$\frac{q}{1-q}. \quad (3.38)$$

A counter-clockwise polygon ending at $(0,1)$ cannot step left of the y-axis or above the line $y = 1$. While it is below this line, it must remain on the east side of its box, and upon reaching the line $y = 1$, it must step west to the y-axis. It must therefore be a bargraph, with generating function

$$\frac{q}{1-2q}. \quad (3.39)$$

A reflection in the y-axis converts a polygon ending at $(-1,0)$ to one ending at $(1,0)$ in the opposite direction, and reverses the direction of a polygon ending at $(0,1)$. So adding together and doubling (3.37), (3.38) and (3.39) will cover all possibilities, and gives the stated result. ∎

We will postpone the singularity analysis of $PA^{(3)}(q)$ until Subsection 3.2.4, as it is quite involved and makes use of a number of results and techniques not found elsewhere in this thesis.
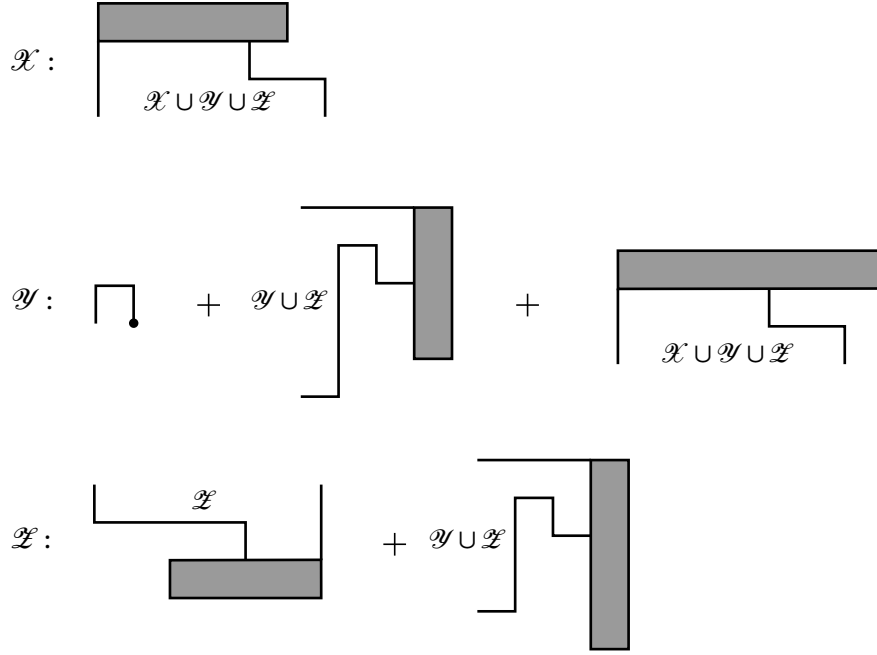
Figure 3.7: The decompositions used to construct 4-sided prudent polygons in $\mathcal{X}, \mathcal{Y}, \mathcal{Z}$ (from top to bottom). The shaded sections are the rows or columns being added, and the labels indicate which classes can accept such additions.

4-sided prudent polygons

As discussed in the last section, 4-sided prudent polygons have an 8-fold symmetry, and thus we only need to enumerate those ending at $(-1, 0)$ in a counter-clockwise direction. We partition these polygons in the same way as in the last subsection – into classes $\mathcal{X}, \mathcal{Y}$ and $\mathcal{Z}$ – but use different catalytic variables. Here, we have the generating functions

$$X(q; u, v) = \sum_{n, i, j \geq 1} x_{n, i, j} q^n u^i v^j$$

$$Y(q; u, v) = \sum_{n, i, j \geq 1} y_{n, i, j} q^n u^i v^j$$

$$Z(q; u, v) = \sum_{n, i, j \geq 1} z_{n, i, j} q^n u^i v^j$$

where

- $x_{n, i, j}$ is the number of polygons in $\mathcal{X}$ with area n, width i and height j,

- $y_{n, i, j}$ is the number of polygons in $\mathcal{Y}$ with area n, height i and width j, and

- $z_{n, i, j}$ is the number of polygons in $\mathcal{Z}$ with area n, width $i + 1$ and height j.

Proposition 3.18. *The generating functions $X(q;u,v)$, $Y(q;u,v)$ and $Z(q;u,v)$ satisfy the functional equations*

$$X(q;u,v) = \frac{qv}{1-q}[X(q;u,v) - X(q;qu,v)] + \frac{qv}{1-q}[Y(q;v,u) - Y(q;v,qu)]$$
$$+ \frac{quv}{1-q}[Z(q;u,v) - qZ(q;qu,v)], \tag{3.40}$$

$$Y(q;u,v) = quv + \frac{qv}{1-q}[Y(q;u,v) - Y(q;qu,v)] + \frac{qv^2}{1-q}[Z(q;v,u) - Z(q;v,qu)]$$
$$+ quv[X(q;qv,u) + Y(q;u,qv) + qvZ(q;qv,u)], \tag{3.41}$$

$$Z(q;u,v) = \frac{qv}{1-q}[Z(q;u,v) - Z(q;qu,v)] + qvY(q;qv,u) + quvZ(q;u,qv). \tag{3.42}$$

The generating function for 4-sided prudent polygons is then given by

$$PA^{(4)}(q) = 8[X(q;1,1) + Y(q;1,1) + Z(q;1,1)]$$
$$= 8q + 24q^2 + 80q^3 + 248q^4 + 736q^6 + 2120q^7 + 5960q^8 + 16464q^9 + \cdots$$

Proof. The construction is the same as for Lemma 3.13, but we explain it in more detail here. In Figure 3.7 we provide an illustration of the construction for the three different classes.

As with the 3-sided polygons in Lemma 3.15, the walk cannot visit any point (x,y) with $x,y < 0$ or with $x < -1$. The walk must approach $(-1,0)$ from above, and must do so immediately upon reaching the line $x = -1$. So every polygon contains the north-west corner of its box. As in the 3-sided case, this leads to a construction involving adding rows to the top of existing polygons.

By definition, a polygon in $\mathscr{X}$ of width i can be constructed by adding a row of length $\leq i$ to the top of any polygon of width i. Adding a row to a polygon in $\mathscr{X}$ gives

$$\sum_{n\geq 1}\sum_{i\geq 1}\sum_{j\geq 1} x_{n,i,j}q^n u^i v^j \cdot v \sum_{k=1}^{i} q^k = qv\sum_{n\geq 1}\sum_{i\geq 1}\sum_{j\geq 1} x_{n,i,j}q^n u^i v^j \cdot \frac{1-q^i}{1-q},$$

which is the first term in the right-hand side of (3.40). Performing similar operations for polygons in $\mathscr{Y}$ and $\mathscr{Z}$ gives the rest of (3.40).

Note. Again, for the remainder of this subsection we give recursive relations purely in terms of the generating functions.

Polygons not in $\mathscr{X}$ must also contain the north-east corner of their box. This leads to another construction involving adding columns to the right-hand side of existing polygons. To obtain a polygon in $\mathscr{Y}$ of height i, a new column of height $\leq i$ should be added to a polygon of height i which contains the north-east corner of its box. So adding a column to a $\mathscr{Y}$ polygon gives

$$\frac{qv}{1-q}Y(q;u,v) - \frac{qv}{1-q}Y(q;qu,v)$$

which is the second term in the right-hand side of (3.41). Performing a similar operation for $\mathscr{Z}$ polygons gives the third term in (3.41).

Adding a new column to a polygon in $\mathscr{X}$ containing its north-east corner can be viewed as adding a sequence of rows on top of one another, and so if the new column has height ≥ 2 then the resulting polygon is actually in $\mathscr{X}$. If the new column has height one, however, the resulting polygon is in $\mathscr{Y}$. Isolating those polygons in $\mathscr{X}$ which contain their north-east corner is difficult; however, we can perform an equivalent construction by adding a row of length $i + 1$ to any polygon of width i. Doing so to a polygon in $\mathscr{X}$ gives

$$quvX(q;qv,u)$$

and combining this with the same for $\mathscr{Y}$ and $\mathscr{Z}$ gives the fourth term in (3.41). The quv term is the unit square.

Polygons in $\mathscr{Z}$ also contain the south-east corner of their box. In a similar fashion to the constructions for $\mathscr{Y}$ and $\mathscr{Z}$, we can add a new row to the bottom of a polygon containing its south-east corner. To do so to a polygon in $\mathscr{Z}$ of width $i + 1$ (remember u measures width -1) requires a new row of width $\leq i$, so we obtain

$$\frac{qv}{1-q}Z(q;u,v) - \frac{qv}{1-q}Z(q;qu,v)$$

which is the first term in the right-hand side of (3.42).

Adding a new row to the bottom of something in $\mathscr{X}$ (containing its south-east corner) will give back something in $\mathscr{X}$, which will have been constructed by an alternate method described above. Adding a new row of length ≥ 2 to a polygon in $\mathscr{Y}$ will result in another polygon in $\mathscr{Y}$, which will also be constructible via alternate means. So we are left only with the possibility of adding a row of length one to the bottom of a polygon in $\mathscr{Y}$. This is analogous to the above description of adding a column of height one to the right of a polygon in $\mathscr{X}$; we now proceed by adding a column of height $i + 1$ to a polygon in $\mathscr{Y}$ or $\mathscr{Z}$ of height i. Doing so gives the final two terms in (3.42). ∎

We are unable to solve the functional equations for X, Y and Z, and so we turn to series analysis. Based on a series of some 800 terms, we strongly conjecture that the exponential growth rate for $pa_n^{(4)}$ is exactly 2, which matches that of 2- and 3-sided prudent polygons (see Subsection 3.2.4 for more details on the 3-sided case). With this assumption in hand, we then estimate that

$$pa_n^{(4)} \sim \alpha 2^n n^\beta$$

where $2.58 < \beta < 2.61$ and α is a positive constant. (In light of the results of the next section, we will eventually revise this estimate – see Table 3.2 and the preceding discussion.)

65

3.2.4   3-sided prudent polygons by area: Analysis and asymptotics

For most lattice object problems, finding and solving the functional equation(s) is the difficult part. Once a generating function has been found, the dominant singularity is often quite obvious, and so the asymptotic form of the coefficients can be easily described. The problem of 3-sided prudent polygons, however, turns out to be rather the opposite. The functional equation (3.32) was not terribly difficult to obtain, and its solution is relatively simple – it only comprises a sum of products of rational functions of q.

The asymptotic behaviour of this model, on the other hand, is considerably more complex than any model we have seen before. The dominant singularity at $q = 1/2$ is not even apparent from the representation of Theorem 3.17. As we shall see, there is in fact an accumulation of poles of the generating function[7] $PA(q)$ towards $q = 1/2$. Accordingly, the nature of the dominant singularity at $q = 1/2$ is rather unusual: a singular expansion as q approaches $1/2$ can be determined, but it involves periodic fluctuations, a strong divergence from the standard simple type $Z^\alpha (\log Z)^\beta$, where $Z := 1 - z/\rho$, with ρ (here equal to $1/2$) the dominant singularity of the generating function under consideration. This is revealed by a Mellin analysis of $PA(q)$ near its singularity, and the periodic fluctuations, which appear to be in a logarithmic scale, eventually echo the geometric speed with which poles accumulate at $1/2$. Then, thanks to a suitable extension to the complex plane, the singular expansion can be transfered to coefficients by the method known as singularity analysis [44, Ch. VI]. The net result is, for the coefficients pa_n, an asymptotic form that involves a standard element $2^n n^g$, but multiplied by a *periodic function* in $\log_2 n$. The presence of oscillations, the transcendental character of the exponent $g = \log_2 3$, and the minute amplitude of these oscillations, about 10^{-9}, are noteworthy features of this asymptotic problem.

Theorem 3.19. *The number $pa_n \equiv pa_n^{(3)}$ of 3-sided prudent polygons of area n satisfies the estimate*

$$pa_n = \left[x_0 + x(\log_2 n) \right] 2^n \cdot n^g + O\left(2^n \cdot n^{g-1} \log n \right), \qquad n \to \infty, \qquad (3.43)$$

where the critical exponent is

$$g = \log_2 3 = 1.58496\ldots$$

and the "principal" constant is

$$x_0 = \frac{\pi}{9 \log 2 \, \sin(\pi g) \Gamma(g+1)} \prod_{j=0}^{\infty} \frac{(1 - \frac{1}{3} 2^{-j})(1 - \frac{3}{2} 2^{-j})}{(1 - \frac{1}{2} 2^{-j})^2} = 0.1083842946\ldots. \qquad (3.44)$$

The function $x(u)$ is a smooth periodic function of u, with period 1, mean value zero, and amplitude approximately $1.54623 \cdot 10^{-9}$, which is determined by its Fourier series representation:

$$x(u) = \sum_{k \in \mathbb{Z}\setminus\{0\}} x_k e^{2ik\pi u}, \qquad with \quad x_k = x_0 \cdot \frac{\sin(\pi g)}{\sin(\pi g + 2ik\pi^2/\log 2)} \cdot \frac{\Gamma(1+g)}{\Gamma(1+g+2ik\pi/\log 2)}.$$

---

[7]Throughout this subsection only dedicated to 3-sided prudent polygons, we omit redundant superscripts and let pa_n and $PA(q)$ represent, respectively, what was denoted by $pa_n^{(3)}$ and $PA^{(3)}(q)$ in Subsection 3.2.3.

Table 3.1: Some of the recurring quantities of this subsection, their reduction at $q = 1/2$ and the relevant equations in the text.

Quantity	$q = 1/2$	Reference
$u = \dfrac{q}{1-q}$	1	(3.48)
$v = \dfrac{1-q+q^2}{1-q}$	$\dfrac{3}{2}$	(3.48)
$a = \dfrac{q^2}{1-q+q^2}$	$\dfrac{1}{3}$	(3.66)
$\gamma = \dfrac{\log v}{\log 1/q}$	$\log_2(3/2)$	(3.74)
$C(q) = \dfrac{2q(3-10q+9q^2-q^3)}{(1-q)(1-2q)^2}$	$\sim \dfrac{1}{4(1-2q)^2}$	(3.46) and (3.51)
$A(q) = \dfrac{2q(1-q)^2}{(1-2q)^2}$	$\sim \dfrac{1}{4(1-2q)^2}$	(3.46) and (3.50)

The proof of the theorem occupies the next subsections, whose organisation reflects the informal description given above. We shall then discuss the fine structure of subdominant terms in the asymptotic expansion of pa_n; cf. Theorem 3.29. Some quantities that appear repeatedly throughout this subsection are tabulated in Table 3.1 for convenience.

Resummations

We start with a minor reorganisation of the formula provided by Theorem 3.17: completion of the finite products that appear there leads to the equivalent *q–hypergeometric* form

$$PA(q) = C(q) + A(q) \cdot Q(1;q) \cdot \sum_{n=1}^{\infty} \frac{(-1)^n q^{2n}}{(1-2q)^n} \cdot \frac{1}{Q(q^n;q)}. \tag{3.45}$$

Here and throughout this subsection, the notations are

$$C(q) := \frac{2q(3-10q+9q^2-q^3)}{(1-q)(1-2q)^2}, \qquad A(q) := \frac{2q(1-q)^2}{(1-2q)^2}, \tag{3.46}$$

and

$$Q(z;q) := Q(z;q;u(q),v(q)), \qquad \text{where} \quad Q(z;q;u,v) = \frac{(vz;q)_\infty}{(quz;q)_\infty}, \tag{3.47}$$

with

$$u(q) = \frac{q}{1-q}, \qquad v(q) = \frac{1-q+q^2}{1-q}. \tag{3.48}$$

In the definition of Q, the notation $(x;q)_n$ represents the usual q-Pochhammer symbol:

$$(x;q)_n = (1-x)(1-qx)\cdots(1-xq^{n-1}).$$

Lemma 3.20. *The function PA(q) is analytic in the open disc $|q| < \sqrt{2} - 1$, where it admits the convergent q-hypergeometric representation*

$$PA(q) = C(q) + A(q) \frac{(v;q)_\infty}{(qu;q)_\infty} \sum_{n=1}^{\infty} (-1)^n \frac{q^{2n}}{(1-2q)^n} \frac{(uq^{n+1};q)_\infty}{(vq^n;q)_\infty}, \qquad (3.49)$$

with $A(q), C(q), u \equiv u(q), v \equiv v(q)$ rational functions given by (3.46) and (3.48).

Proof. The (easy) proof reduces to determining sufficient analyticity regions for the various components of the basic formula (3.45), some of the expansions being also of later use. First, the functions $A(q)$ and $C(q)$ are meromorphic for $|q| < 1$, with only a pole at $q = 1/2$. They can be expanded about the point $q = 1/2$ to give

$$A(q) = \frac{1}{4(1-2q)^2} + \frac{1}{4(1-2q)} - \frac{1}{4} - \frac{1-2q}{4}, \qquad (3.50)$$

$$C(q) = \frac{1}{4(1-2q)^2} + \frac{5}{4(1-2q)} + \frac{3}{4} - \frac{17(1-2q)}{4} + O((1-2q)^2). \qquad (3.51)$$

The function $Q(1;q)$ is analytic for $|q| < 1$ except at the points for which $(uq;q)_\infty = 0$, that is, the points σ for which $1 - \sigma - \sigma^n = 0$ for $n \geq 2$. The smallest of these (in modulus) is $\varphi = (\sqrt{5}-1)/2 = 0.618034...$, a root of $1 - q - q^2$. So $Q(1;q)$ is certainly analytic at $q = 1/2$; the constant term in its expansion about $q = 1/2$ is

$$Q(1;1/2) = \frac{(3/2;1/2)_\infty}{(1/2;1/2)_\infty} = -0.18109782...$$

In similar fashion, $1/Q(z;q)$ is *bivariate analytic* at points (z,q) for which $|q| < 1$, except when $(vz;q)_\infty = 0$. This occurs at points (z_j, q) where $z_j := \frac{1}{vq^j}$, for $j \geq 0$. In particular, for $|q| < \theta$, where[8]

$$\theta = 0.56984... := \text{the unique real root of } 1 - 2x + x^2 - x^3, \qquad (3.52)$$

we have $|z_0| > \theta$, hence $|z_j| > \theta$, for all $j \geq 0$. So, $1/Q(z;q)$ is analytic in the region $\{(z,q) : |z|, |q| < \theta\}$. Thus, for all $n \geq 1$, the functions $1/Q(q^n;q)$ are all analytic and uniformly bounded by a fixed constant, for $|q| < r_0$, where r_0 is any positive number such that $r_0 < \theta$.

From these considerations, it follows that the central infinite sum that figures in (3.45) is, when $|q| < r_1$, dominated in modulus by a positive multiple of the series

$$\sum_n \frac{r_1^{2n}}{(1-2r_1)^n}, \qquad (3.53)$$

provided that $r_1 < \theta$ and $r_1^2/(1-2r_1) < 1$. Any positive r_1 satisfying $r_1 < \sqrt{2} - 1$ is then admissible. In that case, for $|q| < r_1$, the central sum is a normally convergent sum of analytic functions; hence, it is analytic. ∎

---

[8] The function $v(q) = 1 + q^2/(1-q)$, having nonnegative Taylor coefficients, satisfies $|v(q)| \leq v(|q|)$, for $|q| < 1$; thus, $|1/v(q)| \geq 1/v(|q|)$. Also, $1/v(x)$ decreases from 1 to 0 for $x \in [0,1]$. Hence, with θ the real root of $1/v(\theta) = \theta$, it follows that $|z_0| > \theta$ as soon as $|q| < \theta$.

The radius of analyticity of $PA(q)$ is in fact $1/2$. In order to obtain larger regions of analyticity, one needs to improve on the reasoning underlying the derivation of (3.53). This will result from a transformation of the central infinite sum in (3.45), namely,

$$S(q) := \sum_{n \geq 1} (-1)^n \frac{q^{2n}}{(1-2q)^n} \cdot \frac{1}{Q(q^n; q)}. \tag{3.54}$$

Only the bound $1/Q(q^n; q) = O(1)$ was used in the proof of Lemma 3.20, but we have, for instance, $1/Q(q^n; q) = 1 + O(q^n)$, as $n \to \infty$, and a complete expansion exists. Indeed, since $1/Q$ is bivariate analytic in $|z|, |q| < \theta$, its z-expansion at the origin is of the form

$$\frac{1}{Q(z; q)} = 1 + \sum_{\nu \geq 1} d_\nu(q) z^\nu. \tag{3.55}$$

In particular, at $z = q^n$, we have

$$\frac{1}{Q(q^n; q)} = 1 + \sum_{\nu \geq 1} d_\nu(q) q^{\nu n}. \tag{3.56}$$

Now, consider the effect of an individual term $d_\nu(q)$ (instead of $1/Q(q^n; q)$) on the sum (3.54). The identity

$$\sum_{n \geq 1} (-1)^n \frac{q^{2n}}{(1-2q)^n} q^{\nu n} = -\frac{q^{\nu+2}}{1 - 2q + q^{\nu+2}} \tag{3.57}$$

provides an analytic form for the sum on the left, as long as q is not a pole of the right-hand side. Proceeding formally, we then get, with (3.56) and (3.57), upon exchanging summations in the definition (3.54) of $S(q)$, a form of $PA(q)$ that involves *infinitely many meromorphic elements* of the form $1/(1 - 2q + q^{\nu+2})$.

We shall detail validity conditions for the resulting expansion; see (3.59) below. What matters, as seen from (3.57), is the location of poles of the rational functions $(1 - 2q + q^{\nu+2})^{-1}$, for $\nu \geq 1$. Define the quantities

$$\zeta_k := \text{the root in } [0, 1] \text{ of } 1 - 2x + x^{k+2} = 0. \tag{3.58}$$

We have

$$\zeta_0 = 1; \quad \zeta_1 = \frac{\sqrt{5} - 1}{2} \approx 0.618, \quad \zeta_2 \approx 0.543, \quad \zeta_3 \approx 0.518, \ \ldots$$

and $\zeta_k \to \frac{1}{2}$ as k increases. The location of the complex roots of $1 - 2x + x^{k+2} = 0$ is discussed at length in [44, Ex. V.4], as it is related to the analysis of longest runs in binary strings; a consequence of the principle of the argument (or Rouché's Theorem) is that, apart from the positive real root ζ_k, all other complex roots lie outside the disc $|z| < \frac{3}{4}$. The statement below builds upon this discussion and provides an extended analyticity region for $PA(q)$ as well as a justification of the validity of the expansion resulting from (3.56) and (3.57), which is crucial to subsequent developments.

Lemma 3.21. *The generating function $PA(q)$ is analytic at all points of the slit disc*

$$\mathscr{D}_0 := \left\{ q \;:\; |q| < \tfrac{55}{100}; \; q \notin [\tfrac{1}{2}, \tfrac{55}{100}] \right\}.$$

For $q \in \mathscr{D}_0$, the function $PA(q)$ admits the analytic representation

$$PA(q) = C(q) - A(q)\frac{(v;q)_\infty}{(qu;q)_\infty}\left[\frac{q^2}{(1-q)^2} + \sum_{v\geq 1} d_v(q)\frac{q^{v+2}}{1-2q+q^{v+2}}\right], \qquad (3.59)$$

where

$$d_v(q) = [z^v]\frac{1}{Q(z;q)} \equiv [z^v]\frac{(quz;q)_\infty}{(vz;q)_\infty}.$$

In the disc $|z| < \frac{55}{100}$ *punctured at* $\frac{1}{2}$*, the function $PA(q)$ is meromorphic with simple poles at the points $\zeta_2, \zeta_3, \ldots$, with ζ_k as defined in (3.58). Consequently, the function $PA(q)$ is non-holonomic, and, in particular, non-algebraic.*

Proof. The starting point, noted in the proof of Lemma 3.20, is that fact that $1/Q(z;q)$ is bivariate analytic at all points (z,q) such that $|z|, |q| < \theta$, where $\theta \approx 0.56984$ is specified in (3.52). Cauchy's coefficient formula,

$$d_v(q) = \frac{1}{2i\pi}\int_{|z|=\theta_1}\frac{1}{Q(z;q)}\frac{dz}{z^{v+1}},$$

is applicable for any θ_1 such that $0 < \theta_1 < \theta$. Let us set $\theta_1 = \frac{56}{100}$. Then, since $1/Q(z;q)$ is analytic, hence continuous, hence bounded, for $|z| \leq \theta_1$ and $|q| \leq \theta_1$, trivial bounds applied to the Cauchy integral yield

$$|d_v(q)| < C \cdot \theta_1^{-v}, \qquad (3.60)$$

for some absolute constant $C > 0$.

Consider the double sum resulting from the substitution of (3.56) into (3.54),

$$S(q) = \sum_{n\geq 1}(-1)^n\frac{q^{2n}}{(1-2q)^n}\cdot\left(1 + \sum_{v\geq 1}d_v(q)q^{vn}\right).$$

If we constrain q to be small, say $|q| < \frac{1}{10}$, we see from (3.60) that the double sum is absolutely convergent. Hence, the form (3.59) is justified for such small values of q. We can then proceed by analytic continuation from the right-hand side of (3.59). The bound (3.60) grants us the fact that the sum that appears there is indeed analytic in $\mathscr{D}_/$. The statements, relative to the analyticity domain and the alternative expansion (3.59) follow. Finally, since the value $\frac{1}{2}$ corresponds to an accumulation of poles (it is straightforward to check that the points ζ_k are indeed poles, i.e. the coefficient $A(q)(v;q)_\infty/(qu;q)_\infty$ is non-zero at $q = \zeta_k$), the function $PA(q)$ is non-holonomic (see, e.g., [41] for context). ∎

As an immediate consequence of the dominant singularity being at $\frac{1}{2}$, the coefficients pa_n must obey a weak asymptotic law of the form

$$pa_n = 2^n \theta(n), \qquad \text{where} \quad \limsup_{n \to \infty} \theta(n)^{1/n} = 1,$$

that is, $\theta(n)$ is a (currently unknown) subexponential factor.

More precise information requires a better characterisation of the behaviour of $S(q)$, as q approaches the dominant singularity $\frac{1}{2}$. This itself requires a better understanding of the coefficients $d_\nu(q)$. To this end, we state a general and easy lemma about the coefficients of quotients of q-factorials.

Lemma 3.22. *Let a be a fixed complex number satisfying $|a| < 1$ and let q satisfy $|q - \frac{1}{2}| < \frac{1}{10}$. One has, for $\nu \geq 1$*

$$[z^\nu] \frac{(az;q)_\infty}{(z;q)_\infty} = \frac{1}{(q;q)_\infty} \sum_{j=0}^{\infty} \frac{(aq^{-j};q)_\infty}{(q^{-j};q)_j} \cdot q^{j\nu}. \tag{3.61}$$

Proof. The function $h(z) := (az;q)_\infty / (z;q)_\infty$ has simple poles at the points $\overline{z}_j := q^{-j}$, for $j \geq 0$. We have

$$h(z) \underset{z \to \overline{z}_j}{\sim} \frac{e_j(a;q)}{1 - zq^j}, \qquad e_j(a;q) := \frac{(aq^{-j};q)_\infty}{(q^{-j};q)_j (q;q)_\infty}.$$

The usual expansion of coefficients of meromorphic functions [44, Thm. IV.10] immediately implies a terminating form for any $J \in \mathbb{Z}_{\geq 0}$:

$$[z^\nu]h(z) = \sum_{j=0}^{J} e_j(a;q)q^{j\nu} + O(R_J^n), \tag{3.62}$$

where we may adopt $R_J = \frac{3}{2}q^{-J}$.

The last estimate (3.62) corresponds to an evaluation by residues of the Cauchy integral representation of coefficients,

$$[z^\nu]h(z) = \frac{1}{2i\pi} \int_{|z|=R_J} h(z) \frac{dz}{z^{\nu+1}}.$$

Now, let J tend to infinity. The quantity R_J lies approximately midway between two consecutive poles, q^{-J} and q^{-J-1}, and it can be verified elementarily that, throughout $|z| = R_J$, the function $h(z)$ remains bounded in modulus by an absolute constant (this requires the condition $|a| < 1$). It then follows that we can let J tend to infinity in (3.62). For $\nu \geq 1$, the coefficient integral taken along $|z| = R_J$ tends to 0, so that, in the limit, the exact representation (3.61) results. ∎

The formula (3.61) is equivalent to the partial fraction expansion (Mittag-Leffler expansion; see [67, Sec. 7.10]) of the function $h(z)$, which is meromorphic in the whole complex plane:

$$\frac{(az;q)_\infty}{(z;q)_\infty} = 1 + \frac{1}{(q;q)_\infty} \sum_{j=0}^{\infty} \frac{(aq^{-j};q)_\infty}{(q^{-j};q)_j} \frac{zq^j}{1 - zq^j}. \tag{3.63}$$

(The condition $|a| < 1$ ensures the convergence of this expansion.)

A consequence of Lemma 3.22 is an expression for the coefficients $d_\nu(q) = [z^\nu] Q(z;q)^{-1}$, with $Q(z;q)$ defined by (3.47):

$$d_\nu(q) = \frac{1}{(q;q)_\infty} \sum_{j=0}^{\infty} \frac{(quv^{-1}q^{-j};q)_\infty}{(q^{-j};q)_j} \cdot \left(vq^j\right)^\nu, \qquad \nu \geq 1. \tag{3.64}$$

To see this, set

$$a = quv^{-1} = \frac{q^2}{1-q+q^2},$$

and replace z by zv in the definition of $h(z)$. Note that at $q = 1/2$, we have $u = 1$, $v = 3/2$, $a = 1/3$, so that, for $q \approx 1/2$, we expect $d_\nu(q)$ to grow roughly like $(3/2)^\nu$.

Summarizing the results obtained so far, we state:

Proposition 3.23. *The generating function of 3-sided prudent polygons satisfies the identity*

$$PA(q) = D(q) - q^2 A(q) \frac{(a;q)_\infty (v;q)_\infty}{(q;q)_\infty (av;q)_\infty} \sum_{\nu=1}^{\infty} \sum_{j=0}^{\infty} \left[\frac{(aq^{-j};q)_j}{(q^{-j};q)_j} \cdot \frac{v^\nu q^{(j+1)\nu}}{1-2q+q^{\nu+2}} \right], \tag{3.65}$$

where

$$a = \frac{q^2}{1-q+q^2}, \quad v = \frac{1-q+q^2}{1-q}, \quad D(q) = C(q) - \frac{q^2}{(1-q)^2} A(q) \frac{(v;q)_\infty}{(av;q)_\infty}, \tag{3.66}$$

and $A(q), C(q)$ are rational functions defined in (3.46).

Proof. The identity is a direct consequence of the formula (3.64) for $d_\nu(q)$ and of the expression for $PA(q)$ in (3.59), using the equivalence $av = qu$ and the simple reorganization

$$(aq^{-j};q)_\infty = (aq^{-j};q)_j \cdot (a;q)_\infty.$$

Previous developments imply that the identity (3.65) is, in particular, valid in the real interval $(0, \frac{1}{2})$. The trivial equality

$$\frac{(aq^{-j};q)_j}{(q^{-j};q)_j} = \frac{(a-q)(a-q^2)\cdots(a-q^j)}{(1-q)(1-q^2)\cdots(1-q^j)} \tag{3.67}$$

then shows that the expression on the right-hand side indeed represents a *bona fide* formal power series in q, since the q-valuation of the general term of the double sum in (3.65) increases with both j and ν. ∎

The formula (3.65) of Proposition 3.23 will serve as the starting point of the asymptotic analysis of $PA(q)$ as $q \to 1/2$ in the next subsection. Given the discussion of the analyticity of the various components in the proof of Lemma 3.20, the task essentially reduces to estimating the double sum in a suitable complex neighbourhood of $q = 1/2$.

Mellin analysis

Let $T(q)$ be the double sum that appears in the expression (3.65) of $PA(q)$. We shall take it here in the form

$$T(q) = \sum_{j=0}^{\infty} \frac{(aq^{-j};q)_j}{(q^{-j};q)_j} H_j(q) \quad \text{where} \quad H_j(q) := \sum_{\nu=1}^{\infty} \frac{v^\nu q^{(j+1)\nu}}{1 - 2q + q^{\nu+2}}. \tag{3.68}$$

We will now study the functions H_j and propose to show that those of greater index contribute less significant terms in the asymptotic expansion of $PA(q)$ near $q = 1/2$. In this way, a complete asymptotic expansion of the function $PA(q)$, and hence of its coefficients pa_n, can be obtained.

The main technique used here is that of Mellin transforms: we refer the reader to [40] for details of the method. The principles are recalled below. We then proceed to analyse the double sum T of (3.68) when q is *real* and q tends to 1/2. The corresponding expansion is fairly explicit and it is obtained at a comparatively low computational cost. We finally show that the expansion extends to a *sector of the complex plane* around $q = 1/2$.

Principles of the Mellin analysis

Let $f(x)$ be a complex function of the real argument x. Its *Mellin transform*, denoted by $f^\star(s)$ or $\mathcal{M}[f]$, is defined as the integral

$$\mathcal{M}[f](s) \equiv f^\star(s) := \int_0^\infty f(x) x^{s-1} \, dx, \tag{3.69}$$

where s may be complex. It is assumed that $f(x)$ is locally integrable. It is then well known that if f satisfies the two asymptotic conditions

$$f(x) \underset{x \to 0}{=} O(x^\alpha), \qquad f(x) \underset{x \to +\infty}{=} O(x^\beta),$$

with $\alpha > \beta$, then $f^\star$ is an analytic function of s in the strip of the complex plane,

$$-\alpha < \Re(s) < -\beta,$$

also known as a *fundamental strip*. Then, with c any real number of the interval $(-\alpha, -\beta)$, the following inversion formula holds (see [121, §VI.9] for detailed statements):

$$f(x) = \frac{1}{2i\pi} \int_{c-i\infty}^{c+i\infty} f^\star(s) x^{-s} \, ds. \tag{3.70}$$

There are then two essential properties of Mellin transfoms.

($\mathbf{M_1}$) *Harmonic sum property.* If the pairs (λ, μ) range over a denumerable subset of $\mathbb{R} \times \mathbb{R}_{>0}$ then one has the equality

$$\mathcal{M}\left[\sum_{(\lambda,\mu)} \lambda f(\mu x) \right](s) = f^\star(s) \cdot \left(\sum_{(\lambda,\mu)} \lambda \mu^{-s} \right). \tag{3.71}$$

That is to say, the harmonic sum $\sum \lambda f(\mu x)$ has a Mellin transform that decomposes as a product involving the transform of the base function ($f^\star$) and the generalized Dirichlet series ($\sum \lambda \mu^{-s}$) associated with the "amplitudes" λ and the "frequencies" μ. Detailed validity conditions, spelled out in [40], are that the exchange of summation and integral be permissible.

($\mathbf{M_2}$) *Mapping properties.* Poles of transforms are in correspondence with asymptotic expansions of the original function. More precisely, if the Mellin transform $F^\star$ of a function F admits a meromorphic extension beyond the fundamental strip, with a pole of some order m at some point $s_0 \in \mathbb{C}$, with $\Re(s_0) < -\alpha$, then it contributes an asymptotic term of the form $P(\log x) x^{-s_0}$ in the expansion of $F(x)$ as $x \to 0$, where P is a computable polynomial of degree $m - 1$. Schematically:

$$F^\star(s) \underset{s \to s_0}{\asymp} \frac{C}{(s - s_0)^m} \qquad \Longrightarrow \qquad F(x) \underset{x \to 0}{\asymp} P(\log x) x^{-s_0} = \operatorname{Res}\left(f^\star(s) x^{-s}\right)_{s = s_0}. \quad (3.72)$$

Detailed validity conditions, again spelled out in [40], are a suitable decay of the transform $F^\star(s)$, as $\Im(s) \to \pm\infty$, so as to permit an estimate of the inverse Mellin integral (3.70) by residues – in (3.72), the expression is then none other than the residue of $f^\star(s) x^{-s}$ at $s = s_0$.

The power of the Mellin transform for the asymptotic analysis of sums devolves from the application of the mapping property ($\mathbf{M_2}$) to functions $F(x) = \sum \lambda f(\mu x)$ that are harmonic sums in the sense of ($\mathbf{M_1}$). Indeed, the factorisation property (3.71) of ($\mathbf{M_1}$) makes it possible to analyse separately the singularities that arise from the base function (via $f^\star$) and from the amplitude–frequency pairs (via $\sum \lambda \mu^{-s}$); hence an asymptotic analysis results, thanks to ($\mathbf{M_2}$).

Analysis for real values of $q \to 1/2$

Our purpose now is to analyse the quantity T of (3.68) with $q < 1/2$, when $q \to 1/2$. This basically reduces to analysing the quantities $H_j(q)$ of (3.68). Our approach consists of setting $t = 1 - 2q$ and *decoupling*[9] the quantities t and q. Accordingly, we define the function

$$h_j(t) \equiv h_j(t; q, v) := q^{-2} \sum_{v=1}^{\infty} \frac{(v q^j)^v}{1 + t q^{-v-2}}, \qquad (3.73)$$

so that

$$H_j(q) = h_j(t; q, v(q)),$$

with the definition (3.68). We shall let t range over $\mathbb{R}_{\geq 0}$ but restrict the parameter q to a small interval $(1/2 - \epsilon_0, 1/2 + \epsilon_0)$ of $\mathbb{R}$ and the parameter v to a small interval of the form $(3/2 - \epsilon_1, 3/2 + \epsilon_1)$, since $v(1/2) = 3/2$. We shall write such a restriction as

$$q \approx \frac{1}{2}, \qquad v \approx \frac{3}{2},$$

---

[9]An instance of such a decoupling technique appears in De Bruijn's reference text [24].

74

with the understanding that ϵ_0, ϵ_1 can be taken to be suitably small, as the need arises. Thus, *for the time being, we ignore the relations that exist between t and the pair q, v, and we shall consider them as independent quantities.*

As a preamble to the Mellin analysis, we state an elementary lemma.

Lemma 3.24. *Let q be restricted to a sufficiently small interval containing 1/2, and v to a sufficiently small interval containing 3/2. Each function $h_j(t)$ defined by (3.73) satisfies the estimate*

$$h_j(t) \underset{t \to +\infty}{=} O\left(\frac{1}{t}\right), \qquad h_j(t) \underset{t \to 0}{=} \begin{cases} O(1) & \text{if } j \geq 1 \\ O(t^{-\gamma}) & \text{if } j = 0, \text{ with } \gamma = \frac{\log v}{\log(1/q)}. \end{cases} \tag{3.74}$$

For γ, we can also adopt any fixed value larger than $\log_2(4/3) \approx 0.415$, provided q and v are taken close enough to 1/2 and 3/2, respectively.

Proof. Behaviour as $t \to +\infty$. The inequality $(1 + tq^{-\nu-2})^{-1} < t^{-1}q^{\nu+2}$ implies by summation the inequality

$$h_j(t) \leq q^{-2}t^{-1}\sum_{\nu=1}^{\infty} v^{\nu}q^{j\nu}q^{\nu} = O\left(\frac{1}{t}\right), \qquad t \to +\infty,$$

given the convergence of the geometric series $\sum_{\nu} vq^{(j+1)\nu}$, for $v \approx 3/2$ and $q \approx 1/2$.

Behaviour as $t \to 0$. First, for the easy case $j \geq 1$, the trivial inequality $(1 + tq^{-\nu-2})^{-1} \leq 1$ implies

$$h_j(t) = O\left(\sum_{\nu}(vq^j)^{\nu}\right) = O(1), \qquad t \to 0.$$

Next, for $j = 0$, define the function

$$\nu_0(t) := -2 + \frac{\log(1/t)}{\log(1/q)},$$

so that $tq^{-\nu-2} < 1$, if $\nu < \nu_0(t)$, and $tq^{-\nu-2} \geq 1$, if $\nu \geq \nu_0(t)$. Write $\sum_{\nu} = \sum_{\nu_0} + \sum_{\nu \geq \nu_0}$. The sum corresponding to $\nu \geq \nu_0$ is bounded from above as in the case of $t \to +\infty$,

$$\sum_{\nu \geq \nu_0(t)} \frac{v^{\nu}}{1 + tq^{-\nu-2}} \leq \sum_{\nu \geq \nu_0(t)} v^{\nu}t^{-1}q^{\nu+2} = O\left(t^{-1}(vq)^{\nu_0}\right) = O\left(t^{-1}(vq)^{\nu_0}\right), \qquad t \to 0,$$

and the last quantity is $O(t^{-\gamma})$ for $\gamma = (\log v)/\log(1/q)$. The sum corresponding to $\nu < \nu_0$ is dominated by its later terms and is accordingly found to be $O(t^{-\gamma})$. The estimate of $h_0(t)$, as $t \to 0$, results. ∎

We can now proceed with a precise asymptotic analysis of the functions $h_j(t)$, as $t \to 0$. Lemma 3.24 implies that each $h_j(t)$ has its Mellin transform $h_j^\star(s)$ that exists in a non-empty fundamental strip left of $\Re(s) = 1$. In that strip, the Mellin transform is

$$\mathcal{M}[h_j(t)] = q^{-2}\mathcal{M}\left[\frac{1}{1+t}\right] \cdot \left(\sum_{\nu=1}^{\infty}(vq^j)^{\nu}(q^{-\nu-2})^{-s}\right) \qquad \text{(by } \mathbf{M_1})$$

$$= q^{-2}\mathcal{M}\left[\frac{1}{1+t}\right] \cdot \frac{vq^{j+3s}}{1-vq^{j+s}} \tag{3.75}$$

$$= q^{-2}\frac{\pi}{\sin\pi s}\frac{vq^{j+3s}}{1-vq^{j+s}} \qquad \text{(by the classical form of } \mathcal{M}[(1+t)^{-1}]).$$

The Mellin transform of $(1+t)$, which equals $\pi/\sin(\pi s)$, admits $0 < \Re(s) < 1$ as the fundamental strip, so this condition is necessary for the validity of (3.75). In addition, the summability of the Dirichlet series, here plainly a geometric series, requires the condition $|vq^{j+s}| < 1$; that is,

$$\Re(s) > -j + \frac{\log v}{\log 1/q}.$$

In summary, the validity of (3.75) is ensured for s satisfying

$$\lambda < \Re(s) < 1, \qquad \text{with} \quad \lambda := \max\left(0, -j + \frac{\log v}{\log 1/q}\right).$$

Lemma 3.25. *For $q \approx 1/2$ and $v \approx 3/2$ restricted as in Lemma 3.24, the function $h_j(t)$ admits an exact representation, valid for any $t \in (0, q^{-3})$,*

$$h_j(t) = (-1)^j \frac{vq^{3\gamma-2j-2}}{\log 1/q} t^{j-\gamma}\Pi(\log_{1/q} t) + q^{-2}\sum_{r \geq 0}(-1)^r \frac{vq^{j-3r}}{1-vq^{j-r}}t^r. \tag{3.76}$$

Here,

$$\gamma \equiv \gamma(q) := \frac{\log v}{\log 1/q}$$

so that $\gamma \approx \log_2 \frac{3}{2} = 0.415\ldots$, when $q \approx \frac{1}{2}$; the quantity $\Pi(u)$ is an absolutely convergent Fourier series,

$$\Pi(u) := \sum_{k \in \mathbb{Z}} p_k e^{-2ik\pi u}, \tag{3.77}$$

with coefficients p_k given explicitly by

$$p_k = \frac{\pi}{\sin(\pi\gamma + 2ik\pi^2/(\log 1/q))}. \tag{3.78}$$

Observe that the p_k decrease geometrically with k. For instance, at $q = 1/2$, one has

$$p_k = O\left(e^{-2k\pi^2/\log 2}\right) = O\left(4.28 \cdot 10^{-13}\right)^k, \tag{3.79}$$

as is apparent from the exponential form of the sine function. Consequently, even the very first coefficients are small: at $q = 1/2$, typically,

$$|p_1| = |p_{-1}| \approx 2.69 \cdot 10^{-12}, \quad |p_2| = |p_{-2}| \approx 1.15 \cdot 10^{-24}, \quad |p_3| = |p_{-3}| \approx 4.95 \cdot 10^{-37}.$$

Proof. We first perform an *asymptotic* analysis of $h_j(t)$ as $t \to 0^+$. This requires the determination of poles to the left of the fundamental strip of $h_j^\star(s)$, and these arise from two sources.

— The relevant poles of $\pi/\sin \pi s$ are at $s = 0, -1, -2, \ldots$; they are simple and the residue at $s = -r$ is $(-1)^r$.

— The quantity $(1 - vq^{j+s})^{-1}$ has a simple pole at the real point

$$\sigma_0 := -j + \frac{\log v}{\log 1/q}, \tag{3.80}$$

as well as *complex poles* of real part σ_0, due to the complex periodicity of the exponential function $(e^{t+2i\pi} = e^t)$. The set of all poles of $(1 - vq^{j+s})^{-1}$ is then

$$\left\{ \sigma_0 + \frac{2ik\pi}{\log 1/q}, \quad k \in \mathbb{Z} \right\}.$$

The proof of an asymptotic representation (that is, of (3.76), with '$\sim$' replacing the equality sign there) is classically obtained by integrating $h_j^\star(s)t^{-s}$ along a long rectangle with corners at $-d - iT$ and $c + iT$, where c lies within the fundamental strip (in particular, between 0 and 1) and d will be taken to be of the form $-m - \frac{1}{2}$, with $m \in \mathbb{Z}_{\geq 0}$, and smaller than $-j + \gamma$. In the case considered here, there are regularly spaced poles along $\Re(s) = -j + \gamma$, so that one should take values of T that are such that the line $\Im(s) = T$ passes half-way between poles. This, given the fast decay of $\pi/\sin \pi s$ as $|\Im(s)|$ increases and the boundedness of the Dirichlet series $(1 - vq^{j+s})^{-1}$ along $\Im(s) = \pm T$, allows us to let T tend to infinity. By the Residue Theorem applied to the inverse Mellin integral (3.70), we collect in this way the contribution of *all* the poles at $-j + \gamma + 2ik\pi/(\log 1/q)$, with $k \in \mathbb{Z}$, as well as the $m + 1$ initial terms of the sum $\sum_r$ in (3.76). The resulting expansion is of type (3.76) with the sum $\sum_r$ truncated to $m + 1$ terms and an error term that is $O(t^{m+1/2})$.

In general, what the Mellin transform method gives is an *asymptotic* rather than *exact* representation of this type. Here, we have more. We can finally let m tend to infinity and verify that the inverse Mellin integral (3.70) taken along the vertical line $\Re(s) = -m - \frac{1}{2}$ remains uniformly bounded in modulus by a quantity of the form $ct^m q^{-3m}$, for some $c > 0$. In the limit $m \to +\infty$, the integral vanishes (as long as $tq^{-3} < 1$), and the exact representation (3.76) is obtained. ∎

We can now combine the *identity* provided by Lemma 3.25 with the decomposition of the generating function $PA(q)$ as allowed by Equation (3.68), which flows from Proposition 3.23. We recall that $H_j(q) = h_j(t; q, v(q))$.

Proposition 3.26. *The generating function $PA(q)$ of prudent polygons satisfies, for q in a small enough interval*[10] *of the form $(1/2 - \epsilon, 1/2)$ (for some $\epsilon > 0$), the* identity

$$PA(q) = D(q) - q^2 A(q) \frac{(a;q)_\infty (v;q)_\infty}{(q;q)_\infty (av;q)_\infty} T(q), \tag{3.81}$$

---

[10]Numerical experiments suggest that in fact the formula (3.82) remains valid for all $q \in (0, 1/2)$.

where the notations are those of Proposition 3.23, and the function $T(q)$ admits the exact representation

$$T(q) = (1 - 2q)^{-\gamma} \cdot \Pi\left(\frac{\log(1-2q)}{\log 1/q}\right) U(q) + V(q), \qquad \gamma \equiv \frac{\log v}{\log 1/q}, \qquad (3.82)$$

with $\Pi(u)$ given by Lemma 3.25, Equations (3.77) and (3.78). Set

$$t = 1 - 2q.$$

The "singular series" $U(q)$ is

$$U(q) = \frac{v q^{3\gamma - 2}}{\log 1/q} \frac{(-q^{-1}t; q)_\infty}{(-aq^{-2}t; q)_\infty}, \qquad \gamma = \frac{\log v}{\log 1/q}; \qquad (3.83)$$

and the "regular series" $V(q)$ is

$$V(q) = -\frac{(q;q)_\infty}{(a;q)_\infty} \frac{q^{-2}}{1 + q^{-2}t} + q^{-2} \frac{(q;q)_\infty (av;q)_\infty}{(a;q)_\infty (v;q)_\infty} \sum_{r=0}^{\infty} \frac{(a^{-1}v^{-1}q; q)_r}{(v^{-1}q; q)_r} \left(-aq^{-2}t\right)^r. \qquad (3.84)$$

Proof. We start from $T(q)$ as defined by (3.68). The q-binomial theorem is the identity [48, Sec. 1.3]

$$\frac{(\theta z; q)_\infty}{(z; q)_\infty} = \sum_{n=0}^{\infty} \frac{(\theta; q)_n}{(q; q)_n} z^n. \qquad (3.85)$$

Now consider the first term in the expansion (3.76) of Lemma 3.25. Sum the corresponding contributions for all values of $j \geq 0$, after multiplication by the coefficient $\frac{(aq^{-j}; q)_j}{(q^{-j}; q)_j}$, in accordance with (3.68). This gives

$$U(q) = \frac{v q^{3\gamma - 2}}{\log 1/q} \sum_{j=0}^{\infty} \frac{(aq^{-j}; q)_j}{(q^{-j}; q)_j} (-q^2 t)^j = \frac{v q^{3\gamma - 2}}{\log 1/q} \sum_{j=0}^{\infty} \frac{(a^{-1}q; q)_j}{(q; q)_j} (-aq^{-2}t)^j$$

which provides the expression for $U(q)$ of the singular series, via the q-binomial theorem (3.85) taken with $z = -at$ and $\theta = a^{-1}q$.

Summing over j in the second term in the identity (3.76) of Lemma 3.25, we have

$$V(q) = q^{-2} \sum_{r=0}^{\infty} (-q^{-2}t)^r \sum_{j=0}^{\infty} \frac{(aq^{-j}; q)_j}{(q^{-j}; q)_j} \frac{vq^{j-r}}{1 - vq^{j-r}}.$$

Now, the Mittag-Leffler expansion (3.63) associated with Lemma 3.22 can be put in the form

$$\frac{(az; q)_\infty}{(z; q)_\infty} = 1 + \frac{(a; q)_\infty}{(q; q)_\infty} \sum_{j=0}^{\infty} \frac{(aq^{-j}; q)_j}{(q^{-j}; q)_j} \frac{zq^j}{1 - zq^j}.$$

An application of this identity to $V(q)$, with $z = vq^{-r}$, shows that

$$V(q) = q^{-2} \frac{(q;q)_\infty}{(a;q)_\infty} \sum_{r=0}^{\infty} (-q^{-2}t)^r \left(\frac{(avq^{-r}; q)_\infty}{(vq^{-r}; q)_\infty} - 1\right),$$

which is equivalent to the stated form of $V(q)$. Note that this last form is a q-hypergeometric function of type ${}_2\phi_1$; see [48].

So far, we have proceeded formally and left aside considerations of convergence. It can be easily verified that all the sums, single or double, involved in the calculations above are absolutely (and uniformly) convergent, provided t is taken small enough (i.e., q is sufficienty close to 1/2), given that all the involved parameters, such as a, u, v, then stay in suitably bounded intervals of the real line. ∎

Analysis for complex values of $q \to 1/2$

We now propose to show that the "transcendental" expression of $PA(q)$ provided by Proposition 3.26 is actually valid in certain regions of the complex plane that extend beyond an interval of the real line. The regions to be considered are dictated by the requirements of the singularity analysis method to be deployed in the next subsection.

Let θ_0 be a number in the interval $(0, \pi/2)$, called the *angle*, and r_0 a number in $\mathbb{R}_{>0}$, called the *radius*. We define a *sector* (anchored at 1/2) to be the set of all complex numbers $z = 1/2 + re^{i\theta}$ such that

$$0 < r < r_0 \qquad \text{and} \qquad \theta_0 < \theta < 2\pi - \theta_0.$$

We stress the fact that the angle should be strictly smaller than $\pi/2$, so that a sector in the sense of the definition always includes a part of the line $\Re(s) = 1/2$. The smallness of a sector will be measured by the smallness of r_0. That is to say:

Proposition 3.27. *There exists a sector $\mathscr{S}_0$ (anchored at 1/2), of angle[11] $\theta_0 < \pi/2$ and radius $r_0 > 0$, such that the identity expressed by (3.81) and (3.82) holds for all $q \in \mathscr{S}_0$.*

Proof. The proof is a simple consequence of analytic continuation. We first observe that an infinite product such as $(c; q)_\infty$ is an analytic function of both c and q, for arbitrary c and $|q| < 1$. Similarly, the inverse $1/(c; q)_\infty$ is analytic provided $cq^j \neq 1$, for all c. For instance, taking $c = a$ where $a = a(q) = q^2/(1 - q + q^2)$ and noting that $a(1/2) = 1/3$, we see that $1/(a; q)$ is an analytic function of q in a small complex neighbourhood of $q = 1/2$. This reasoning can be applied to the various Pochhammer symbols that appear in the definition of $T(q), U(q), V(q)$. Similarly, the hypergeometric sum that appears in the regular series $V(q)$ is seen to be analytic in the three quantities $a \approx 1/3$, $v \approx 3/2$, and $t = 1 - 2q \approx 0$. In particular, *the functions $U(q)$ and $V(q)$ are analytic in a complex neighbourhood of $q = 1/2$.*

Next, consider the quantity

$$(1 - 2q)^{-\gamma} = \exp(-\gamma \log(1 - 2q)).$$

[11] A careful examination of the proof of Proposition 3.27 shows that *any* angle $\theta_0 > 0$, however small, is suitable, but only the existence of *some* $\theta_0 < \pi/2$ is needed for singularity analysis.

It is clear that the function $\gamma \equiv \gamma(q)$ is analytic in a neighbourhood of $q = 1/2$, since it equals $(\log v)/(\log 1/q)$. The logarithm, $\log(1 - 2q)$, is analytic in any *sector* anchored at $1/2$. By composition, there results that $(1 - 2q)^{-\gamma}$ is analytic in a small sector anchored at $1/2$. It only remains to consider the Π factor in (3.81). A single Fourier element, $p_k e^{-2ik\pi u}$, with $u = \log_{1/q} t$ and $t = 1 - 2q$, is also analytic in a small sector (anchored at $1/2$), as can be seen from the expression

$$p_k e^{-2ik\pi u} = p_k \exp\left(-2ik\pi \frac{\log(1 - 2q)}{\log 1/q} \right). \qquad (3.86)$$

Note that, while $\Re(\log(1 - 2q)) \to \infty$ as $q \to 1/2$, the complex exponential $\exp(2ik\pi \log_2(1 - 2q))$ remains uniformly bounded, since $\Im(\log(1 - 2q))$ is bounded for q in a sector. Then, given the fast geometric decay of the coefficients p_k at $q = 1/2$ (in particular $p_k = O(e^{-2k\pi^2/\log 2})$; cf. (3.78)), it follows that $\Pi(\log_2 t)$ is also analytic in a sector. A crude adjustment of this argument (see (3.95) and (3.96) below for related expansions) suffices to verify that the geometric decay of the terms composing (3.86) persists in a sector anchored at $1/2$, so that $\Pi(\log_{1/q} t)$ is also analytic in such a sector.

Finally, the auxiliary quantities $D(q), A(q)$ are meromorphic at $q = 1/2$, with at most a double pole there; in particular, they are analytic in a small enough sector anchored at $1/2$. We can then choose for $\mathscr{S}_0$ a small sector that satisfies this as well as all the previous analyticity constraints. Then, by *unicity of analytic continuation*, the expression on the right-hand side of (3.81), with $T(q)$ as given by (3.82), *must* coincide with (the analytic continuation of) $PA(q)$ in the sector $\mathscr{S}_0$. ∎

Singularity analysis and transfer

If we drastically reduce all the non-singular quantities that occur in the main form (3.81) of Proposition 3.26 by letting $q \to 1/2$, we are led to infer that $PA(q)$ satisfies, in a sector around $q = 1/2$, an estimate of the form

$$PA(q) = \xi_0 (1 - 2q)^{-\gamma_0 - 2} \Pi(\log_2(1 - 2q)) + O\left((1 - 2q)^{-3/2}\right), \qquad \gamma_0 := \log_2(3/2), \qquad (3.87)$$

where

$$\xi_0 = -\frac{1}{16} U(1/2) \frac{(1/3; 1/2)_\infty (3/2; 1/2)_\infty}{(1/2; 1/2)_\infty (1/2; 1/2)_\infty}, \qquad U(1/2) = \frac{16}{9 \log 2}, \qquad (3.88)$$

and $U(q)$ is the singular series of (3.83). Let us ignore for the moment the oscillating terms and simplify $\Pi(u)$ to its constant term p_0, with p_k given by (3.78). This provides a numerical approximation $\widehat{PA}(q)$ of $PA(q)$. With the general asymptotic approximation (derived from Stirling's formula)

$$[q^n](1 - 2q)^{-\lambda} \underset{n \to +\infty}{\sim} \frac{1}{\Gamma(\lambda)} 2^n n^{\lambda - 1}, \qquad \lambda \notin \mathbb{Z}_{\leq 0}, \qquad (3.89)$$

it is easily seen that $[q^n]\widehat{PA}(q)$ is asymptotic to the quantity $x_0 2^n n^g$ of Equation (3.43) in Theorem 3.19, which is indeed the "principal" asymptotic term of $pa_n = [q^n]PA(q)$, where $g = \gamma_0 + 1 = \log_2 3$.

A rigorous justification and a complete analysis depend on the general singularity analysis theory [44, Ch. VI] applied to the expansion of $PA(q)$ near $q = 1/2$ (see also Chapter 1 for a brief overview). We recall that a Δ-*domain with base* 1 is defined to be the intersection of a disc of radius strictly larger than 1 and of the complement of a sector of the form $-\theta_0 < \arg(z - 1) < \theta_0$ for some $\theta_0 \in (0, \pi/2)$. A Δ-domain with base ρ is obtained from a Δ-domain with base 1 by means of the homothetic transformation $z \mapsto \rho z$. Singularity analysis theory is then based on two types of results.

(S_1) The coefficients of functions in a basic asymptotic scale have known asymptotic expansions [44, Thm. VI.1]. In the case of the scale $(1 - z)^{-\lambda}$, the expansion, which extends (3.89), is of the form

$$[z^n](1 - z)^{-\lambda} \underset{n \to +\infty}{\sim} n^{\lambda - 1}\left(1 + \sum_{k \geq 1} \frac{e_k}{n^k}\right), \qquad \lambda \in \mathbb{C} \setminus \mathbb{Z}_{\geq 0},$$

where e_k is a computable polynomial in λ of degree $2k$. Observe that this expansion is valid for *complex* values of the exponent λ, and if $\lambda = \sigma + i\tau$, then

$$n^{\lambda - 1} = n^{\sigma - 1} \cdot n^{i\tau} = n^{\sigma - 1} e^{i\tau \log n}.$$

Thus, the real part (σ) of the singular exponent drives the asymptotic regime; the imaginary part, as soon as it is nonzero, induces *periodic oscillations* in the scale of $\log n$. A noteworthy feature is that smaller functions at the singularity $z = 1$ have asymptotically smaller coefficients.

(S_2) An approximation of a function near its singularity can be transferred to an approximation of coefficients according to the rule

$$f(z) \underset{z \to 1}{=} O\left((1 - z)^{-\lambda}\right) \qquad \Longrightarrow \qquad [z^n]f(z) \underset{n \to +\infty}{=} O\left(n^{\lambda - 1}\right).$$

The condition is that $f(z)$ be analytic in a Δ–domain and that the O–approximation holds in such a Δ-domain, as $z \to 1$; see [44, Thm. VI.3]. Once more, smaller error terms are associated with smaller coefficients.

Equipped with these principles, it is possible to obtain a complete asymptotic expansion of $[q^n]PA(q)$ once a complete expansion of $PA(q)$ in the vicinity of $q = 1/2$ has been obtained (set $q = z/2$, so that $z \approx 1$ corresponds to $q = 1/2$). In this context, Proposition 3.23 precisely grants us the analytic continuation of $PA(q)$ in a Δ-domain anchored at $1/2$, with any opening angle arbitrarily small; Proposition 3.26, together with Proposition 3.27, describe in a precise manner the asymptotic form of $PA(q)$ as $q \to 1/2$ in a Δ-domain and it is a formal exercise to transform them into a standard asymptotic expansion, in the form required by singularity analysis theory.

Proposition 3.28. *As $q \to 1/2$ in a Δ-domain, the function $PA(q)$ satisfies the expansion*

$$PA(q) \underset{q \to 1/2}{\sim} \frac{1}{(1-2q)} + R(q) + \sum_{j \geq 1} (1-2q)^{-\gamma_0 - 2 + j} \sum_{\ell=0}^{j} (\log(1-2q))^\ell \, \Pi^{(j,\ell)}(\log_2(1-2q)), \quad (3.90)$$

with $\gamma_0 = \log_2(3/2)$. Here $R(q)$ is analytic at $q = 1/2$ and each $\Pi^{(j,\ell)}(u)$ is a Fourier series

$$\Pi^{(j,\ell)}(u) := \sum_{k \in \mathbb{Z}} p_k^{(j,\ell)} e^{2ik\pi u},$$

with a computable sequence of coefficients $p_k^{(j,\ell)}$.

Proof. From Proposition 3.26, we have

$$PA(q) = PA^{\text{reg}}(q) + PA^{\text{sing}}(q), \quad (3.91)$$

where the two terms correspond, respectively, to the "regular" part (involving the regular series $V(q)$) and the "singular part" (involving the singular series $U(q)$ as well as the factor $(1-2q)^{-\gamma}$ and the oscillating series).

Regarding the regular part, we have, with the notations of Proposition 3.26,

$$PA^{\text{reg}}(q) = D(q) - q^2 A(q) \frac{(a;q)_\infty (v;q)_\infty}{(q;q)_\infty (q;q)_\infty} V(q). \quad (3.92)$$

We already know that $A(q)$ and $D(q)$ are meromorphic at $q = 2$ with a double pole, while $V(q)$ and the Pochhammer symbols are analytic at $q = 1/2$. Thus, this regular part has at most a double pole at $q = 1/2$. A simple computation shows that the coefficient of $(1-2q)^{-2}$ reduces algebraically trivially – in the sense that no q-identity is involved – to 0. Thus, the regular part involves only a simple pole at $q = 1/2$, as is expressed by the first two terms of the expansion (3.91), where $R(q)$ is analytic at $q = 1/2$. (Note that the coefficient of $(1-2q)^{-1}$ is exactly 1, again for trivial reasons.)

The singular part is more interesting and it can be analysed by the method suggested at the beginning of this subsection. Whenever convenient, we freely use the abbreviation $t = 1 - 2q$. The function $\gamma(q) = (\log v)/(\log 1/q)$ is analytic at $q = 1/2$, where

$$\gamma(q) = \log_2 \frac{3}{2} + 2 \frac{\log 3}{(\log 2)^2} (q - \tfrac{1}{2}) + \cdots$$

$$\approx 0.58496 + 4.5732(q - \tfrac{1}{2}) + 16.317(q - \tfrac{1}{2})^2 + 39.982(q - \tfrac{1}{2})^3 + 86.991(q - \tfrac{1}{2})^4 + \cdots$$

The function $(1-2q)^{-\gamma}$ can then be expanded as

$$(1-2q)^{-\gamma(q)} = (1-2q)^{-\gamma_0} e^{-(\gamma(q)-\gamma_0)\log t}, \qquad \text{with} \quad \gamma_0 = \gamma(1/2) = \log_2 \frac{3}{2}$$

$$= (1-2q)^{-\gamma_0} \left(1 + \frac{\log 3}{(\log 2)^2} t \log t + t^2 P_2(\log t) + t^3 P_3(\log t) + \cdots \right),$$

$$\quad (3.93)$$

for a computable family of polynomials $P_2, P_3, \ldots$, where $\deg P_\ell = \ell$ and $P_\ell(0) = 0$. For instance, we have, with $y := \log t$:

$$(1 - 2q)^{-(\gamma - \gamma_0)\log t} \approx 1 + 2.28ty + t^2(-4.07y + 2.61y^2) + t^3(4.99y - 9.32y^2 + 1.99y^3) + \cdots$$

The singular series $U(q)$ of (3.83) is analytic at $q = 1/2$ and its coefficients can be determined, both numerically and, in principle, symbolically in terms of Pochhammer symbols and their logarithmic derivatives (which lead to q-analogues of harmonic numbers). Numerically, they can be estimated to high precision, by bounding the infinite sum and products to a finite but large value. (The validity of the process can be checked empirically by increasing the values of this threshold, the justification being that all involved sums and products converge geometrically fast – we found that replacing $+\infty$ by 100 in numerical computations provides estimates that are at least correct to 25 decimal digits.) In this way, we obtain, for instance, the expansion of the function $V(q)$, which is of the form

$$U(q) \approx \frac{16}{9\log 2} + 9.97\,t + 21.5\,t^2 + 35.8\,t^3 + 51.9\,t^4 + \cdots. \tag{3.94}$$

Finally, regarding $\Pi(u)$ taken at $u = \log_{1/q}(1 - 2q)$, we note that the coefficients p_k of (3.78) can be expanded around $q = 1/2$ and pose no difficulty, while the quantities $e^{2ik\pi u}$ can be expanded by a process analogous to (3.93). Indeed, we have

$$p_k \equiv p_k(q) = \frac{\pi}{\sin(\pi\gamma_0 + 2ik\pi^2/\log 2)} \cdot \exp\left(1 + e_1(k)t + e_2(k)t^2 + \cdots\right), \tag{3.95}$$

where the e_k only grow polynomially with k. Also, at $u = \log_{1/q}(1 - 2q)$, one has

$$e^{2ik\pi u} = (1 - 2q)^{2ik\pi/\log 2} \exp\left[2ik\pi\log_2 t\left(g_1 t + g_2 t^2 + \cdots\right)\right], \tag{3.96}$$

where the coefficients g_j are those of $(\log 1/q)^{-1} - (\log 2)^{-1}$ expanded at $q = 1/2$ and expressed in terms of $t = 1 - 2q$.

We can now recapitulate the results of the discussion of the singular part: from (3.94), (3.95), and (3.96), we find that the terms appearing in the singular expansion of $PA(q)$ are of the form, for $j = 0, 1, 2, \ldots$,

$$(1 - 2q)^{-\gamma_0 - 2} t^j (\log t)^\ell t^{2ik\pi/\log 2},$$

with ℓ such that $0 \le \ell \le j$ and $k \in \mathbb{Z}$. The terms at fixed j, ℓ add up to form the Fourier series $\Pi^{(j,\ell)}$, whose coefficients exhibit a fast decrease with $|k|$, similar to that encountered in (3.79). Consequently, a finite version of (3.90) at any order holds, so that the statement results. ∎

With this last proposition, we can conclude the proof of Theorem 3.19.

Proof of Theorem 3.19. The analytic term $R(q)$ in (3.90) leaves no trace in the asymptotic form of coefficients. Thus the global contribution of the *regular* part to coefficients pa_n reduces to 2^n (with coefficient 1 and no power of n), corresponding to the term $(1 - 2q)^{-1}$ in (3.90).

The transfer to coefficients of each term of the *singular* part of (3.90) is permissible, given the principles of singularity analysis recalled above. Only an amended form allowing for logarithmic factors is needed, but this is covered by the general theory: for the translation of the coefficients of the basic scale $(1-z)^{-\lambda}\log^k(1-z)$, see [44, Sec. VI.2]. From a computational point of view, one may conveniently operate [44, Note VI.7] with

$$[z^n](1-z)^{-\lambda}(\log(1-z))^k = (-1)^k\frac{\partial}{\partial\lambda}\frac{\Gamma(n+\lambda)}{\Gamma(\lambda)\Gamma(n+1)},$$

then replace the Gamma factors of large argument by their complete Stirling expansion.

We can now complete the proof of Theorem 3.19. It suffices to retain the terms corresponding to $j=0$ in (3.90), in which case the error term is of the form $O\left((1-2q)^{-\gamma_0-1}\log(1-2q)\right)$, which corresponds to a contribution that is $O(n^{\gamma_0}\log n)=O(n^{g-1}\log n)$ for pa_n.

Next, regarding the Fourier element of index $k=0$, the function-to-coefficient correspondence yields

$$(1-2q)^{-\gamma_0-2} \quad\Longrightarrow\quad \frac{n^{\gamma_0+1}}{\Gamma(\gamma_0+2)}\left(1+O\left(\frac{1}{n}\right)\right).$$

Thus, the coefficient x_0 in (3.44) has value (cf (3.87) and (3.88)) given by

$$x_0 = \xi_0\cdot p_0|_{q=1/2}\cdot\frac{1}{\Gamma(\gamma_0+2)}, \qquad \gamma_0 = \frac{\log 3/2}{\log 2},$$

with ξ_0 as in (3.88). This, given the form (3.78) of p_k at $k=0$, is equivalent to the value of x_0 stated in Theorem 3.19 (where $g := \gamma_0+1 = \log_2 3$).

For a Fourier element of index $k\in\mathbb{Z}$, we have similarly

$$(1-2q)^{-\gamma_0-2-i\chi_k} \quad\Longrightarrow\quad \frac{n^{\gamma_0+1+i\chi_k}}{\Gamma(\gamma_0+2+i\chi_k)}(1+O(1/n)), \qquad \text{where} \quad \chi_k := \frac{2k\pi}{\log 2}.$$

We finally observe that

$$n^{\gamma_0+1+i\chi_k} = n^{\gamma_0+1}e^{i\chi_k\log n},$$

so that all the terms, for $k\in\mathbb{Z}$, are of the same asymptotic order (namely, $O(n^{\gamma_0+1})$) and their sum constitutes a Fourier series in $\log n$. The Fourier coefficient x_k then satisfies, from the discussion above:

$$x_k = \xi_0\cdot p_k|_{q=1/2}\cdot\frac{1}{\Gamma(\gamma_0+2+i\chi_k)}.$$

Thus finally, with $g\equiv\gamma_0+1$:

$$x_k = \frac{\pi}{9\log 2\,\sin(\pi g+2ik\pi^2/\log 2)\Gamma(g+1+2ik\pi/\log 2)}\prod_{j=0}^{\infty}\frac{(1-\frac{1}{3}2^{-j})(1-\frac{3}{2}2^{-j})}{(1-\frac{1}{2}2^{-j})^2}. \tag{3.97}$$

This completes the proof of Theorem 3.19. ∎

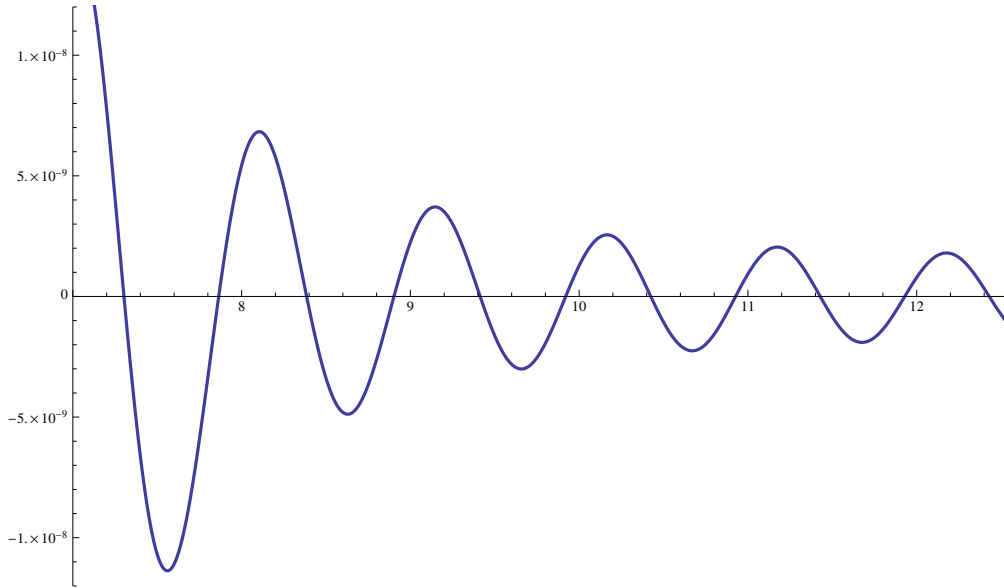The same method shows the existence of a complete asymptotic expansion for pa_n.

Figure 3.8: The difference $(pa_n - \Omega_6)2^{-n}n^{-g}$ against $\log_2 n$, where Ω_6 is the six-term extension of (3.98), for n up to 6000. The diagram confirms the presence of oscillations that, asymptotically, have amplitude of the order of 10^{-9}.

Theorem 3.29. *The number of 3-sided prudent polygons satisfies a complete asymptotic expansion,*

$$pa_n = 2^n + 2^n \cdot n^g \left(\Xi_{0,0} + \frac{1}{n} \left(\log n \cdot \Xi_{1,1} + \Xi_{1,0} \right) + \frac{1}{n^2} \left(\log^2 n \cdot \Xi_{2,2} + \log n \cdot \Xi_{2,1} + \Xi_{2,0} \right) \cdots \right),$$

where $\Xi_{j,\ell}$ is an absolutely convergent Fourier series in $\log n$.

The non-oscillating form obtained by retaining only the constant terms of each Fourier series is computed by a symbolic manipulation system such as MAPLE or MATHEMATICA in a matter of seconds and is found to start as

$$
\begin{aligned}
\frac{\Omega_5}{2^n} \approx\ & 1 + 0.1083842947 \cdot n^g \\
& + (-0.3928066917 L + 0.5442458535) \cdot n^{g-1} \\
& + \left(0.2627062704\, L^2 + 0.6950193894\, L + 0.6985601031 \right) \cdot n^{g-2} \\
& + \left(0.08310555463\, L^3 - 0.02188678892\, L^2 - 1.570478457\, L - 1.18810811075202 \right) \cdot n^{g-3} \\
& + \big(0.06722511293\, L^4 + 0.05494834609\, L^3 - 3.297513638\, L^2 - 4.663711650\, L \\
& \hspace{6cm} - 4.156441653 \big) \cdot n^{g-4}, \quad (3.98)
\end{aligned}
$$

where $L = \log n$. In principle, all the coefficients have explicit forms in terms of the basic quantities that appear in Theorem 3.19 (augmented by derivatives of q-Pochhammer symbols at small rational values). However, the corresponding formulae blow up exponentially, so that we only mention here the next coefficient $-0.39280\ldots$ in (3.98), whose exact value turns out to be

$$-x_0 \frac{\log 3}{\log^2 2}\, g.$$

85

Figure 3.8 displays the difference between pa_n and the six-term extension Ω_6 of (3.98). It is piquant to note that all the terms given in (3.98) are needed in order to succeed in bringing the fluctuations of Figure 3.8 out of the closet.

Conclusions

Classification of prudent polygons. Our first conclusion is that the present study permits us to advance the classification of prudent walks and polygons: the generating functions and their coefficient asymptotics are now known in all cases up to 3-sided (walks by length; polygons by either perimeter or area). Functional equations are also known for 4-sided prudent walks and polygons, from which it is possible to distill plausible estimates.

In particular, recall from Subsection 3.2.3 that we estimated that the number $pa_n^{(4)}$ of 4-sided prudent polygons of area n satisfies

$$pa_n^{(4)} \sim \alpha 2^n n^\beta$$

where $2.58 < \beta < 2.61$. In light of the results of this subsection, it seems plausible that the multiplier α might not be constant, and might instead have an oscillating nature similar to that found in the 3-sided case. Moreover, it is tempting to conjecture that the exponent β is precisely one greater than that of the 3-sided model, that is, $\beta = 1 + \log_2 3 = 2.5849625\ldots$.

We have investigated this conjecture with some success. First, we considered the Hadamard quotient of the series for 4-sided polygons and that of the derivative of the series for 3-sided polygons. (Differentiation increases the exponent by 1.) If the conjecture is true, the coefficients of the Hadamard quotient should tend to a constant. Extrapolating the quotient

$$h_n = n \frac{pa_n^{(3)}}{pa_n^{(4)}},$$

we find a limit of 3.25 ± 0.05. Next, we tested the proposal that the asymptotics for the 4-sided case are given by the derivative of the 3-sided case. We fitted successive quartets of coefficients $a_j, a_{j+1}, a_{j+2}, a_{j+3}$ to

$$pa_n^{(4)} = 2^n(x n^g + x_1 n^2 + x_2 n^{g-1} \log n + x_3 n^{g-1})$$

where $g = \log_2 3 + 1$. Estimates of x are well-converged to ≈ 0.03341, implying that the "amplitude" of the 3-sided polygons should be ≈ 0.108, which agrees well with the exact value of $0.1083842947\ldots$ calculated above. On balance, we believe the "conjecture" is more likely to be true than not.

We can then summarise the present state of knowledge with Table 3.2 (we mark conjectural results with a (?)).

(The oscillating coefficient $C_3(n)$ is expressible in terms the Fourier series $x(u)$ of Theorem 3.19.)

Table 3.2: Summary of known results for enumerations of prudent polygons by area.

Prudent polygons	Generating function	Asymptotic number	References
2-sided	rational	2^n	Proposition 3.14
3-sided	non-holonomic	$C_3(n) \cdot 2^n n^{\log_2 3}$	Theorem 3.19
4-sided	functional equation	$C_4(n) \cdot 2^n n^{1+\log_2 3}$ (?)	Proposition 3.18.

Methodology. The generating function of 3-prudent polygons has been found to be a q-hypergeometric function, with the argument and parameters subject to a rational substitution. The methods developed here should clearly be useful in a number of similar situations. Note that the asymptotic enumeration of prudent walks and polygons by length and perimeter is in a way easier since the *dominant* singularity is polar or algebraic. (Bousquet-Mélou [17] however exhibits an interesting situation where the *complete* singular structure has a complex geometry.)

Estimates involving periodic oscillations are not unheard of in combinatorial asymptotics [44, 91, 92]. What is especially interesting in the case of 3-sided prudent polygons is the pattern of singularities that accumulate geometrically fast to the dominant singularity. This situation is prototypically encountered in the already evoked problem of the longest run in strings: the classical treatment is via real analytic methods followed by a Mellin analysis of the expressions obtained; see [79]. In fact, the chain

$$\text{Coefficient asymptotics} \rightsquigarrow \text{Mellin transform} + \text{Singularity analysis} \qquad (3.99)$$

is applicable for moment analyses. For instance, the analysis of the expected longest run of the letter a in a random binary sequence over the alphabet $\{a, b\}$ leads to the generating function [44, Ex. V.4]

$$\Phi(z) = (1-z) \sum_{k \geq 0} \frac{z^k}{1 - 2z + z^{k+1}},$$

to which the chain (3.99) can be applied.

Another source of similar phenomena is the analysis of digital trees [82, 108], when these are approached via ordinary generating functions (rather than the customary exponential or Poisson generating functions). Typically, in the simplest case of node-depth in a random digital tree, one encounters the generating function

$$\Psi(z) = \frac{1}{1-z} \sum_{k \geq 0} \frac{2^{-k}}{1 - z(1 - 2^{-k})},$$

where the geometric accumulation of poles towards 1 is transparent, so that the chain (3.99) can once more be applied [43].

We should finally mention that "critical" exponents similar to $g = \log_2 3$ surface at several places in mathematics, usually accompanied by oscillation phenomena, but they do so for reasons essentially simpler than in our chain (3.99). For instance, in fractal geometry, the Hausdorff

dimension of the triadic Cantor set is $1/g$, see [38], while that of the familiar Sierpiński gasket is g, so that g occurs as critical exponent in various related integer sequences [42]. The same exponent $g = \log_2 3 \approx 1.58$ is otherwise known to most students of computer science, since it appears associated with the recurrent sequence $f_n = n + 3f_{\lfloor n/2 \rfloor}$, which serves to describe the complexity of Karatsuba multiplication [81] (where, recursively, the product of *two* double-precision numbers is reduced to *three* single-precision numbers). In such cases, the exponent g is eventually to be traced to the singularity (at $s = g$, precisely) of the Dirichlet series

$$\omega(s) = \frac{1}{1 - 3 \cdot 2^{-s}},$$

which is itself closely related to the Mellin transforms of (3.75). See also the studies by Drmota and Szpankowski [30], Dumas [32], as well as [42] for elements of a general theory.

Comparison with other SAW and SAP models. We know of no other SAW or SAP models, solvable or otherwise, which display oscillatory behaviour resembling (3.98). (Excepting rather trivial examples, like the fact that a SAP on the square or honeycomb lattice must have even perimeter.) We were only able to detect the oscillating component of $pa_n^{(3)}$ after solving and analysing the generating function; if another model displayed similar behaviour, it is highly unlikely that the oscillations could be detected from numerical data alone, and so one would need to be able to solve the generating function in order to state definitively that oscillatory behaviour were present.

At this stage the reason for the oscillatory behaviour of $pa_n^{(3)}$ is a mystery to us, and for this reason we are reluctant to suggest that any other SAW or SAP models might display such behaviour. The one exception is 4-sided prudent polygons, whose asymptotic form (see Table 3.2 above) appears to be very closely related to the 3-sided case. We are especially hesitant to make any conjectures about general SAW and SAP models, as there are already a number of well-supported conjectures regarding those models which do not feature oscillatory behaviour. (See Chapter 1 for some further discussion.) In particular, in five or more dimensions it has been proven (again, refer to Chapter 1) that

$$c_n \sim A\mu^n$$

directly contradicting the possiblity of oscillatory behaviour.

3.3 Other square lattice models

We saw in the last section that prudent walks are, in a sense, a generalisation of partially directed walks – the latter can be considered as 1-sided prudent walks, and these can then be extended to 2-sided and so on. In this section we will examine some further generalisations, namely *perimeter walks* and *quasi-prudent* walks, so that we have a chain which looks like

$$directed \subset prudent \subset perimeter \subset quasi\text{-}prudent.^{12}$$

We will also briefly mention some other solvable models which display interesting behaviour.

3.3.1 Perimeter walks and polygons

Recall that the constructions we used for prudent walks in Subsection 3.2.1 used the property that the endpoint of any prudent walk always lies on the boundary of its bounding box (the smallest rectangle on the lattice which contains the entire walk). This property suggests that we can define a new class of walks which are characterised entirely by this restriction, without necessarily being prudent. That is, a *perimeter walk* is a SAW which, after every step, ends on the boundary of the current bounding box. This definition leads to the same sub-classification we used for prudent walks:

- A *1-sided perimeter walk* must, after every step, end on the east side of its current bounding box. These are the same as 1-sided prudent walks, or equivalently NES-directed walks.

- A *2-sided perimeter walk* must, after every step, end on the east or north side of its current bounding box.

- A *3-sided perimeter walk* must, after every step, end on the east, north or west side of its current bounding box, and in addition cannot step from the south-east to south-west corner (or vice versa) when the box has width one.

- A *4-sided* or *unrestricted perimeter walk* may end on any side of its bounding box.

See Figure 3.9 for examples of 2-, 3- and 4-sided perimeter walks.

Analogously to the prudent case, we define a *k-sided perimeter polygon* to be a *k*-sided perimeter walk which ends at a vertex adjacent to the origin.

2-sided perimeter walks

The difficulty of solving perimeter walks, in comparison to their prudent counterparts, becomes clear as soon as we consider the 2-sided case. Here, we will generate the walks in the same way as for prudent walks – by considering the last inflating step (LIS), which must be east or north. However, we will need to use two catalytic variables, and thus find that the 2-sided walks are the most general subclass of perimeter walks that we can solve exactly.

---

[12]If we define perimeter walks in three dimensions in the obvious way (i.e. the endpoint of a walk must lie on the smallest rectangular prism containing the entire walk), then this hierarchy breaks down – there exist prudent walks in three dimensions which are not perimeter walks. Numerical calculations by Garoni [9] suggest the growth rate for three-dimensional prudent walks is about 4.4913, while for perimeter walks it is only 4.33.
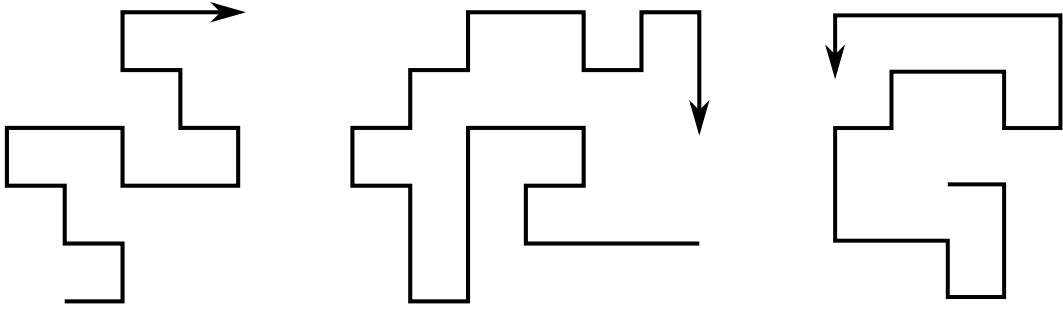
Figure 3.9: 2-sided, 3-sided and 4-sided (unrestricted) perimeter walks on the square lattice.
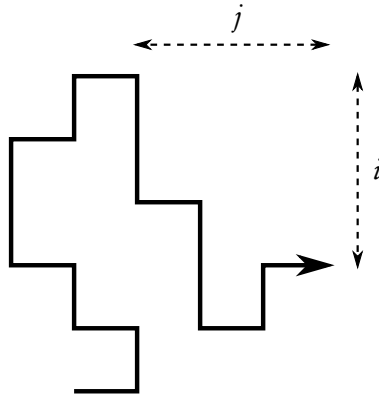
Figure 3.10: A 2-sided perimeter walk with the distances i and j measured by the variables u and v indicated. Note that a walk which ends at the north-east corner of its box will have $i = j = 0$.

Since we have a reflective symmetry in the line $y = x$, we only need to count walks ending on the east side of their bounding box. So define the generating function

$$R(t; u, v) = \sum_{n,i,j \geq 0} r_{n,i,j} t^n u^i v^j$$

where $r_{n,i,j}$ is the number of 2-sided perimeter walks of length n which end on the east side of their bounding box, with distance i from the endpoint to the north-east corner of the bounding box and distance j between the north-east corner and the nearest occupied vertex on the north side of the box. (If a walk ends at the north-east corner of its box, it will have $i = j = 0$.) See Figure 3.10 for an illustration.

Lemma 3.30. *The generating function $R(t; u, v)$ satisfies the functional equation*

$$\left(1 - tv - \frac{t^2 v}{u - t} - \frac{t^2 uv}{1 - tu}\right) R(t; u, v) = \frac{1}{1 - tu} - \frac{tuv}{u - t} R(t; t, v)$$

$$+ \frac{t(2 - tu)}{1 - tu} R(t; t, 1) - \frac{t^2 u}{1 - tu} R(t; t, tu) + \frac{t^2 u(1 - v)}{1 - tu} R(t; 0, 1). \quad (3.100)$$

Proof. There are three cases to consider:

- **No LIS:** The walk must be empty or contain only south steps; the generating function is thus

$$\frac{1}{1-tu}. \tag{3.101}$$

- **LIS east:** After the inflating step, the walk can take up to i steps north or unbounded steps south. If it does not step all the way to the north-east corner, j increases by one; if it does then j becomes zero. The generating function is then

$$\sum_{n,i,j\geq 0} r_{n,i,j} t^n u^i v^j \cdot tv \left(1 + \sum_{l=1}^{i-1} t^l u^{-l} + t^i u^{-i} v^{-j-1} + \sum_{l=1}^{\infty} t^l u^l\right)$$

$$= t \sum_{n,i,j\geq 0} r_{n,i,j} t^n \left(u^i v^{j+1} + \frac{v^{j+1}(tu^i - ut^i)}{u - t} + t^i + \frac{t u^{i+1} v^{j+1}}{1 - tu}\right)$$

$$= tv R(t;u,v) + \frac{tv}{u-t}(t R(t;u,v) - u R(t;t,v)) + t R(t;t,1) + \frac{t^2 uv}{1-tu} R(t;u,v). \tag{3.102}$$

However, the above fails in the case that $i = 0$, i.e. the inflating step is from the north-east corner of the box. (In that case j does not increase.) We correct this by subtracting off the offending terms,

$$-tv R(t;0,v) - \frac{tv}{u-t}(t R(t;0,v) - u R(t;0,v)) - t R(t;0,1) - \frac{t^2 uv}{1-tu} R(t;0,v), \tag{3.103}$$

and then adding back on the corrected versions (without the extra factor of v):

$$t R(t;0,v) + \frac{t^2 u}{1-tu} R(t;0,v). \tag{3.104}$$

We note here that $r_{n,0,j} \neq 0$ only when $j = 0$, and so $R(t;0,v)$ has no dependence on v. Thus we can replace any $R(t;0,v)$ term in (3.103) and (3.104) with $R(t;0,1)$.

- **LIS north:** After the inflating step, the walk must step east to the north-east corner of the box. If the inflating step was not from the north-east corner, the walk may be able to then take south steps down the east side of the box (while remaining self-avoiding, of course). The generating function is thus

$$\sum_{n,i,j\geq 0} r_{n,i,j} t^n \cdot t^{i+1} \sum_{l=0}^{j} t^l u^l$$

$$= t \sum_{n,i,j\geq 0} r_{n,i,j} t^{n+i} \frac{1 - (tu)^{j+1}}{1 - tu}$$

$$= \frac{t}{1-tu}(R(t;t,1) - tu R(t;t,tu)). \tag{3.105}$$

Now every 2-sided perimeter walk ending on the east side of its bounding box is covered by exactly one of the above three cases; so by adding (3.101)–(3.105) we obtain $R(t;u,v)$. Rearranging then gives (3.100). ∎

Setting $u = 0$ and $v = 1$ in (3.100) gives

$$R(t; 0, 1) = 1 + 2t R(t; t, 1),$$ \hfill (3.106)

which can then be substituted into (3.100):

$$\left(1 - tv - \frac{t^2 v}{u - t} - \frac{t^2 u v}{1 - tu}\right) R(t; u, v) = \frac{1 + t^2 u - t^2 u v}{1 - tu} - \frac{tuv}{u - t} R(t; t, v)$$

$$- \frac{t^2 u}{1 - tu} R(t; t, tu) + \frac{t(2 - tu + 2t^2 u - 2t^2 u v)}{1 - tu} R(t; t, 1). \quad (3.107)$$

From here we will apply the iterated kernel method to determine the solution to $R(t; 1, 1)$. There are some intermediate steps, so we split the overall result into several parts.

Lemma 3.31. *Define*

$$S \equiv S(t; v) = \frac{1 - tv + t^2 + t^3 v - \sqrt{(1 - tv + t^2 + t^3 v)^2 - 4t^2}}{2t}$$

and then, recursively,

$$S_n \equiv S_n(t; v) = S_{n-1}(t; t S(t; v)) = S(t; t S_{n-1}(t; v))$$

with $S_0 = v, S_1 = t S(t; v)$, etc. Then define

$$A_n(t; v) = \frac{(S_{n+1} - t^2)(1 + t S_{n+1} - t S_n S_{n+1})}{t S_n S_{n+1}(1 - S_{n+1})},$$

$$B_n(t; v) = \frac{-(S_{n+1} - t^2)}{S_n(1 - S_{n+1})},$$

$$C_n(t; v) = \frac{(S_{n+1} - t^2)(2 - S_{n+1} + 2t S_{n+1} - 2t S_n S_{n+1})}{S_n S_{n+1}(1 - S_{n+1})}.$$

Then the generating function $R(t; t, 1)$ has the solution

$$R(t; t, 1) = \frac{\displaystyle\sum_{n=0}^{\infty} A_n(t; 1) \prod_{k=0}^{n-1} B_k(t; 1)}{1 - \displaystyle\sum_{n=0}^{\infty} C_n(t; 1) \prod_{k=0}^{n-1} B_k(t; 1)}. \hfill (3.108)$$

Proof. Define the kernel

$$K(t; u, v) = 1 - tv - \frac{t^2 v}{u - t} - \frac{t^2 u v}{1 - tu}.$$

The equation $K(t; u, v) = 0$ has one root in v and two in u, but only one of these, $u = S$, is a power series in t. Substituting $u = S$ into (3.107) gives

$$R(t; t, v) = \frac{S - t}{tvS}\left(\frac{1 + t^2 S - t^2 v S}{1 - tS} - \frac{t^2 S}{1 - tS} R(t; t, tS) + \frac{t(2 - tS + 2t^2 S - 2t^2 Sv)}{1 - tS} R(t; t, 1)\right),$$

\hfill (3.109)

or equivalently,

$$R(t; t, S_0) = A_0 + B_0 R(t; t, S_1) + C_0 R(t; t, 1). \tag{3.110}$$

Sending $v \mapsto S_n$ gives

$$R(t; t, S_n) = A_n + B_n R(t; t, S_{n+1}) + C_n R(t; t, 1), \tag{3.111}$$

and with back-substitution we arrive at

$$R(t; t, S_0) = \sum_{n=0}^{N} A_n \prod_{k=0}^{n-1} B_k + \prod_{n=0}^{N} B_n \cdot R(t; t, S_{N+1}) + \sum_{n=0}^{N} C_n \prod_{k=0}^{n-1} B_k \cdot R(t; t, 1). \tag{3.112}$$

If we define $\bar{S}_n \equiv \bar{S}_n(t) = S_n(t; 1)$, then setting $v = 1$ in (3.112) gives

$$R(t; t, 1) = \sum_{n=0}^{N} A_n(t; 1) \prod_{k=0}^{n-1} B_k(t; 1) + \prod_{n=0}^{N} B_n(t; 1) \cdot R(t; t, \bar{S}_{N+1})$$
$$+ \sum_{n=0}^{N} C_n(t; 1) \prod_{k=0}^{n-1} B_k(t; 1) \cdot R(t; t, 1). \tag{3.113}$$

Now, we wish to take the limit $N \to \infty$. All the terms in (3.113) are power series in t, with $A_n(t; v) = 1 + t^2 + O(t^3)$, $B_n(t; v) = -t^3 + O(t^4)$ and $C_n(t; v) = 2t + t^3 + O(t^4)$, whenever v is a power series in t. In addition, $\bar{S}_N$ converges to a fixed power series as $N \to \infty$, namely $t^2 + t^5 + 2t^8 + t^{10} + O(t^{11})$. So $R(t; t, \bar{S}_{N+1})$ converges to a fixed power series as $N \to \infty$, namely $1 + 2t + 5t^2 + 27t^4 + 64t^5 + O(t^6)$.

Hence we see that the first and third terms on the RHS of (3.113) converge to fixed power series in t as $N \to \infty$, while the second term disappears. So we obtain

$$R(t; t, 1) = \sum_{n=0}^{\infty} A_n(t; 1) \prod_{k=0}^{n-1} B_k(t; 1) + \sum_{n=0}^{\infty} C_n(t; 1) \prod_{k=0}^{n-1} B_k(t; 1) \cdot R(t; t, 1), \tag{3.114}$$

and the result of the lemma follows. ∎

Note that in the proof above we only consider the functions as formal power series in t – we have not yet taken radii of convergence, etc. into account. We will finish by writing down the solution to $R(t; 1, 1)$ before considering the singularities of the generating functions.

Theorem 3.32. *Define A_n, B_n and C_n as in Lemma 3.31. Then the generating function $R(t; 1, 1)$ has the solution*

$$R(t; 1, 1) = \frac{1}{1 - 2t - t^2} \left[1 - t^2 \sum_{n=0}^{\infty} A_n(t; t) \prod_{k=0}^{n-1} B_k(t; t) \right.$$
$$\left. + \left(t - t^2 - t^2 \sum_{n=0}^{\infty} C_n(t; t) \prod_{k=0}^{n-1} B_k(t; t) \right) R(t; t, 1) \right]. \tag{3.115}$$

Proof. Define $\tilde{S}_n \equiv \tilde{S}_n(t) = S_n(t; t)$. Setting $v = t$ in (3.112) gives

$$R(t; t, t) = \sum_{n=0}^{N} A_n(t; t) \prod_{k=0}^{n-1} B_k(t; t) + \prod_{n=0}^{N} B_n(t; t) \cdot R(t; t, \tilde{S}_{N+1})$$

$$+ \sum_{n=0}^{N} C_n(t; t) \prod_{k=0}^{n-1} B_k(t; t) \cdot R(t; t, 1). \quad (3.116)$$

Similar arguments to those used in Lemma 3.31 provide for convergence as $N \to \infty$, and so we obtain

$$R(t; t, t) = \sum_{n=0}^{\infty} A_n(t; t) \prod_{k=0}^{n-1} B_k(t; t) + \sum_{n=0}^{\infty} C_n(t; t) \prod_{k=0}^{n-1} B_k(t; t) \cdot R(t; t, 1). \quad (3.117)$$

Setting $u = v = 1$ in (3.107) and substituting (3.117) gives the theorem. ∎

Combining walks which end on the east side of their bounding box with those ending on the north side and applying the inclusion-exclusion principle allows us to find the overall generating function of 2-sided perimeter walks.

Corollary 3.33. *The generating function for 2-sided perimeter walks is*

$$E^{(2)}(t) = \sum_{n \geq 0} e_n^{(2)} t^n = 2R(t; 1, 1) - R(t; 0, 1), \quad (3.118)$$

where $R(t; 0, 1)$ can be obtained from $R(t; t, 1)$ via (3.106).

From here, it is natural to wish to know the singular behaviour of $E^{(2)}(t)$, as it is this which determines the asymptotic behaviour of the coefficients $e_n^{(2)}$. Unfortunately, the structure of $E^{(2)}(t)$ is such that we have been unable to determine the exact value of the dominant singularity, though we can estimate it numerically with a high level of precision.

Conjecture 3.34. *The dominant singularity of $E^{(2)}(t)$ is a simple pole at $t = \rho \approx 0.399361698853$, with uncertainty confined to the last digit. The number of 2-sided perimeter polygons of length n is asymptotically*

$$e_n^{(2)} \sim A\tau^n$$

where $\tau = \rho^{-1} \approx 2.50399575841$ and $A \approx 2.27287$.

We are quite confident as to the truth of this conjecture, but given the fact that we can only obtain a numerical estimate, as well as the complexity of the generating functions, we will not make any serious attempt to prove its validity. The value of ρ appears to be the point at which

$$\sum_{n=0}^{\infty} C_n(t; 1) \prod_{k=0}^{n-1} B_k(t; 1) = 1. \quad (3.119)$$

It can be easily shown that the component functions $A_n(t; v), B_n(t; v)$ and $C_n(t; v)$ are analytic for $|t| < \sqrt{2} - 1$ and $|v| \leq 1$. To prove that ρ is indeed the dominant singularity of $E^{(2)}(t)$,

we also need to show that the various infinite sums featuring in its solution are convergent for $|t| < \rho'$ for some $\rho' > \rho$, and also that there is no other value of t with $|t| \leq \rho$ satisfying (3.119). The complexity of these functions seems to make proving facts like these quite difficult.

It seems very likely that the generating function $E^{(2)}(t)$ has countably many singularities in $\mathbb{C}$, and is thus not D-finite. For example, every $S_n(t; 1)$ term for $n \geq 0$ appears in the generating function, and these terms are singular when

$$t^2 S_{n-1}(t; 1) = \frac{1 \pm t}{1 \mp t}.$$

This equation will have different solutions for each n. However, these solutions do not lie on the interval $0 < t < \sqrt{2} - 1$, and thus the power series representation of $S_{n-1}(t; 1)$ is of limited use.

We note here that, while it seems impossible to generate 2-sided perimeter walks without using two extra variables (in addition to the length variable t), the kernel (3.100) of the resulting functional equation is quadratic in only one of those variables (and so, in a sense, that variable is the only "true" catalytic variable). This suggests that there may in fact be a simpler solution than the one we obtain above in Theorem 3.32. We have attempted to investigate this further, but thus far all alternative methods to the one described above have proven futile.

Based on Monte Carlo studies by Tim Garoni [9], it has been conjectured that the mean squared end-to-end distance of 2-sided perimeter walks of length n grows like $O(n^2)$, so that the exponent $\nu = 1$.

3-sided perimeter walks

Given the difficulty of solving 2-sided perimeter walks and then analysing their generating function, it should come as no surprise that we are unable to determine the exact solution of 3-sided perimeter walks. We will present here the functional equations satisfied by generating functions of 3-sided perimeter walks. We omit the proof, as the method used here is the same as for 2-sided walks.

Define the generating functions

$$T(t; u, v, w, x) = \sum_{n,i,j,k,l \geq 0} t_{n,i,j,k,l} t^n u^i v^j w^k x^l$$

$$R(t; u, v, w, x) = \sum_{n,i,j,k,l \geq 0} r_{n,i,j,k,l} t^n u^i v^j w^k x^l$$

where $t_{n,i,j,k,l}$ is the number of n-step 3-sided perimeter walks ending on the north side of their box, with

- distance i from the endpoint to the north-east corner of the box,

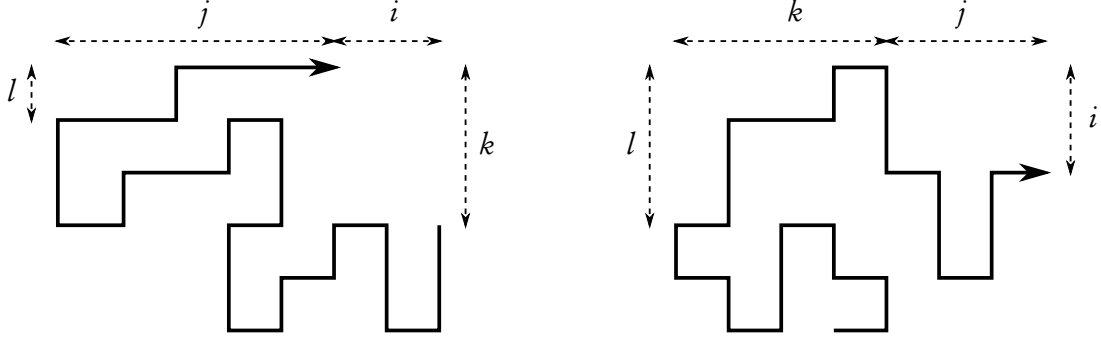- distance j from the endpoint to the north-west corner of the box,

Figure 3.11: 3-sided perimeter walks, ending on the north and east sides of their boxes, with the distances i, j, k, l measured by the catalytic variables u, v, w, x indicated.

- distance k from the north-east corner of the box to the nearest occupied vertex on the east side of the box, and

- distance l from the north-west corner of the box to the nearest occupied vertex on the west side of the box.

Similarly, $r_{n,i,j,k,l}$ is the number of n-step 3-sided perimeter walks ending on the east side of their box, with

- distance i from the endpoint to the north-east corner of the box,

- distance j from the north-east corner of the box to the nearest occupied vertex on the north side of the box,

- total width $j + k$, and

- distance l from the north-west corner of the box to the nearest occupied vertex on the west side of the box.

See Figure 3.11 for an illustration of these measurements for the two types of 3-sided perimeter walks.

Proposition 3.35. *The generating functions $T(t; u, v, w, x)$ and $R(t; u, v, w, x)$ satisfy the functional equations*

$$\left(1 - twx - \frac{t^2 vwx}{u - tv} - \frac{t^2 uwx}{v - tu}\right) T(t; u, v, w, x) = 1 - \frac{t(w-1)(x-1)}{1-t} - \frac{tuwx}{u - tv} T(t; tv, v, w, x)$$

$$+ tx T(t; tv, v, 1, x) - \frac{tvwx}{v - tu} T(t; u, tu, w, x) + tw T(t; u, tu, w, 1) - \frac{t^2 ux(w-1)}{v - tu} T(t; 0, v, 1, x)$$

$$+ \frac{tvx(w-1)}{v - tu} T(t; 0, tu, 1, x) - t(w-1) T(t; 0, tu, 1, 1) - \frac{t^2 vw(x-1)}{u - tv} T(t; u, 0, w, 1)$$

$$+ \frac{tuw(x-1)}{u - tv} T(t; tv, 0, w, 1) - t(x-1) T(t; tv, 0, 1, 1) + \frac{tv^2}{v - tu} R(t; t, v, v, x)$$

$$- \frac{t^2 uv}{v - tu} R(t; t, tu, v, x) + \frac{tu^2}{u - tv} R(t; t, u, u, w) - \frac{t^2 uv}{u - tv} R(t; t, tu, u, w) \quad (3.120)$$

96

and

$$\left(1 - tv - \frac{t^2 v}{u - t} - \frac{t^2 uv}{1 - tu}\right) R(t; u, v, w, x) = \frac{1}{1 - tu} - \frac{tuv}{u - t} R(t; t, v, w, x) + twR(t; t, w, w, x)$$

$$- \frac{t^2 u(v - w)}{1 - tu} R(t; 0, 1, w, x) + \frac{tx}{1 - tu} T(t; tw, w, 1, x) - \frac{t^2 ux}{1 - tu} T(t; tw, w, tu, x)$$

$$- \frac{t(x - 1)}{1 - tu} T(t; tw, 0, 1, 1) + \frac{t^2 u(x - 1)}{1 - tu} T(t; tw, 0, tu, 1). \quad (3.121)$$

The generating function for 3-sided perimeter walks is then

$$E^{(3)}(t) = T(t; 1, 1, 1, 1) + 2R(t; 1, 1, 1, 1) - 2R(t; 0, 1, 1, 1).$$

While we are not able to solve these equations, they do enable us to generate relatively long series in polynomial time. A numerical analysis strongly suggests that the dominant singularity of $E^{(3)}(t)$ is at the same point as $E^{(2)}(t)$ (i.e. at $t = \rho \approx 0.399361698853$), and is again a simple pole. We estimate the amplitude to be approximately 6.33.

4-sided perimeter walks

It is possible to write down a functional equation for 4-sided perimeter walks, but the generating functions involved would use even more catalytic variables than the 3-sided case, and so realistically we have no hope of solving any such equation. For this reason we will only mention numerical results.

As with 2-sided and 3-sided perimeter walks, the dominant singularity of unrestricted perimeter walks appears to be a simple pole at $t = \rho \approx 0.399361698853$. The amplitude is expected to be approximately 16.12. Monte Carlo studies by Garoni [9] suggest that the mean squared end-to-end distance of unrestricted perimeter walks grows like $O(n^2)$.

2-sided perimeter polygons by perimeter

Since 1-sided perimeter polygons are the same as the trivial case of 1-sided prudent polygons, we now turn our attention to 2-sided perimeter polygons, enumerated by half-perimeter. Such objects are either prudent or not; since we know the half-perimeter generating function of 2-sided prudent polygons (Lemma 3.11), we only need to consider those polygons which contain a non-prudent step.

We will consider the generating function

$$C(t; u) = \sum_{n, i \geq 0} c_{n, i} t^n u^i,$$

where $c_{n, i}$ is the number of 2-sided perimeter polygons ending at $(0, 1)$ in a counter-clockwise direction which contain at least one non-prudent step, with half-perimeter n and rightmost column of height $i - 1$. (Equivalently, i is the distance from the north-east corner of the bounding box to the *farthest* occupied vertex on the east side of the box.) In Figure 3.12 we provide an illustration of a 2-sided perimeter polygon together with the distance i.

Figure 3.12: A 2-sided perimeter polygon, as counted by $C(t; u)$, with the distance i measured by the catalytic variable u indicated.

Lemma 3.36. *The generating function $C(t; u)$ satisfies the functional equation*

$$\left(1 + \frac{tu}{1-u} - \frac{t^2 u}{1-tu}\right) C(t; u) = \frac{tu^2}{1-u}\widehat{B}(t; 1) - \frac{tu(1-t)}{(1-u)(1-tu)}\widehat{B}(t; u) + \frac{tu^2}{1-u}C(t; 1), \quad (3.122)$$

where

$$\widehat{B}(t; u) = \sum_{n,i \geq 0} \widehat{b}_{n,i} t^n u^i$$

is the generating function of bargraphs with rightmost column of height at least 2, with t conjugate to half-perimeter and u conjugate to the height of the last column.

Proof. We begin by noting that with only minor modifications it is easy to use Schwerdtfeger's method [105] to show that

$$\widehat{B}(t; u) = \frac{t^2 u^2 (1 - t - tu + t^2)}{1 - u + tu^2 - t^2 u} \cdot \frac{V - 1}{1 - tV} - \frac{t^3 u^2 (u - t)}{1 - u + tu^2 - t^2 u}, \quad (3.123)$$

where

$$V \equiv V(t) = \frac{1 + t^2 - \sqrt{(1-t)(1 - 3t - t^2 - t^3)}}{2t}$$

is the power series root of $1 - V + tV - t^2 V = 0$.

The only possibility for a non-prudent step for this class of polygons is a step west towards the origin. After some number of such steps, the walk can take a north step, and can then only take west steps to the final point $(0, 1)$. So these polygons can be viewed as bargraphs with a piece (i.e. a row of cells) taken out of the north-east corner. We construct them by adding columns to the right side of other polygons; in particular, we can add a column to a bargraph with rightmost column of height at least 2, or to an existing perimeter polygon. For the former case, the generating function is given by

$$\sum_{n,i \geq 0} \widehat{b}_{n,i} t^n \cdot t \left(\sum_{l=2}^{i} u^l + u^i \sum_{l=1}^{\infty} u^l t^l \right)$$

$$= t \sum_{n,i \geq 0} \widehat{b}_{n,i} t^n \left(\frac{u(u - u^i)}{1 - u} + \frac{t u^{i+1}}{1 - t u} \right)$$

$$= \frac{t u}{1 - u} [u \widehat{B}(t; 1) - \widehat{B}(t; u)] + \frac{t^2 u}{1 - t u} \widehat{B}(t; u) \tag{3.124}$$

Using exactly the same method, the polygons generated by adding a column to an existing perimeter polygon have the generating function

$$\frac{t u}{1 - u} [u C(t; 1) - C(t; u)] + \frac{t^2 u}{1 - t u} C(t; u). \tag{3.125}$$

Adding (3.124) and (3.125) gives $C(t; u)$, and the lemma follows. ∎

From here, the kernel method can be applied to (3.122).

Theorem 3.37. *The generating function $C(t; 1)$ has the solution*

$$C(t; 1) = \frac{t^3 V^2 (1 - t)^2}{(1 - t V)(1 + t^2 - 2 t V)} + t^2 - \frac{t(1 - t)(V - 1)}{1 - t V}, \tag{3.126}$$

where $V \equiv V(t)$ is as defined in the proof of Lemma 3.36. The half-perimeter generating function for 2-sided perimeter polygons is then

$$EP^{(2)}(t) = 2C(t; 1) + 4B(t)$$

$$= 1 - 4t + t^2 + \frac{(-1 + 5t + t^2 + t^3)\sqrt{1 - t}}{\sqrt{1 - 3t - t^2 - t^3}} \tag{3.127}$$

$$= 4t^2 + 8t^3 + 22t^4 + 64t^5 + 192t^6 + 588t^7 + 1828t^8 + \cdots$$

where $B(t) = t(V - 1)/(1 - t V)$ is the half-perimeter generating function of bargraphs.

Proof. The substitution $u = V$ cancels the kernel on the LHS of (3.122), enabling us to write $C(t; 1)$ in terms of $\widehat{B}(t; 1)$ and $\widehat{B}(t; V)$. Simplifying then gives (3.126).

The generating function of 2-sided perimeter polygons ending at $(0, 1)$ is $C(t; 1) + 2B(t)$, as we have to account for the prudent polygons ending at $(0, 1)$ in a counter-clockwise direction, as well as the polygons which walk clockwise. By Schwerdtfeger [105] the former class is counted by $B(t)$; it is not hard to see that the polygons in the latter class are just vertically-oriented bargraphs. We finally multiply by two to account for those polygons ending at $(1, 0)$. ∎

Corollary 3.38. *The dominant singularity of $EP^{(2)}(t)$ is a pole of order $1/2$ at $t = \sigma \approx 0.296$, a root of $1 - 3\sigma - \sigma^2 - \sigma^3 = 0$. Thus the number of 2-sided perimeter polygons of half-perimeter n is asymptotically*

$$e p_n^{(2)} \sim \alpha \lambda^n n^{-1/2} \tag{3.128}$$

where $\lambda = \sigma^{-1} \approx 3.38$ and $\alpha \approx 0.2623$.

Note that the location of the singularity, at $t = \sigma$, is the same point as for 2-sided prudent polygons (see Lemma 3.11), but with a different exponent.

3-sided perimeter polygons by perimeter

Unsurprisingly, the progression from 2-sided to 3-sided perimeter polygons brings many of the same problems as for perimeter walks. In particular, we can derive functional equations for the generating functions, but these require at least three catalytic variables. The complexity of this class of polygons seems to require the use of several sets of generating functions. For illustrative purposes we present the functional equations for one such set, namely those polygons which end at $(-1, 0)$ in a counter-clockwise direction and contain the north-west corner of their bounding box.

We define two classes of these polygons: $\mathcal{X}$ contains polygons for which the removal of the top row does not change the width of the polygon or leave two or more disconnected pieces, and $\mathcal{Y}$ contains those polygons not in $\mathcal{X}$. These two classes have the generating functions

$$X(t; w, v, x) = \sum_{n,i,j,k \geq 0} x_{n,i,j,k} t^n w^i v^j x^k$$

and

$$Y(t; w, v, x) = \sum_{n,i,j,k \geq 0} y_{n,i,j,k} t^n w^k v^j x^k$$

where $x_{n,i,j,k}$ is the number of 3-sided perimeter polygons in $\mathcal{X}$, with

- half-perimeter n,

- distance i from the north-west corner of the box to the farthest occupied vertex on the north side of the box,

- total width $i + j$, and

- distance k from the north-east corner of the box to the nearest occupied vertex on the east side of the box.

Similarly, $y_{n,i,j,k}$ is the number of 3-sided perimeter polygons in $\mathcal{Y}$, with

- half-perimeter n,

- distance i from the north-west corner of the box to the farthest occupied vertex on the north side of the box,
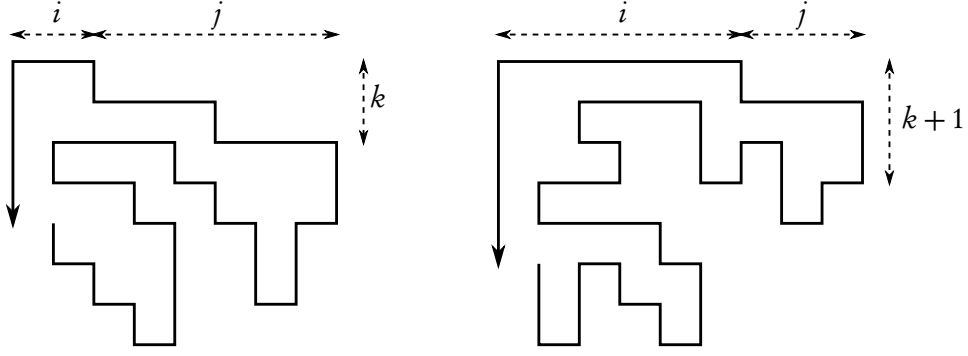
- total width $i + j$, and

Figure 3.13: 3-sided perimeter polygons, in $\mathcal{X}$ and $\mathcal{Y}$ respectively, with the distances i, j, k measured by the catalytic variables w, v, x indicated.

- distance $k+1$ from the north-east corner of the box to the farthest occupied vertex on the east side of the box.

See Figure 3.13 for examples of polygons in $\mathcal{X}$ and $\mathcal{Y}$, together with the distances measured by the catalytic variables.

Proposition 3.39. *The generating functions $X(t; w, v, x)$ and $Y(t; w, v, x)$ satisfy the functional equations*

$$\left(1 + \frac{twx}{v-w} + \frac{t^2wx}{v-tw}\right)X(t; w, v, x) = \frac{twx}{v-w}X(t; v, v, x) - \frac{tvx}{v-tw}X(t; w, tw, x)$$

$$+ tX(t; w, tw, 1) + \frac{twx^2}{v-w}Y(t; v, v, 1) - \frac{twvx^2(1-t)}{(v-w)(v-tw)}Y(t; w, v, 1)$$

$$+ \frac{t(v-tw-vx^2)}{v-tw}Y(t; w, tw, 1) + \frac{twx(1-x)}{v-w}Y(t; v, 0, 1) - \frac{tvx(1-x)}{v-w}Y(t; w, 0, 1) \quad (3.129)$$

and

$$\left(1 + \frac{tvx}{1-x} - \frac{t^2vx}{1-tx}\right)Y(t; w, v, x) = t^2w + \frac{tvx}{1-x}Y(t; w, v, 1) + \frac{tw}{1-x}Y(t; w, 0, 1)$$

$$- \frac{twx(1-t)}{(1-x)(1-tx)}Y(t; w, 0, x) - \frac{t^2vx}{1-tx}Y(t; w, 0, 0) + \frac{t^2w}{1-tx}X(t; w, tw, 1)$$

$$- \frac{t^2w}{1-tx}X(t; w, tw, tx) + t^2wX(t; w, 0, 0) + t^2wY(t; w, tw, 1). \quad (3.130)$$

We omit the proof, but the recursive constructions work as follows:

- A polygon in $\mathcal{X}$ can be constructed by

    - adding a row of width i to the top of a polygon in $\mathcal{X}$ or $\mathcal{Y}$ of width $\geq i$.

- A polygon in $\mathcal{Y}$ is either a single square, or can be constructed by

    - adding a column to the right side of a polygon in $\mathcal{Y}$,

101

– adding a row of width i to a polygon in $\mathscr{X}$ of width i, plus a column of height ≥ 1 to the right hand side (while remaining self-avoiding, of course), or

– adding a row of width $i + 1$ to the top of a polygon in $\mathscr{Y}$ of width i.

While we are unable to solve these equations (or those satisfied by the other subclasses of 3-sided perimeter polygons), we can use them to generate relatively long (approx. 170 terms) series. A numerical analysis of these series then leads us to conjecture the asymptotic form of $e\,p_n^{(3)}$, the number of 3-sided perimeter polygons with half-perimeter n.

Conjecture 3.40. *The number of 3-sided perimeter polygons with half-perimeter n is asymptotically*

$$e\,p_n^{(3)} \sim x\omega^n n^{-1/2},$$

where $\omega \approx 4.10$ is the same growth rate as for 3-sided prudent polygons, as described in Lemma 3.12, and x is a positive constant.

4-sided perimeter polygons by perimeter

For the full, unrestricted case of 4-sided perimeter polygons, we can write down (very complicated) recursive constructions, and use these to generate series. Unfortunately we are only able to compute data for polygons of half-perimeter up to 25, which is not nearly large enough for us to obtain accurate estimates of growth rates or exponents. However, our (very rough) estimate for the exponential growth rate is 4.41, which along with the results for 2- and 3-sided polygons, leads us to tentatively conjecture that it is the same value as for 4-sided prudent polygons [47]. As for the exponent, the series is simply too short for us to obtain any reliable estimate.

2-sided perimeter polygons by area

As we did with prudent polygons, we now turn our attention to the enumeration of perimeter polygons by area. The case of 1-sided polygons is trivial (and the same as 1-sided prudent polygons), so we begin with 2-sided perimeter polygons. The interesting case is again those polygons ending at $(0, 1)$ in a counter-clockwise direction. The non-prudent polygons are then very easy to construct: we simply take any bargraph whose rightmost column has height at least two, and attach another bargraph on the right side, with the top of the second bargraph one unit lower than that of the first bargraph.

The area generating function of bargraphs is

$$B(q) = \frac{q}{1 - 2q},$$

and so the generating function of bargraphs with rightmost column of height at least two is

$$\widehat{B}(q) = \frac{q^2}{1 - 2q}.$$

So the area generating function of 2-sided perimeter (and not prudent) polygons ending at $(0, 1)$ in a counter-clockwise direction is

$$\widehat{B}(q)B(q) = \frac{q^3}{(1-2q)^2}.$$ (3.131)

Lemma 3.41. *The area generating function of 2-sided perimeter polygons is*

$$EA^{(2)}(q) = \sum_{n \geq 1} ea_n^{(2)} q^n = \frac{2q(2-4q+q^2)}{(1-2q)^2}$$ (3.132)

and so

$$ea_n^{(2)} = \begin{cases} 4 & n = 1 \\ (n+6)2^{n-2} & n \geq 2. \end{cases}$$

Proof. Polygons ending at $(0, 1)$ in a counter-clockwise direction are either prudent (and so have generating function $B(q)$), or are non-prudent and then have the generating function given in (3.131). Polygons ending at $(0, 1)$ clockwise are just vertically-oriented bargraphs, and so these also have generating function $B(q)$. By symmetry, polygons ending at $(1, 0)$ have the same generating function as those ending at $(0, 1)$. So the overall generating function is

$$2\left(\frac{q^3}{(1-2q)^2} + 2B(q)\right) = \frac{2q(2-4q+q^2)}{(1-2q)^2}.$$ ∎

3-sided perimeter polygons by area

The enumeration of 3-sided perimeter polygons by area is not as complicated as the enumeration by perimeter, but unfortunately we are still not able to solve the functional equations. We present here the functional equations for the same subset of 3-sided perimeter polygons we described earlier – namely, those which end at $(-1, 0)$ in a counter-clockwise direction and contain the north-west corner of their box. We subdivide these polygons slightly differently to before: $\mathcal{X}$ remains the same, and contains polygons for which the removal of the top row does not change the width of the polygon or leave two or more disconnected pieces; $\mathcal{Y}$ is now those polygons not in $\mathcal{X}$ which contain the north-east corner of their bounding box; and $\mathcal{Z}$ contains all polygons not in $\mathcal{X}$ or $\mathcal{Y}$.

Define the generating functions

$$X(q; w, v) = \sum_{n,i,j \geq 0} x_{n,i,j} q^n w^i v^j$$

$$Y(q; w, v) = \sum_{n,i,j \geq 0} y_{n,i,j} q^n w^i v^j$$

and

$$Z(q; w) = \sum_{n,i \geq 0} z_{n,i} q^n w^i$$

where $x_{n,i,j}$ is the number of 3-sided perimeter polygons in $\mathcal{X}$, with

- area n,

- width i, and

- distance j from the north-east corner of the box to the nearest occupied vertex on the east side of the box.

Similarly, $y_{n,i,j}$ is the number of 3-sided perimeter polygons in $\mathscr{Y}$, with

- area n,

- width i, and

- distance $k+1$ from the north-east corner of the box to the farthest occupied vertex on the east side of the box.

Finally, $z_{n,i}$ is the number of 3-sided perimeter polygons in $\mathscr{Z}$, with

- area n, and

- width i.

Proposition 3.42. *The generating functions $X(q;w,v), Y(q;w,v)$ and $Z(q;w)$ satisfy the equations*

$$\left(1 - \frac{qv}{1-q}\right) X(q;w,v) = X(q;qw,1) - \frac{v}{1-q} X(q;qw,v) + \frac{qv}{1-q} Y(q;w,1)$$
$$+ \frac{1-q-v}{1-q} Y(q;qw,1) + \frac{qv^2}{1-q} Z(q;w) + \frac{1-q-v^2}{1-q} Z(q;w) \quad (3.133)$$

$$Y(q;w,v) = qw + \frac{qw}{1-qv} X(q;qw,1) - \frac{qw}{1-qv} X(q;qw,qv) + qwX(q;qw,0)$$
$$+ qwY(q;qw,1) + qwZ(q;qw) + \frac{qw}{1-qv} Y(q;w,1) \quad (3.134)$$

and

$$\left(1 - \frac{qw}{1-q}\right) Z(q;w) = \frac{qw}{1-q} Y(q;w,1) - \frac{qw}{1-q} Y(q;w,0). \quad (3.135)$$

We again omit the proof, but the constructions work as follows:

- A polygon in $\mathscr{X}$ can be constructed by

  – adding a row of width i to the top of any polygon of width $\geq i$.

- A polygon in $\mathscr{Y}$ is either a single square, or can be constructed by

  – adding a column to the right side of a polygon in $\mathscr{Y}$,

- adding a row of width i to the top of a polygon in $\mathscr{X}$ of width i, plus a column of height ≥ 1 to the right side (while remaining self-avoiding), or

- adding a row of width $i+1$ to the top of a polygon in $\mathscr{Y}$ or $\mathscr{Z}$ of width i.

- A polygon in $\mathscr{Z}$ can be constructed by

  - adding a column to the right side of a polygon in $\mathscr{Y}$ or $\mathscr{Z}$, with the top of the new column one unit lower than the north side of the box.

We can use this recursive construction to generate a long series (800 terms) for 3-sided perimeter polygons by area. It is clear that the exponential growth rate is 2, just as it is for prudent polygons. Our best estimate for the exponent is 4.242, so that we have the approximate result

$$ae_n^{(3)} \sim x 2^n n^{4.242}$$

for some positive constant x. We note that while this estimate for the exponent is indistinguishable from $3\sqrt{2}$, we have no compelling reason to believe that to be the true value.

4-sided perimeter polygons by area

We have generated 44 terms in the sequence of 4-sided perimeter polygons, enumerated by area. We are confident that the growth rate is once again 2, but our estimate for the exponent is very imprecise. Our best estimate is that

$$ae_n^{(4)} \sim x 2^n n^{6.2}$$

for a positive constant x.

3.3.2 Quasi-prudent walks and polygons

As we mentioned at the start of this section, we can consider perimeter walks to be a generalisation of prudent walks, which are in turn a generalisation of directed walks. In this subsection we will briefly describe a further generalisation of all of these, namely *quasi-prudent walks*.

We believe that quasi-prudent walks were proposed by Jim Propp in response to a seminar on prudent walks. On the square lattice, these are SAWs for which, after every step, it is possible to draw a semi-infinite ray parallel to a lattice axis from the endpoint without intersecting the walk. Equivalently, quasi-prudent walks are always able to "escape" to infinity[13] by taking an infinite sequence of steps in the same lattice direction.[14]

Clearly prudent and perimeter walks are also quasi-prudent – a ray drawn from the endpoint, perpendicular to the side of the bounding box will suffice. However, in general, quasi-prudent

---

[13] In the sense of a lattice – perhaps "move infinitely far from the origin" is more precise.

[14] This definition also makes sense for walks on the triangular lattice, but is more complicated on the honeycomb lattice, where walks cannot take consecutive collinear steps. We will explore a related problem in Section 3.5.
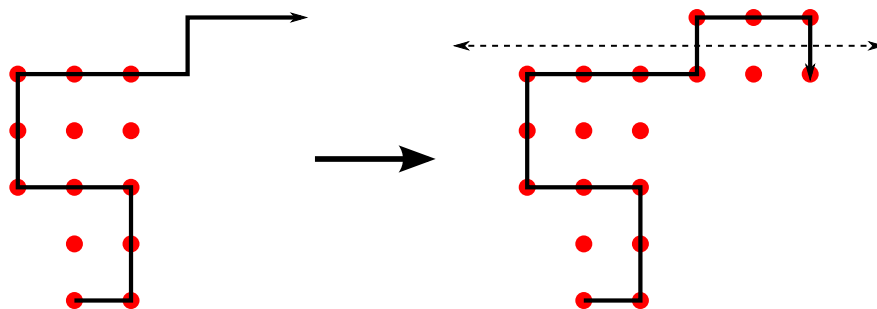
Figure 3.14: A quasi-prudent walk, before and after a step which adds points to the hull (the dots). The perpendicular bisector of that step is also indicated.

walks need not end on their bounding box; this means that the constructions involving inflating steps that we used for prudent and perimeter walks will no longer be of use here.

Instead, Mireille Bousquet-Mélou has suggested the use of a *hull*. It is defined as follows:

- The empty walk has no hull. (That is, its hull is the empty set.)

- After each step of the walk, consider the perpendicular bisector of that last step. WLOG assume the step was in the positive x direction, say from (x_0, y_0) to $(x_0 + 1, y_0)$. Then the bisector is the line $x = x_0 + 1/2$.

- If the bisector does not cross any other steps of the walk, the hull does not change. (That is, no new points are added to the hull.)

- Otherwise, say the bisector does cross another step of the walk, between vertices $(x_0, y_0 + \delta)$ and $(x_0 + 1, y_0 + \delta)$ with $\delta > 0$. Then every vertex (x, y) such that

$$x \in \{x_0, x_0 + 1\} \text{ and } y_0 \leq y \leq y_0 + \delta$$

  is added to the hull. (If $\delta < 0$ then the above inequalities should be reversed in the obvious way.) This process is illustrated in Figure 3.14.

- If the hull is now not connected via edges of the lattice, there will be a sub-path of the walk which connects the two pieces of the hull. Add the vertices of this sub-path to the hull. This process is illustrated in Figure 3.15.

We will denote the hull of a quasi-prudent SAW ω by $\mathscr{H}(\omega)$. See Figures 3.14 and 3.17 for examples of a quasi-prudent walks and their hulls.

Lemma 3.43. *The endpoint of a quasi-prudent walk is always adjacent to an unoccupied vertex which is not in its hull.*

Proof. Assume for a contradiction that a quasi-prudent walk ω has ended with a step from $v' = (x_0 - 1, y_0)$ to $v = (x_0, y_0)$ with all the vertices $a = (x_0 + 1, y_0)$, $b = (x_0, y_0 + 1)$ and $c = (x_0, y_0 - 1)$
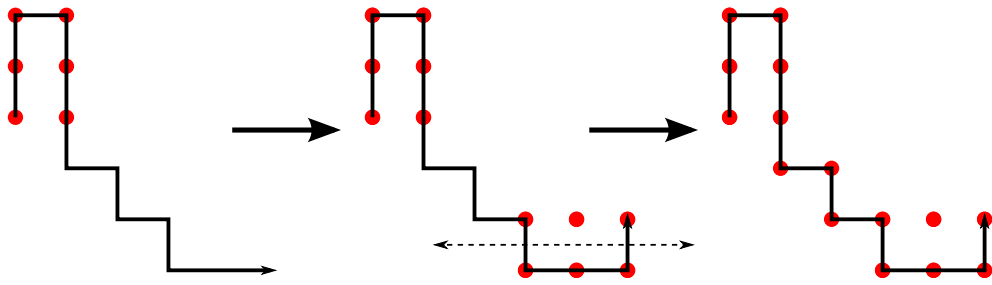
106

Figure 3.15: Another quasi-prudent walk, before and after a step which adds points to the hull (the dots). In this case, after checking the perpendicular bisector of the new step (the centre figure) and adding points between the steps it crosses, the hull was disconnected, and so we add points from the walk between the two pieces to connect the hull.
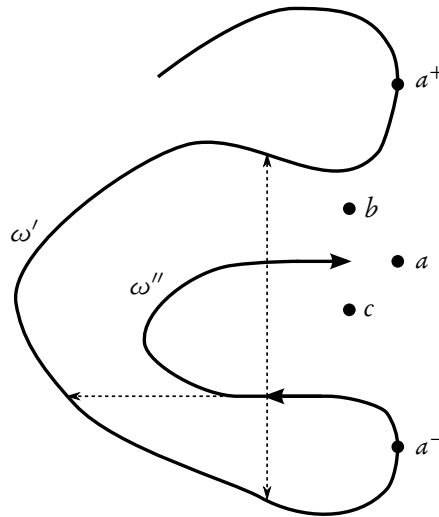

Figure 3.16: An illustration of the argument used in the proof of Lemma 3.43. The subwalk ω'' must contain a west step (indicated), but such a step cannot be quasi-prudent.

either occupied by ω or in the hull of ω. (If instead the last step was vertical or in the negative x direction, we can rotate/reflect the walk as necessary.) Clearly if they are all occupied by ω then this would contradict the quasi-prudent condition, so at least one must be unoccupied.

First, say there is a ray from v to infinity in the positive x direction (i.e. through a) which does not intersect ω. So a is unoccupied, but in the hull. This situation can only occur there if there are vertices both above and below a which are in ω. We denote these by vertices by $a^+ = (x_0 + 1, y_0 + \delta^+)$ and $a^- = (x_0 + 1, y_0 - \delta^-)$, with $\delta^+, \delta^- > 0$. Then a^+ and a^- must be connected by a subwalk ω' of ω which cannot cross the ray from v to infinity in the positive x direction, and one of a^+ and a^- must also be connected to v' via a subwalk ω''. But then there must be at least one west step in ω'', and a quick sketch makes it clear that after such a step, the walk would have been cut off in the north, west and south directions by ω', contradicting the quasi-prudent condition. See Figure 3.16 for an illustration.

Instead, suppose there is a ray from v to infinity in the positive y direction (i.e. through b) which does not intersect ω. So b is unoccupied, but in the hull. Then there are vertices $b^+ = (x_0 + \delta^+, y_0 + 1)$ and $b^- = (x_0 - \delta^-, y_0 + 1)$, with $\delta^+, \delta^- > 0$, which have been visited by ω. So there is a subwalk ω' of ω which connects b^+ and b^- and does not pass through the ray from v to infinity in the positive y direction, and another subwalk ω'' which connects one of b^+ and b^- to v'. Then ω'' must contain a south step, and after such a step the walk would have been cut off in the west, south and east directions by ω', contradicting the quasi-prudent condition.

The argument for the final case, when there is a non-intersecting ray from v to infinity in the negative y direction, is simply a reflection of the previous case.

So now there cannot be a ray from v in any direction which does not not intersect ω, a contradiction of the quasi-prudent condition. ∎

We define the *boundary* of the hull of a quasi-prudent walk ω to be the vertices in $\mathscr{H}(\omega)$ which are adjacent to at least one vertex not in $\mathscr{H}(\omega)$.

Corollary 3.44. *The endpoint of a quasi-prudent walk ω must either lie on the boundary of the hull of ω, or not be in the hull at all.*

So we see that, in some respects, the hull of a quasi-prudent walk functions like the bounding box of a prudent or perimeter walk – a quasi-prudent walk either ends on the boundary of its hull, or has stepped off its hull and not added any more vertices since. The second case is not as complicated as it sounds, since a walk that has not added any vertices to its hull must be very simple (in particular, it must be fully directed). This gives us a way to classify quasi-prudent walks in a similar manner to prudent and perimeter walks.

First, we define 8 (not necessarily distinct) points on the hull of a walk ω (assuming that the hull is non-empty):

$$N^+(\omega) = \max_x \max_y \{(x,y) \in \mathscr{H}(\omega)\} \qquad N^-(\omega) = \min_x \max_y \{(x,y) \in \mathscr{H}(\omega)\}$$

$$E^+(\omega) = \max_y \max_x \{(x,y) \in \mathscr{H}(\omega)\} \qquad E^-(\omega) = \min_y \max_x \{(x,y) \in \mathscr{H}(\omega)\}$$

$$S^+(\omega) = \max_x \min_y \{(x,y) \in \mathscr{H}(\omega)\} \qquad S^-(\omega) = \min_x \min_y \{(x,y) \in \mathscr{H}(\omega)\}$$

$$W^+(\omega) = \max_y \min_x \{(x,y) \in \mathscr{H}(\omega)\} \qquad W^-(\omega) = \min_y \min_x \{(x,y) \in \mathscr{H}(\omega)\}$$

Then we can define 8 "sides" of the hull (some or all of these might be a single vertex):

$$N(\omega) = \{(x,y) : y = N_y^+, N_x^- \le x \le N_x^+\}$$

$$E(\omega) = \{(x,y) : x = E_x^+, E_y^- \le y \le E_y^+\}$$

$$S(\omega) = \{(x,y) : y = S_y^+, S_x^- \le x \le S_x^+\}$$

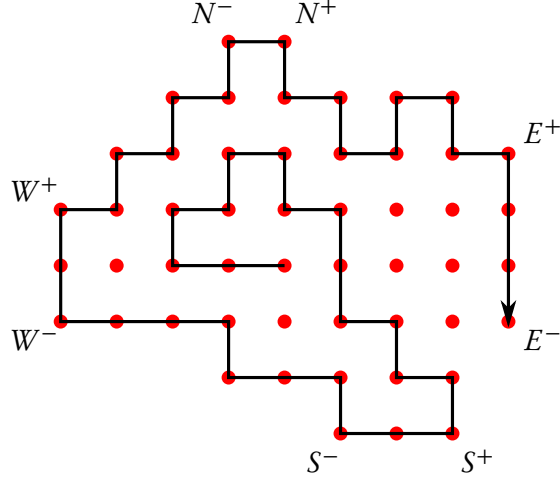$$W(\omega) = \{(x,y) : x = W_x^+, W_y^- \le y \le W_y^+\}$$

Figure 3.17: A quasi-prudent walk with its hull (the dots) and 'corner' points indicated.

$$NE(\omega) = \{(x,y) \in \mathscr{H}(\omega) : N_x^+ \le x \le E_x^+, E_y^+ \le y \le N_y^+, (x+1, y+1) \notin \mathscr{H}(\omega)\}$$

$$SE(\omega) = \{(x,y) \in \mathscr{H}(\omega) : S_x^+ \le x \le E_x^-, S_y^+ \le y \le E_y^-, (x+1, y-1) \notin \mathscr{H}(\omega)\}$$

$$SW(\omega) = \{(x,y) \in \mathscr{H}(\omega) : W_x^- \le x \le S_x^-, S_y^- \le y \le W_y^-, (x-1, y-1) \notin \mathscr{H}(\omega)\}$$

$$NW(\omega) = \{(x,y) \in \mathscr{H}(\omega) : W_x^+ \le x \le N_x^-, W_y^+ \le y \le N_y^-, (x-1, y+1) \notin \mathscr{H}(\omega)\}$$

where N_x^+ is the x-coordinate of $N^+(\omega)$, etc. In Figure 3.17 we identify the 8 "corner" points on the hull of a quasi-prudent walk.

We can now define a sub-classification of quasi-prudent walks according to the location of their endpoint in relation to the hull.

- A quasi-prudent walk ω is *1-sided* if, after every step, the endpoint of the current walk ω' is in $E(\omega')$ or not in $\mathscr{H}(\omega')$.

- A quasi-prudent walk ω is *2-sided* if, after every step, the endpoint of the current walk ω' is in $E(\omega') \cup NE(\omega')$ or not in $\mathscr{H}(\omega')$.

- A quasi-prudent walk ω is *3-sided* if, after every step, the endpoint of the current walk ω' is in $E(\omega') \cup NE(\omega') \cup N(\omega')$ or not in $\mathscr{H}(\omega')$.

- We continue in this fashion, adding more sides in a counter-clockwise direction, up to *8-sided* or *unrestricted* quasi-prudent walks.

Unfortunately, while this sub-classification looks pleasing on paper, it turns out to be largely useless; the only subclass of quasi-prudent walks that we have been able to solve (or even obtain functional equations for) is the 1-sided case. In this subsection, then, we will just solve the 1-sided case and mention numerical results for unrestricted quasi-prudent walks.

1-sided quasi-prudent walks

These are sufficiently simple that we can construct them from directed walks, though we will ultimately still need to use the kernel method. There are several distinct possibilities (we remind the reader that a walk is *fully directed* if it steps in at most two directions on the lattice):

- Any fully directed walk has no hull and is thus classified as 1-sided.

- A NES-directed walk containing all three allowed steps (i.e. one that actually has a hull) ends on the east side of its hull.

- A NW-directed walk containing at least one west step can be followed by a north and then east step, and will then have a non-empty hull. After this we can attach a NES-directed walk (maintaining self-avoidance, of course).

- Similarly, a SW-directed walk containing at least one west step can be followed by a south and then east step, and then a NES-directed walk.

The third case above is the only one we need to address; the first two are already solved, and the fourth is simply a reflection of the third. See Figure 3.18 for an illustration of this type of walk.

Now such a walk contains a first west step, say from $(0, y_0)$ to $(-1, y_0)$, and a last east step, say $(x_1 - 1, y_1)$ to (x_1, y_1). Then let $x' = \min\{0, x_1\}$. If we remove all steps to the east of $x = x'$, as well as any vertical steps along the line $x = x'$, then we are left with a particular kind of column-convex polygon [16] which is missing its easternmost vertical edges. Specifically, these are column-convex polygons whose southern boundary is a NW-directed walk and whose northern boundary is a NES-directed walk. We will solve the perimeter generating function of such polygons and then use it to compute the generating function for this class of 1-sided quasi-prudent walks.

Define the generating function

$$P(t; u, x) = \sum_{n, i, j \geq 0} p_{n, i, j} t^n u^i x^j,$$

where $p_{n, i, j}$ is the number of column-convex polygons with the following properties:

- the southern boundary is a NW-directed walk;

- the northern boundary is a NES-directed walk;

- the perimeter[15] is n;

- the height of the westernmost column is i; and

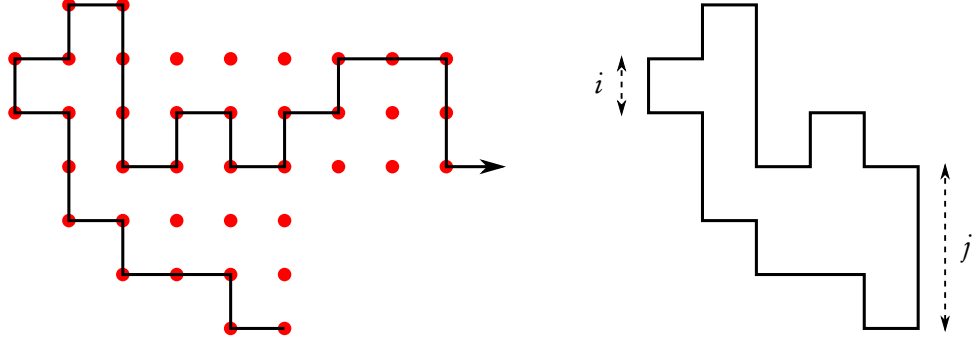- the height of the easternmost column is j.

110

Figure 3.18: A 1-sided quasi-prudent walk with its hull indicated, and the associated FP-polygon with the distances i and j indicated.

For lack of a better name we will call these *FP-polygons* (**F**ully-directed **P**artially-directed polygons). See Figure 3.18 for the FP-polygon corresponding to a given 1-sided quasi-prudent walk.

Lemma 3.45. *The generating function $P(t; u, x)$ satisfies the functional-differential equation*

$$\left(1 - \frac{t^2 u^2}{(1-u)^2} + \frac{t^4 u^2}{(1-u)(1-t^2 u)}\right) P(t; u, x) = \frac{t^4 u x}{1 - t^2 u x} - \frac{t^2 u^2 (1-t^2)}{(1-u)^2 (1-t^2 u)} P(t; 1, x)$$

$$+ \frac{t^2 u}{1-u} \left(\frac{\partial}{\partial u} P(t; u, x)\right)\bigg|_{u=1}. \quad (3.136)$$

Proof. We construct FP-polygons recursively by adding successive columns to the left side of existing FP-polygons.

- Polygons comprising only a single column have the generating function

$$\frac{t^4 u x}{1 - t^2 u x}. \quad (3.137)$$

- Polygons whose leftmost column is not longer than their second-leftmost column have the generating function

$$\sum_{n,i,j \geq 0} p_{n,i,j} t^n x^j \cdot t^2 \left(\sum_{k=0}^{i-1} \sum_{l=1}^{i-k} u^l\right)$$

$$= t^2 \sum_{n,i,j \geq 0} p_{n,i,j} t^n x^j \cdot \frac{u(i - u - iu + u^{i+1})}{(1-u)^2}$$

$$= \frac{t^2 u^2}{(1-u)^2} [P(t; u, x) - P(t; 1, x)] + \frac{t^2 u}{1-u} \left(\frac{\partial}{\partial u} P(t; u, x)\right)\bigg|_{u=1}. \quad (3.138)$$

- Polygons whose leftmost column is longer than their second-leftmost column have the

---

[15] *Not* the half-perimeter.

111

generating function

$$\sum_{n,i,j\geq 0} p_{n,i,j}\, t^n x^j \cdot t^2 \left(\sum_{k=1}^{i} u^k \sum_{l=1}^{\infty} t^{2l} u^l \right)$$

$$= t^2 \sum_{n,i,j\geq 0} p_{n,i,j}\, t^n x^j \cdot \frac{t^2 u^2 (1-u^i)}{(1-u)(1-t^2 u)}$$

$$= \frac{t^4 u^2}{(1-u)(1-t^2 u)} [P(t;1,x) - P(t;u,x)]. \tag{3.139}$$

Since all polygons are described by exactly one of the above three cases, adding (3.137), (3.138) and (3.139) gives $P(t;u,x)$, and the lemma follows. ∎

For every other polygon and walk model that we have examined in this chapter, the kernel of the functional equation is either quadratic in one or more of the catalytic variables (e.g. (3.122)), or is of no use when solving the equation (e.g. (3.133)). Fortunately, (3.136) is different – the kernel

$$K(t;u) = 1 - \frac{t^2 u^2}{(1-u)^2} + \frac{t^4 u^2}{(1-u)(1-t^2 u)}$$

is *cubic* in u, and two of the three roots of $K(t;u)=0$ are power series in t. This enables us to solve for $P(t;1,x)$ directly from (3.136), without having to find another equation satisfied by $P(t;1,x)$ and $\frac{\partial}{\partial u}P(t;u,x)|_{x=1}$.

Lemma 3.46. *Denote the two power series roots of $K(t;u)=0$ by $M \equiv M(t)$ and $N \equiv N(t)$, with*

$$M = 1 + t + t^2 + t^3 + \frac{3}{2}t^4 + \frac{5}{2}t^5 + \frac{9}{2}t^6 + \frac{65}{8}t^7 + 15t^8 + O(t^9) \tag{3.140}$$

$$N = 1 - t + t^2 - t^3 + \frac{3}{2}t^4 - \frac{5}{2}t^5 + \frac{9}{2}t^6 - \frac{65}{8}t^7 + 15t^8 + O(t^9). \tag{3.141}$$

Then

$$P(t;1,x) = \frac{-t^2 x(1-t^2 x)(1-M)(1-t^2 M)(1-N)(1-t^2 N)}{(1-t^2)(1-t^2 xM)(1-t^2 xN)(1-t^2 MN)} \tag{3.142}$$

$$= t^4 x + t^6(x+x^2) + t^8(2x+3x^2+x^3) + t^{10}(6x+7x^2+6x^3+x^4) + O(t^{12})$$

Proof. Substituting $u = M$ and $u = N$ into (3.136) gives two independent equations in $P(t;1,x)$ and $\frac{\partial}{\partial u}P(t;u,x)|_{x=1}$, which can then be combined to obtain the solution to $P(t;1,x)$. ∎

From here, the generating function for 1-sided quasi-prudent walks can be easily calculated.

Theorem 3.47. *The generating function $Q^{(1)}(t)$ for 1-sided quasi-prudent walks is given by*

$$Q^{(1)}(t) = \frac{1-t-t^2-3t^3-2t^4}{(1-t)(1-2t)(1-2t-t^2)} + \frac{2(1+t)(1-2t-t^3)}{(1-t)^2(1-2t)(1-2t-t^2)}R(t;1)$$

$$- \frac{2(1-2t-t^3)}{(1-t)^2(1-2t)(1-2t-t^2)}R(t;t) \tag{3.143}$$

$$= 1 + 4t + 12t^2 + 32t^3 + 80t^4 + 198t^5 + 482t^6 + 1172t^7 + 2832t^8 + O(t^9)$$

where

$$R(t;x) = \sum_{n,j \geq 0} r_{n,j} t^n x^j = P(t;1,t^{-1}x)$$

$$= \frac{-tx(1-tx)(1-M)(1-t^2M)(1-N)(1-t^2N)}{(1-t^2)(1-txM)(1-txN)(1-t^2MN)},$$

and M and N are as defined in Lemma 3.46.

Proof. The original motivation for solving the FP-polygon generating function $P(t;1,x)$ was that we obtained FP-polygons after deleting certain edges from 1-sided quasi-prudent walks. We now wish to reverse this process and add those edges back in.

Given an FP-polygon p, we begin by removing its rightmost vertical edges. The generating function of all such objects is exactly $R(t;x)$, where x is still conjugate to the height of the rightmost column. Now p will have precisely two rightmost vertices a and b, which were respectively the north-east and south-east corners of its rightmost column. To reverse the edge-deletion procedure, we will attach a NES-directed walk w_1 to a and/or a SE-directed walk w_2 to b, subject to the constraint that w_1 and w_2 cannot both contain an east step.

- If w_2 does contain an east step then w_1 cannot. The generating function of all possible w_2 walks is then

$$\frac{2t}{1-2t} - \frac{t}{1-t} = \frac{t}{(1-t)(1-2t)},$$

and so the generating function of 1-sided quasi-prudent walks obtained in this fashion is

$$\frac{t}{(1-t)(1-2t)} \sum_{n,j \geq 0} r_{n,j} t^n \cdot \left(\sum_{l=0}^{j-1} t^l + \sum_{l=1}^{\infty} t^l \right).$$

(The pair of sums over l correspond to w_1 containing south and north steps respectively.)

$$= \frac{t}{(1-t)(1-2t)} \sum_{n,j \geq 0} r_{n,j} t^n \cdot \left(\frac{1-t^j}{1-t} + \frac{t}{1-t} \right)$$

$$= \frac{t(1+t)}{(1-t)^2(1-2t)} R(t;1) - \frac{t}{(1-t)^2(1-2t)} R(t;t). \tag{3.144}$$

- Otherwise, w_2 does not contain any east steps. The generating function of possible w_2 walks is then just

$$\frac{1}{1-t},$$

and so the generating function of 1-sided quasi-prudent walks obtained in this fashion is

$$\frac{1}{1-t} \sum_{n,j \geq 0} r_{n,j} t^n \cdot \left(\sum_{l=0}^{j-1} t^l + \sum_{l=1}^{\infty} t^l \right) \cdot \left(1 + \frac{t(1+t)}{1-2t-t^2} \right).$$

(The pair of sums over l correspond to w_1 starting with south or north steps respectively. The last factor is the generating function of the NES-directed walks which can then be attached.)

$$= \frac{1}{1-t} \sum_{n,j \geq 0} r_{n,j} t^n \cdot \left(\frac{1-t^j}{1-t} + \frac{t}{1-t} \right) \cdot \left(\frac{1-t}{1-2t-t^2} \right)$$

$$= \frac{1+t}{(1-t)(1-2t-t^2)} R(t;1) - \frac{1}{(1-t)(1-2t-t^2)} R(t;t). \qquad (3.145)$$

Adding (3.144) and (3.145) gives the generating function for the 1-sided quasi-prudent walks generated from FP-polygons. We then need to multiply this by two to account for those walks which start with a SW-directed walk and then turn south and continue as a NES-directed walk.

We finally need to account for the directed walks which are also 1-sided quasi-prudent walks. These include NES-directed walks, which have generating function

$$\frac{1+t}{1-2t-t^2},$$

and the NW- and SW-directed walks which are not also NES-directed, which have generating function

$$\frac{t(1+2t)}{(1-t)(1-2t)}.$$

Adding all these generating functions together gives the result of the theorem. ∎

By inspection it is clear that $Q^{(1)}(t)$ is algebraic. It is straightforward to verify that the dominant singularity of $Q^{(1)}(t)$ is a simple pole at $t = \sqrt{2} - 1 = 0.414\ldots$. The roots M and N are analytic on $|t| < \tau$ where $\tau = 0.462\ldots$ is a root of $4 - 16\tau^2 - 12\tau^4 - 3\tau^6 = 0$, and all of the values of t which cancel the denominators of $R(t;1)$ or $R(t;t)$ are outside the region $|t| \leq \sqrt{2} - 1$.

Corollary 3.48. *The number $q_n^{(1)}$ of 1-sided quasi-prudent walks of length n is asymptotically*

$$q_n^{(1)} \sim x(1+\sqrt{2})^n,$$

where $x = 2.38551\ldots$.

Unrestricted quasi-prudent walks

The result of Corollary 3.48 is somewhat disappointing, as it shows that for the only class of quasi-prudent walks that we can solve exactly, the exponential growth rate is the same as for partially directed walks and far less than even that of 2-sided prudent walks. The problem seems to be that, while the notion of the hull is useful for devising a sub-classification of quasi-prudent walks, its irregular shape means that it offers little assistance when trying to recursively generate such walks.

To study quasi-prudent walks of greater generality than the 1-sided case, we then have to resort to numerical studies. We have used a backtracking algorithm to generate walks of length up to 32, and we find that asymptotically the number of 8-sided (unrestricted) quasi-prudent walks of length n is

$$q_n^{(8)} \sim \varkappa \mu^n n^g,$$

where $\mu \approx 2.609$ and $g \approx 1.0$, and $\varkappa$ is a positive constant.

Tim Garoni [9] has performed a Monte Carlo study of very long quasi-prudent walks (up to 10000 steps), and predicts that the mean squared end-to-end distance of quasi-prudent walks of length n is asymptotically αn^2 for a constant α. Equivalently, he conjectures that the exponent $\nu = 1$ for quasi-prudent walks.

1-sided quasi-prudent polygons by perimeter

As with prudent and perimeter walks, we can define a k-sided quasi-prudent polygon to be a k-sided quasi-prudent walk which ends at a vertex adjacent to the start of the walk. We can compute the half-perimeter generating function of 1-sided quasi-prudent polygons, but have not made any study of the more general cases.

Fortunately, most of the work is already done – the generating function can be obtained with only minor modifications to the function $R(t; x)$ we calculated in Theorem 3.47.

Lemma 3.49. *The half-perimeter generating function for 1-sided quasi-prudent polygons is given by*

$$QP^{(1)}(t) = \frac{2(1 - 2t + 2t^2)}{(1-t)^2} R(\sqrt{t}; \sqrt{t}) + \frac{2t^{3/2}}{1-t} \left(\frac{\partial}{\partial x} R(\sqrt{t}; x) \right)\Bigg|_{x=\sqrt{t}} + \frac{2t^2}{1-t} \tag{3.146}$$

$$= 4t^2 + 8t^3 + 24t^4 + 80t^5 + 284t^6 + 1052t^7 + 4020t^8 + O(t^9).$$

Proof. There are three main cases to consider:

- The NES-directed walks which form polygons are trivial – they are just comprised of a single column above or below the x-axis, and have the generating function

$$\frac{2t^2}{1-t}. \tag{3.147}$$

- The walks which do not step to the east of the y-axis are exactly the FP-polygons (and their reflections) we considered in Lemmas 3.45 and 3.46. Their perimeter generating function is $2P(t; 1, 1) = 2R(t; t)$, and so the half-perimeter generating function is $2R(\sqrt{t}; \sqrt{t})$.

- The non-trivial walks which do step east of the y-axis are FP-polygons (and their reflections) with a column attached to the right side, with the constraint that the bottom of the new column is no higher than the bottom of the FP-polygon. (For the reflections, the top

of the new column must be no lower than the top of the FP-polygon.) Their generating function is thus

$$\frac{2}{1-t} \sum_{n,j\geq 0} r_{n,j} \sqrt{t}^{\,n} \cdot t \left(\sum_{l=0}^{j-1} \sqrt{t}^{\,l} \sqrt{t}^{\,j-l} + \sum_{m=1}^{\infty} \sqrt{t}^{\,m} \sqrt{t}^{\,j+m} \right)$$

(The sum over l corresponds to the cases when the top of the new column is no higher than the top of the FP-polygon's rightmost column; the sum over m covers the other cases.)

$$= \frac{2t}{1-t} \sum_{n,j\geq 0} r_{n,j} \sqrt{t}^{\,n} \left(j \sqrt{t}^{\,j} + \frac{\sqrt{t}^{\,j+2}}{1-t} \right)$$

$$= \frac{2t^{3/2}}{1-t} \left(\frac{\partial}{\partial x} R(\sqrt{t};x) \right)\Big|_{x=\sqrt{t}} + \frac{2t^2}{(1-t)^2} R(\sqrt{t};\sqrt{t}) \tag{3.148}$$

Adding together the three cases above gives the result of the lemma. ∎

The dominant singularity of $QP^{(1)}(t)$ is only marginally more complicated than that of $Q^{(1)}(t)$; we find that here it is a square-root singularity occurring in $M(\sqrt{t})$ and $N(\sqrt{t})$, at the point $t = \rho = 0.213862\ldots$, a root of $4 - 16\rho - 12\rho^2 - 3\rho^3 = 0$.

Corollary 3.50. *The number $qp_n^{(1)}$ of 1-sided quasi-prudent polygons of half-perimeter n is asymptotically*

$$qp_n^{(1)} \sim \varkappa \mu^n n^{-3/2},$$

where $\mu = \rho^{-1} = 4.675891\ldots$ and $\varkappa = 0.392267\ldots$.

1-sided quasi-prudent polygons by area

The area generating function for 1-sided quasi-prudent polygons is even easier to calculate than the perimeter generating function, though of course we cannot use the $P(t;u,x)$ function we derived earlier.

Instead, define the generating function

$$A(q;u) = \sum_{n,i\geq 0} a_{n,i} q^n u^i$$

where $a_{n,i}$ is the number of FP-polygons with area n and left column of height i.

Lemma 3.51. *The function $A(q;u)$ satisfies the equation*

$$A(q;u) = \frac{qu}{1-qu} + \frac{qu}{1-qu} \left(\frac{\partial}{\partial u} A(q;u) \right)\Big|_{u=1}. \tag{3.149}$$

Proof. There are two cases to consider:

- Those polygons comprising only a single column have the generating function

$$\frac{qu}{1-qu}.$$

- Otherwise, an FP-polygon can be constructed by adding a column to the left side of another FP-polygon. These then have the generating function

$$\sum_{n,i\geq 0} a_{n,i} q^n \cdot \left(\sum_{l=0}^{i-1} \sum_{k=1}^{\infty} q^k u^k \right)$$

$$= \sum_{n,i\geq 0} a_{n,i} q^n \cdot \frac{iqu}{1-qu}$$

$$= \frac{qu}{1-qu} \left(\frac{\partial}{\partial u} A(q;u) \right) \Bigg|_{u=1}.$$

Adding together the two above cases gives the result of the lemma. ∎

Lemma 3.52. *The area generating function for 1-sided quasi-prudent polygons is*

$$QA^{(1)}(q) = \frac{4q(1-q)}{1-3q+q^2} \tag{3.150}$$

$$= 4q + 8q^2 + 20q^3 + 52q^4 + 136q^5 + 356q^6 + 932q^7 + 2440q^8 + O(q^9).$$

Thus the number $qa_n^{(1)}$ of 1-sided quasi-prudent polygons of area n is asymptotically

$$qa_n^{(1)} \sim \varkappa \tau^n$$

where $\tau = (3+\sqrt{5})/2 = 2.61803\ldots$ and $\varkappa = (10-2\sqrt{5})/5 = 1.10557\ldots$.

Proof. Taking the partial derivative of (3.149) with respect to u and then setting $u=1$ gives

$$\frac{\partial}{\partial u} A(q;u) \Bigg|_{u=1} = \frac{q}{1-3q+q^2}, \tag{3.151}$$

which upon substitution into (3.149) gives

$$A(q;1) = \frac{q(1-q)}{1-3q+q^2}. \tag{3.152}$$

As was discussed for the perimeter case, there are now three main cases for 1-sided quasi-prudent polygons:

- Polygons comprising a single column above or below the x-axis have the generating function

$$\frac{2q}{1-q}. \tag{3.153}$$

- Polygons which do not step to the right of the y-axis are exactly FP-polygons (or their reflections). These have generating function $2A(q;1)$.

- Non-trivial polygons which step to the right of the y-axis are FP-polygons (or their reflections) with a column attached to the right side, with the constraint that the bottom of the new column must be no higher than the bottom of the FP-polygon. (For the reflections, the top of the new column must be no lower than the top of the FP-polygon.) Their generating function is thus

$$\frac{2q}{(1-q)^2}A(q;1). \tag{3.154}$$

Adding together the three cases above gives $QA^{(1)}(q)$.

The dominant singularity is clearly a simple pole at a root of $1-3q+q^2=0$. The smallest such root is $\tau^{-1}=(3-\sqrt{5})/2=0.381966\dots$. $\blacksquare$

3.3.3 Other models

As mentioned at the start of this chapter, one of the primary goals in studying subclasses of SAWs and SAPs is finding solvable models which display behaviour similar to that of the general cases. In practice, this usually boils down to two criteria:

- Looking for a model, enumerated by say $\{a_n\}$, such that the asymptotic behaviour of a_n closely resembles (1.7) (for walks) or (1.11) (for polygons), with the growth constant and exponent close to the known/conjectured values for the lattice; or

- Looking for a model for which some measurement of the mean 'size' (e.g. mean squared end-to-end distance of walks of length n) resembles (1.29), with the exponent equal or close to the known/conjectured value for the lattice.

In this subsection we will briefly mention two (existing) results which are set apart from the other models presented in this chapter, in ways which are related to the above two points.

Weakly directed walks

Recall from Chapter 1 that the current best estimate for the growth constant of the square lattice is [73, 75]

$$\mu_{\text{square}} \approx 2.63815853031(3).$$

Of all the *solvable* models described in this chapter thus far, the one whose exponential growth rate is the closest to this estimate is the model of 2-sided perimeter walks, for which the growth rate is

$$\tau \approx 2.50399575841.$$

(The conjectured value for quasi-prudent walks is much higher than this, at around 2.609, but we are unable to solve that model.)

However, Bacher and Bousquet-Mélou [2] have recently defined and solved a subclass of SAWs with a growth rate considerably larger than that of 2-sided perimeter walks. They call them *weakly directed walks*, and they have the following very simple definition: a SAW is weakly directed if, between two visits to any given horizontal line, the walk is partially directed.

Bacher and Bousquet-Mélou show that weakly directed walks are easily constructed by concatenating irreducible bridges, and that these bridges must be partially directed. They use this construction to solve for the generating function of weakly directed walks, and find that the number w_n of these walks of length n is asymptotically

$$w_n \sim \varkappa \mu^n$$

where $\mu \approx 2.5447$ and $\varkappa$ is a positive constant.

As Bacher and Bousquet-Mélou note [2], this value for the growth constant of a solvable subclass of SAWs is currently the largest known, except for somewhat unsatisfying models like SAWs confined to a strip of small width [122] or SAWs constructed by concatenating irreducible bridges of small lengths [74].

Spiral walks

In Chapter 1 we mentioned that the mean squared end-to-end distance for SAWs of length n, denoted $\langle R_e^2 \rangle_n$, is expected to behave asymptotically as

$$\langle R_e^2 \rangle_n \sim E n^{2\nu},$$

where E is a lattice-dependent positive constant and ν depends only on the dimensionality of the lattice. It is expected that in two dimensions, $\nu = 3/4$.

For all of the solved (and unsolved, for that matter) walk models presented in this chapter thus far, the mean squared end-to-end distance of walks of length n has been shown or is expected to behave asymptotically like $\bar{E} n^2$ for some constant $\bar{E}$. That is, if we define the exponent ν for subclasses of SAWs just as for general SAWs, then for all the subclasses considered in this chapter so far, ν is (or is believed to be) one.

However, there exists a solvable subclass of SAWs, namely *spiral walks*, for which the exponent ν is less than 1. These are simply SAWs for which a left turn is forbidden;[16] that is, the consecutive pairs of steps north-west, east-north, south-east and west-south are forbidden.[17] If we denote by s_n the number of such walks of length n, then Blöte and Hilhorst [15] and

---

[16] By symmetry we could of course forbid right turns instead; there appears to be no clear consensus in the literature.

[17] Such a restriction is known as a *two-step rule* [57].

119

Guttmann and Wormald [59] independently found that

$$s_n \sim \frac{\pi}{4 \cdot 3^{5/4}} n^{-7/4} \exp(2\pi(n/3)^{1/2}).$$

This means that, from an enumerative perspective, spiral walks are far removed from general SAWs – they don't even have an exponential growth rate! Instead, the dominant factor in the asymptotic form of s_n is $\mu^{\sqrt{n}}$ for a positive constant μ.

However, Blöte and Hilhorst also showed that the mean squared end-to-end distance of spiral walks of length n, which we will denote by $\langle S_e^2 \rangle_n$, has the asymptotic form

$$\langle S_e^2 \rangle_n \sim \frac{3}{4\pi^2} n (\log n)^2.$$

While this is still quite different from the $E n^{3/2}$ we expect for general two-dimensional SAWs, it is significant in that it is different from the $O(n^2)$ behaviour we observe in all other solvable models. Ignoring the logarithmic correction term in $\langle S_e^2 \rangle_n$, we note that for spiral walks the exponent $\nu = 1/2$. To our knowledge there are no known solved models for which $1/2 < \nu < 1$.

3.4   Triangular lattice

In this section and the next we will briefly consider the other regular two-dimensional lattices: the triangular and honeycomb (hexagonal) lattices[18]. Historically there seems to have been less interest in finding solvable models on these lattices than on the square lattice, and this is reflected in the fact that we have spent relatively little time studying solvable models on these other lattices. The only models we will consider here are adaptations of prudent walks from the square lattice.

The definition of a prudent walk carries over from the square lattice to the triangular lattice in the obvious way: a SAW on the triangular lattice is prudent if it never takes a step towards an occupied lattice vertex.[19] The notion of the bounding box, however, does not translate quite so easily – in general, the 'box' becomes an irregular six-sided shape, where one or more of the sides can have length 0 (i.e. consist of only a single vertex). See Figure 3.19 for an example of a prudent walk on the triangular lattice and its irregular 'box'.

Bousquet-Mélou [17] considers prudent walks on the triangular lattice[20] with the added restriction that every walk must end on the boundary of its *bounding triangle* – the smallest

---

[18]For the honeycomb lattice, 'regular' is somewhat of a misnomer, as the lattice is not vertex-transitive. However, viewed as a graph, the honeycomb lattice is 3-regular, and moreover it is face-transitive, so we do not feel too uncomfortable referring to it as 'regular'.

[19]As we will see in Section 3.5, things are somewhat more complicated on the honeycomb lattice.

[20]Bousquet-Mélou orients the lattice so that it contains horizontal edges (corresponding with e.g. Figure 3.19), and we will follow this convention.

north-pointing equilateral triangle containing the entire walk.[21] She solves the generating function of such walks, finding it to be non-D-finite, with dominant singularity a simple pole at $\rho = (\sqrt{17} - 3)/4 = 0.280\ldots$. Thus the number pt_n of these triangular prudent walks of length n satisfies

$$pt_n \sim x \left(\frac{3 + \sqrt{17}}{2} \right)^n$$

for some positive constant x, where the growth rate $(3 + \sqrt{17})/2 \approx 3.56$. (We remind the reader here that the number of SAWs on the triangular lattice is expected to behave like

$$c_n \sim An^{\gamma - 1}\mu^n,$$

where $\gamma = 43/32$ and $\mu \approx 4.150797226$.)

We will remove the bounding triangle restriction and instead consider prudent walks on the triangular lattice in terms of the irregular 'box'. In an abuse of mathematical language, we shall refer to this shape as the *bounding hexagon*, or just the *hexagon* of a walk. To be more precise, the hexagon of a walk ω is the smallest convex lattice polygon which contains ω. It will have six sides: north, north-east, south-east, south, south-west and north-west; some (or all, in the case of the empty walk) may consist of single vertices. In the same way that a prudent walk on the square lattice must end on its bounding box, a prudent walk on the triangular lattice must end on its hexagon. We can then define six subclasses of prudent walks on the triangular lattice:

- A prudent walk is *1-sided* if, after every step, the endpoint of the current walk is on the north-east side of its hexagon.

- A prudent walk is *2-sided* if, after every step, the endpoint of the current walk is on the north-east or north sides of its hexagon.

- We continue adding sides in a counter-clockwise fashion for 3-sided through to 6-sided (unrestricted) prudent walks.

In Figure 3.20 we show examples of 1- and 2-sided prudent walks.

1-sided prudent walks

These are essentially the triangular lattice-analogue of partially directed walks on the square lattice – they are unable to take west or south-west steps, and as long as we forbid immediate reversals then all such walks are prudent. Define the generating function

$$T^{(1)}(t) = \sum_{n \geq 0} t_n^{(1)} t^n,$$

where $t_n^{(1)}$ is the number of 1-sided prudent walks of length n.

---

[21]We will refer to such walks here and in Chapter 4 as *equilateral* prudent walks.
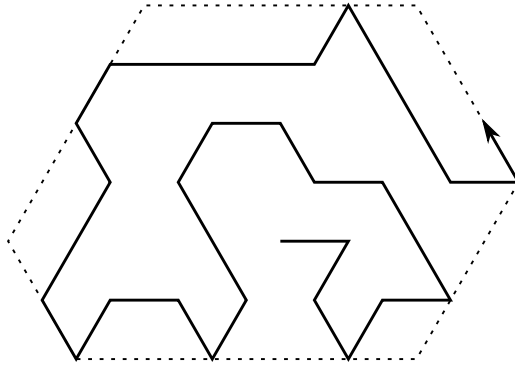
Figure 3.19: A prudent walk on the triangular lattice with its hexagon indicated.

Lemma 3.53. *The generating function $T^{(1)}(t)$ of 1-sided prudent walks on the triangular lattice is*

$$T^{(1)}(t) = \frac{1+t}{1-3t-2t^2}$$
$$= 1 + 4t + 14t^2 + 50t^3 + 178t^4 + 634t^5 + 2258t^6 + O(t^7). \qquad (3.155)$$

Thus the number of 1-sided prudent walks of length n satisfies

$$t_n^{(1)} \sim \left(\frac{17 + 5\sqrt{17}}{34}\right) \cdot \left(\frac{3 + \sqrt{17}}{2}\right)^n.$$

Proof. As we did on the square lattice, we divide walks according to their *last inflating step*. Here, any north-east or east step is inflating.

- Walks which have no inflating steps contain only north-west or south-east steps; their generating function is thus

$$1 + \frac{2t}{1-t}.$$

- Otherwise, walks which have taken an inflating step can be split into three parts: what came before the last inflating step (any 1-sided walk), the last inflating step (north-east or east), and what came after the last inflating step (nothing, north-west steps or south-east steps). Their generating function is thus

$$T^{(1)}(t) \cdot 2t \cdot \left(1 + \frac{2t}{1-t}\right).$$

Adding the above two cases together gives

$$T^{(1)}(t) = 1 + \frac{2t}{1-t} + T^{(1)}(t) \cdot 2t \cdot \left(1 + \frac{2t}{1-t}\right), \qquad (3.156)$$

the solution to which is exactly the result of the lemma. The dominant singularity is a simple pole at a root of $1 - 3t - 2t^2 = 0$. ∎
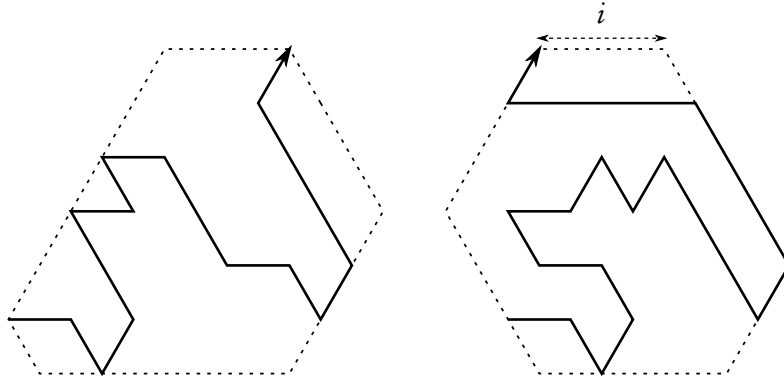
Figure 3.20: 1-sided and 2-sided prudent walks on the triangular lattice, with their respective hexagons and the distance i (for the 2-sided walk) indicated.

Note that the growth rate of 1-sided prudent walks is the same as that of Bousquet-Mélou's prudent walks which end on their bounding triangle. This serves to highlight how strong the bounding triangle restriction is, and suggests that the average walk behaves much like a 1-sided walk (or a reflection of one).

2-sided prudent walks

The construction and solution for 2-sided prudent walks looks much like that of 2-sided prudent walks on the square lattice. By symmetry, we only need to count those walks ending on the north side of their hexagon. Define the generating function by

$$N(t;u) = \sum_{n,i \geq 0} n_{n,i} t^n u^i$$

where $n_{n,i}$ is the number of 2-sided prudent walks of length n which end on the north side of their hexagon, with the endpoint at distance i from the easternmost vertex of the north side of the hexagon. (See Figure 3.20 for an example of a 2-sided prudent walk with the distance i indicated.)

Lemma 3.54. *The generating function $N(t;u)$ satisfies the functional equation*

$$\left(1 - \frac{t(1-t^2)(1+u)}{(u-t)(1-tu)}\right) N(t;u) = \frac{1}{1-tu} + \frac{(1+t)(u-2t)}{u-t}N(t;t) - \frac{1-t+t^2-tu}{1-tu}N(t;0).$$

(3.157)

Proof. As usual, we count walks according to their last inflating step.

- Walks with no inflating step contain only west steps, and have the generating function

$$\frac{1}{1-tu}.$$

123

- Walks whose last inflating step was north-west are constructed by adding an inflating step to any walk counted by $N(t;u)$, and then possibly east or west steps. Their generating function is

$$\sum_{n,i\geq 0} n_{n,i} t^n \cdot t \left(\sum_{k=1}^{i} t^k u^{i-k} + \sum_{l=0}^{\infty} t^l u^{i+l} \right)$$

(The sum over k is for walks with east steps following the inflating step, while the sum over l accounts for west steps.)

$$= t \sum_{n,i\geq 0} n_{n,i} t^n \cdot \left(\frac{t(u^i - t^i)}{u - t} + \frac{u^i}{1 - tu} \right)$$

$$= \frac{t^2}{u - t} [N(t;u) - N(t;t)] + \frac{t}{1 - tu} N(t;u).$$

- Similarly, walks with last inflating step north-east (from the north side of the hexagon) have generating function

$$\sum_{n,i\geq 0} n_{n,i} t^n \cdot t \left(\sum_{k=0}^{i-1} t^k u^{i-1-k} + \sum_{l=1}^{\infty} t^l u^{i-1+l} \right)$$

(The sum over k is for walks with east steps following the inflating step, while the sum over l considers west steps.)

$$= t \sum_{n,i\geq 0} n_{n,i} t^n \cdot \left(\frac{u^i - t^i}{u - t} + \frac{t u^i}{1 - tu} \right)$$

$$= \frac{t}{u - t} [N(t;u) - N(t;t)] + \frac{t^2}{1 - tu} N(t;u),$$

except we are now missing the term for when the inflating step is from the north-east corner and no west steps follow. So we must add on an extra $tN(t;0)$.

- We likewise consider walks with last inflating step north-east (from the north-east side of the hexagon), which have generating function

$$N(t;t) - N(t;0).$$

(The subtracted term is because walks with last inflating step from the corner were previously counted.)

- Finally, walks with last inflating step east have generating function

$$tN(t;t).$$

Adding all the above together and equating with $N(t;u)$ gives the result of the lemma. ∎

Lemma 3.55. *Define*

$$G \equiv G(t) = \frac{1 - t + t^2 + t^3 - (1 + t)\sqrt{1 - 4t + 2t^2 + t^4}}{2t}$$

$$X(t;u) = (u - t)(2 + tu - t^2)$$

$$Y(t;u) = (1 + t)(u - 4t + 2t^3 + tu + 2t^2 u - 2t^3 u - tu^2 + t^2 u^2)$$

$$K(t;u) = u - 2t + t^3 - tu + t^2 u + t^3 u - tu^2.$$

Then the generating function $N(t;u)$ has the solution

$$N(t;u) = \frac{1}{K(t;u)(3 - t)}\left(X(t;u) - \frac{Y(t;u)X(t;G)}{Y(t;G)}\right), \tag{3.158}$$

and the generating function for 2-sided prudent walks is

$$T^{(2)}(t) = 2N(t;1) - N(t;0)$$

$$= \frac{3 - 10t - t^2 + 6t^3 - 2t^4 + 2t(1 - t)\sqrt{1 - 4t + 2t^2 + t^4}}{(1 - 3t - 2t^2)(3 - 14t + 11t^2 - 4t^3)} \tag{3.159}$$

Proof. Substituting $u = 0$ into (3.157) gives

$$N(t;0) = \frac{1}{3 - t} + \frac{2(1 + t)}{3 - t}N(t;t), \tag{3.160}$$

and then using (3.160) to eliminate $N(t;0)$ from (3.157) leaves

$$K(t;u)N(t;u) = \frac{X(t;u)}{3 - t} + \frac{Y(t;u)}{3 - t}N(t;t). \tag{3.161}$$

Now G is the power series root of $K(t;u) = 0$, so substituting $u = G$ into (3.161) gives

$$N(t;t) = -\frac{X(t;G)}{Y(t;G)},$$

and then the solution to $N(t;u)$ follows. To obtain $T^{(2)}(t)$, we add together the generating functions of walks ending on the north and north-east sides of the hexagon, and then subtract the walks ending at the corner which have been counted twice. ∎

The generating function is algebraic, and thus its singular behaviour is easily verified.

Corollary 3.56. *The number $t_n^{(2)}$ of 2-sided prudent walks of length n satisfies*

$$t_n^{(2)} \sim \varkappa \tau^n$$

where $\tau = 3.792560\ldots$, a root of $4 - 11\tau + 14\tau^2 - 3\tau^3 = 0$, and $\varkappa = 1.39239\ldots$.

3-sided prudent walks

We are able to construct functional equations for the 3-sided case, but unfortunately our attempts to solve these equations have been unsuccessful. For brevity we will omit the proof; it follows in much the same way as the 2-sided case.

Define the generating functions

$$A(t;u,v) = \sum_{n,i,j \geq 0} a_{n,i,j} t^n u^i v^j$$

$$B(t;u,v) = \sum_{n,i,j \geq 0} b_{n,i,j} t^n u^i v^j$$

where $a_{n,i,j}$ is the number of 3-sided prudent walks ending on the north side of their hexagon, with

- length n,

- distance i from the endpoint to the easternmost vertex of the north side of the hexagon, and

- distance j from the endpoint to the westernmost vertex of the north side of the hexagon.

Similarly, $b_{n,i,j}$ is the number of 3-sided prudent walks ending on the north-east side of their hexagon, with

- length n,

- distance i from the endpoint to the easternmost vertex of the north side of the hexagon, and

- the north side of the hexagon of length j.

In Figure 3.21 we illustrate examples of the two types of 3-sided prudent walk, together with the distances measured by the catalytic variables.

Proposition 3.57. *The generating functions $A(t;u,v)$ and $B(t;u,v)$ satisfy the functional equations*

$$\left(1 - \frac{t(1-t^2)(u+v)}{(u-tv)(v-tu)}\right) A(t;u,v) = 1 - \frac{t(1+t)}{v-tu} A(t;u,tu) - \frac{t(1+t)}{u-tv} A(t;tv,v)$$

$$- \frac{t^2(1-u)}{v-tu} A(t;0,v) + \frac{t^2(1-u)}{v-tu} A(t;0,tu) - \frac{t^2(1-v)}{u-tv} A(t;u,0) + \frac{t^2(1-v)}{u-tv} A(t;tv,0)$$

$$+ v(1+t)B(t;t,v) - (v-t)B(t;0,v) + u(1+t)B(t;t,u) - (u-t)B(t;0,u) \quad (3.162)$$

$$\left(1 - \frac{tv(1-t^2)(1+u)}{(u-t)(1-tu)}\right) B(t;u,v) = \frac{1}{1-tu} - \frac{tv(1+t)}{u-t} B(t;t,v) + \frac{t(1-tv)}{1-tu} B(t;0,v)$$

$$+ \frac{1+t}{v} A(t;tv,v) - \frac{1}{v} A(t;0,v) - \frac{t(1-v)}{v} A(t;tv,0). \quad (3.163)$$
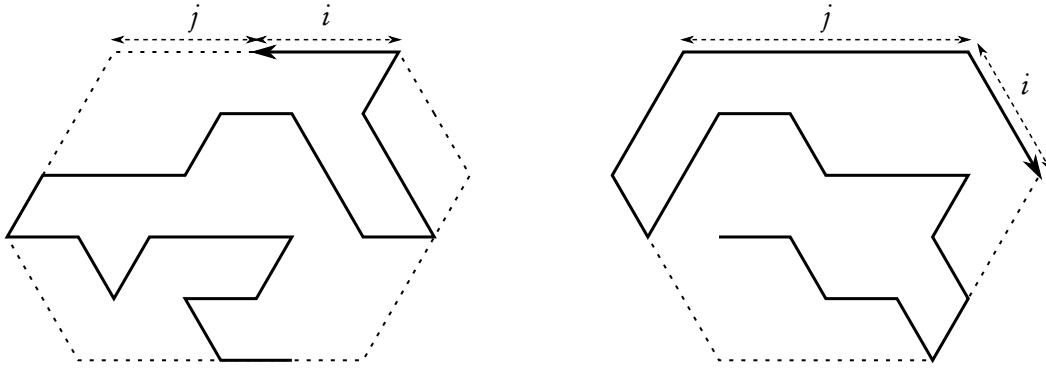
Figure 3.21: 3-sided prudent walks on the triangular lattice, ending on the north and north-east sides of their hexagons, with the distances i and j indicated.

The construction used to generate these walks is very similar to the one used for 3-sided prudent walks on the square lattice [31, 17], and so the resulting functional equations (3.162) and (3.163) do seem very similar. However, matters are complicated here by the fact that there is more nuanced behaviour occurring at the corners of the hexagon – in particular, a north-east step from the north-east corner of the hexagon moves *both* the north and north-east sides of the box (and likewise for a north-west step from the north-west corner of the hexagon). This results in a number of extra terms featuring in these functional equations which are not needed for the square lattice case – terms like $A(t; u, 0)$ and $B(t; 0, u)$.

The generating function for 3-sided prudent walks is $T^{(3)}(t) = A(t; 1, 1) + 2B(t; 1, 1) - 2B(t; 0, 1)$. Numerical evidence suggests that the growth rate of 3-sided prudent walks is approximately 3.84138, and that the singularity is a simple pole.

It should be possible to write down functional equations for the 4-, 5- and 6-sided cases as well, but given that we have little chance of being able to solve such equations, we will forego this task for the present time.

3.5 Honeycomb lattice

As we mentioned at the start of the last section, there are some factors which make prudent walks on the honeycomb lattice somewhat more complicated than on the triangular (and square) lattice. Most obviously, there are two types of vertices on the honeycomb lattice:[22] those with edges in the west, north-east and south-east directions, and those with edges in the east, north-west and south-west directions. This means that we will need to split the walks into two groups, according to which type of vertex they end on, and use separate generating functions for each group.

More subtle is the fact that walks on the honeycomb lattice cannot actually step in straight

---

[22]Throughout this section we will orient the honeycomb lattice as in Figure 3.22, that is, so that the lattice contains horizontal edges.
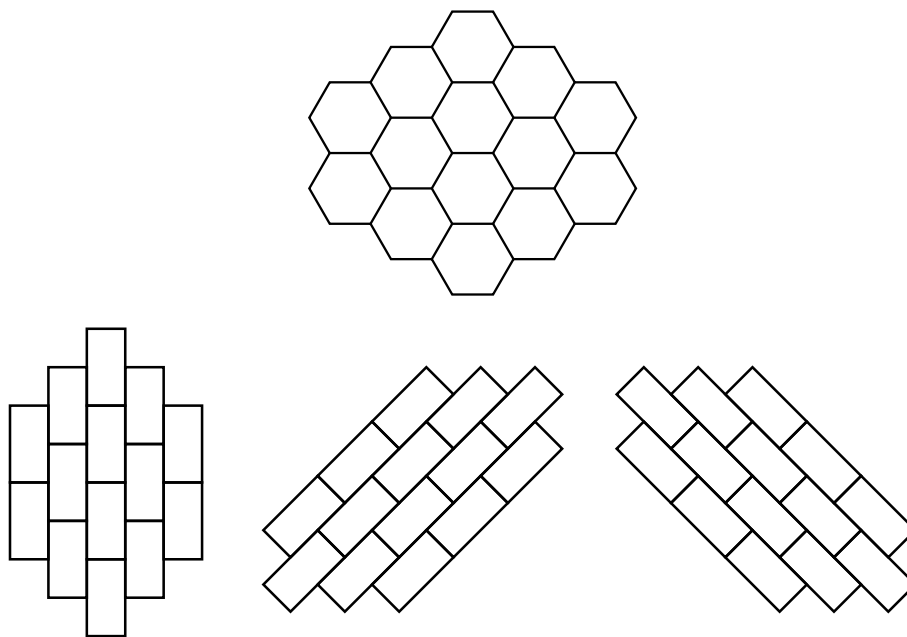
Figure 3.22: A section of the honeycomb lattice and the three orientations of equivalent brick-work lattices: vertical, positive and negative respectively.

lines – they are forced to turn through an angle of $\pm\pi/3$ at each vertex. This makes the normal definition of 'prudent' rather peculiar – we wish to forbid a walk from stepping towards an occupied vertex, but it wouldn't be able to take more than one step at a time in a given direction anyway.

Instead, we propose a slight modification of the prudent condition; it is most easily visualised in terms of *brickwork lattices*. A brickwork lattice is simply a deformation of the honeycomb lattice which results in all edges lying in one of two perpendicular orientations. A honeycomb lattice is equivalent to three different orientations of a brickwork lattice;[23] we will refer to the three orientations as *vertical*, *positive* and *negative*. (See Figure 3.22.)

We present the following definition of a prudent walk on the honeycomb lattice:[24]

- Any walk of length 0 or 1 is considered to be prudent.

- Otherwise, a walk $\omega = (\omega_0, \omega_1, \ldots, \omega_{n-1}, \omega_n)$ is prudent if

  - the walk $(\omega_0, \omega_1, \ldots, \omega_{n-1})$ is prudent, and
  - after deforming the lattice into the brickwork lattice with the property that the last

---

[23]By equivalent, we mean that they are the same when considered as *graphs*, and in particular there is a one-to-one mapping between SAWs on each lattice. Measurements like end-to-end distance will change from one lattice to another.

[24]We could of course define prudent walks in the more obvious way, by simply forbidding any steps towards occupied vertices. It is not difficult to show that a walk which meets that criterion is also prudent by our definition, so that our prudent walks are a superset of that class.
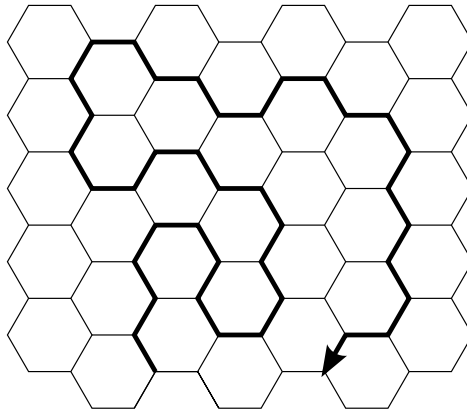
Figure 3.23: A prudent walk on the honeycomb lattice.

two steps of ω are collinear, we have that the last step of ω was not directed towards an already occupied vertex.

In essence, after each step we deform the lattice so that the walk *can* step in a straight line, and we then apply the prudent condition.[25] See Figure 3.23 for an example.

Our sub-classification of walks into 1-sided, 2-sided, etc. also depends on the brickwork lattice orientations. We say that a prudent walk $\omega = (\omega_0, \omega_1, \ldots, \omega_{n-1}, \omega_n)$ on the honeycomb lattice is *k-sided*, with $1 \leq k \leq 6$, if

- the walk $(\omega_0, \omega_1, \ldots, \omega_{n-1})$ is k-sided, and

    - $k = 1$: in the vertical orientation there is no occupied vertex further east than ω_n;

    - $k = 2$: in the vertical orientation there is no occupied vertex further east than ω_n, or in the negative orientation there is no occupied vertex further north-east than ω_n;

    - $k = 3$: in the vertical orientation there is no occupied vertex further east than ω_n, or in the negative orientation there is no occupied vertex further north-east than ω_n, or in the positive orientation there is no vertex further north-west than ω_n;

    - we continue this in a counter-clockwise direction for $k = 4, 5, 6$.

We will discuss here only the 1- and 2-sided models, as these are the ones we have been able to solve.

1-sided prudent walks

We split the walks into two groups $\mathscr{L}$ and $\mathscr{R}$, depending on which type of vertex they end at. Walks ending at a vertex with an edge in the east direction are in $\mathscr{R}$, and the others are in $\mathscr{L}$.

---

[25]We thank one of the examiners for the following alternative phrasing of the definition of a prudent walk: Any walk of length 0 or 1 is prudent; otherwise, a SAW of length $n \geq 2$ is prudent if its last two steps could be repeated indefinitely without creating self-intersections.
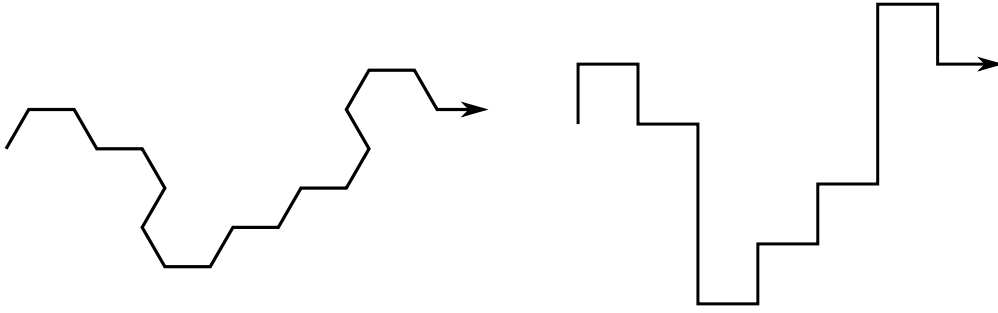
Figure 3.24: A 1-sided prudent walk on the honeycomb lattice, and the same walk on the vertical brickwork lattice.

(So the walk illustrated in Figure 3.24 is in $\mathscr{L}$.) We will arbitrarily assign the starting vertex to be an $\mathscr{L}$ vertex.[26] So define the two generating functions

$$L(t) = \sum_{n \geq 0} l_n t^n$$

$$R(t) = \sum_{n \geq 0} r_n t^n$$

where l_n (resp. r_n) is the number of 1-sided prudent walks of length n in $\mathscr{L}$ (resp. $\mathscr{R}$).

Lemma 3.58. *The generating functions $L(t)$ and $R(t)$ are given by*

$$L(t) = \frac{1 + t^2}{1 - 3t^2}, \qquad R(t) = \frac{2t}{1 - 3t^2}, \tag{3.164}$$

and so the generating function of 1-sided prudent walks is

$$H^{(1)}(t) = L(t) + R(t) = \frac{(1 + t)^2}{1 - 3t^2} \tag{3.165}$$

$$= 1 + 2t + 4t^2 + 6t^3 + 12t^4 + 18t^5 + 36t^6 + 54t^7 + 108t^8 + O(t^9).$$

The number $h_n^{(1)}$ of 1-sided prudent walks of length n satisfies

$$h_n^{(1)} \sim \varkappa_n 3^{n/2},$$

where

$$\varkappa_n = \begin{cases} 4/3 & n \text{ even} \\ \frac{2}{\sqrt{3}} & n \text{ odd.} \end{cases}$$

Proof. We count walks according to their last east step. Walks in $\mathscr{L}$ with no east step have the generating function

$$1 + \frac{2t^2}{1 - t^2},$$

---

[26]This choice does affect the generating functions slightly, but will not change the dominant singular behaviour.

while those in $\mathscr{R}$ have the generating function

$$\frac{2t}{1-t^2}.$$

Otherwise, the last east step of a walk must have been appended to a walk in $\mathscr{R}$. The generating function of walks in $\mathscr{L}$ with at least one east step is then

$$R(t) \cdot t \cdot \left(1 + \frac{2t^2}{1-t^2}\right),$$

and walks in $\mathscr{R}$ have the generating function

$$R(t) \cdot t \cdot \frac{2t}{1-t^2}.$$

Thus, we have

$$L(t) = 1 + \frac{2t^2}{1-t^2} + \left(t + \frac{2t^3}{1-t^2}\right) R(t)$$

$$R(t) = \frac{2t}{1-t^2} + \frac{2t^2}{1-t^2} R(t), \tag{3.166}$$

and the solution to these equations is exactly the result of the lemma.

The function $H^{(1)}(t)$ has two dominant singularities: simple poles at $t = \pm 1/\sqrt{3}$. This leads to the asymptotic behaviour

$$h_n^{(1)} \sim \left(\frac{2}{3} + \frac{1}{\sqrt{3}}\right)(\sqrt{3})^n + \left(\frac{2}{3} - \frac{1}{\sqrt{3}}\right)(-\sqrt{3})^n,$$

which is equivalent to the second part of the lemma. ∎

2-sided prudent walks

Based on the definition of 2-sided prudent walks, we further subdivide these walks into two classes: $\mathscr{A}$ contains walks which, in the vertical orientation, have not stepped further east than their endpoint, and $\mathscr{B}$ contains walks which, in the negative orientation, have not stepped further north-east than their endpoint. We will thus have to work with four generating functions, which will count walks in $\mathscr{A} \cap \mathscr{L}$, $\mathscr{A} \cap \mathscr{R}$, $\mathscr{B} \cap \mathscr{L}$ and $\mathscr{B} \cap \mathscr{R}$. (The fact that we have to start at either an $\mathscr{L}$ or an $\mathscr{R}$ vertex prevents us from being able to exploit a reflective symmetry between $\mathscr{A}$ and $\mathscr{B}$.) We define them as

$$A_L(t; u) = \sum_{n,i \geq 0} al_{n,i} t^n u^i \qquad\qquad B_L(t; u) = \sum_{n,i \geq 0} bl_{n,i} t^n u^i$$

$$A_R(t; u) = \sum_{n,i \geq 0} ar_{n,i} t^n u^i \qquad\qquad B_R(t; u) = \sum_{n,i \geq 0} br_{n,i} t^n u^i$$

where
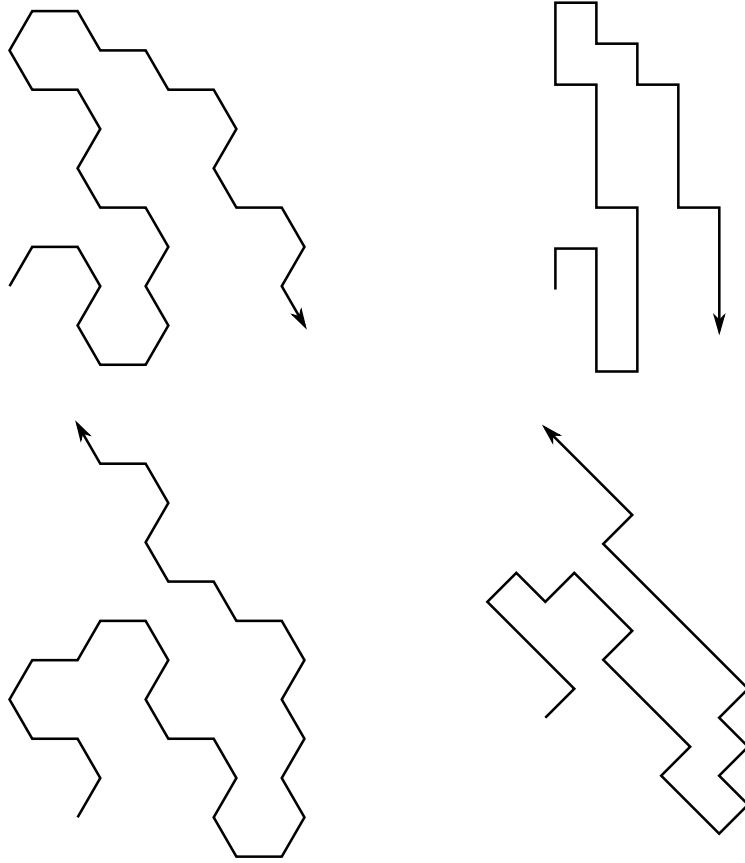
Figure 3.25: Top: A 2-sided prudent walk in $\mathscr{A}$, on the honeycomb and vertical brickwork lattice. For this walk $i = 2$. Bottom: A 2-sided prudent walk in $\mathscr{B}$, on the honeycomb and negative brickwork lattice. For this walk $i = 3$.

- $al_{n,i}$ (resp. $ar_{n,i}$) is the number of 2-sided prudent walks in $\mathscr{A} \cap \mathscr{L}$ (resp. $\mathscr{A} \cap \mathscr{R}$) with length n and requiring a further i *north-east* steps in order to be a part of $\mathscr{B}$; and

- $bl_{n,i}$ (resp. $br_{n,i}$) is the number of 2-sided prudent walks in $\mathscr{B} \cap \mathscr{L}$ (resp. $\mathscr{B} \cap \mathscr{R}$) with length n and requiring a further i *east* steps in order to be a part of $\mathscr{A}$.

In Figure 3.25 we provide examples of 2-sided prudent walks in $\mathscr{A}$ and $\mathscr{B}$.

Lemma 3.59. *The generating functions $A_L(t;u), A_R(t;u), B_L(t;u)$ and $B_R(t;u)$ satisfy the equations*

$$A_L(t;u) = \frac{1}{1-t^2u} + \frac{tu(1-t^4)}{(u-t^2)(1-t^2u)}A_R(t;u) - \frac{t^3}{u-t^2}A_R(t;t^2)$$
$$+ B_L(t;t^2) - (1-t^2)B_L(t;0) \quad (3.167)$$

$$\left(1 - \frac{t^2(1-t^2)(1+u)}{(u-t^2)(1-t^2u)}\right)A_R(t;u) = \frac{t}{1-t^2u} - \frac{t^2}{u-t^2}A_R(t;t^2) + tB_L(t;t^2) \quad (3.168)$$

132

$$\left(1 - \frac{t^2(1-t^2)(1+u)}{(u-t^2)(1-t^2u)}\right) B_L(t;u) = \frac{1}{1-t^2u} - \frac{t^2}{u-t^2}B_L(t;t^2) + tA_R(t;t^2) \qquad (3.169)$$

$$B_R(t;u) = t + \frac{tu}{1-t^2u} + \frac{tu(1-t^4)}{(u-t^2)(1-t^2u)}B_L(t;u) - \frac{t^3}{u-t^2}B_L(t;t^2)$$
$$+ A_R(t;t^2) - (1-t^2)A_R(t;0). \quad (3.170)$$

Proof. Though we've not properly defined the hexagon of a walk, and thus we cannot really talk about the *sides* of the hexagon, it is still clear what constitutes an inflating step in this model: any step which moves one of the relevant boundaries (the eastern boundary in the vertical orientation or the north-eastern boundary in the negative orientation) is inflating. So an inflating step must be an east step from a walk in $\mathscr{A} \cap \mathscr{R}$ or a north-east step from a walk in $\mathscr{B} \cap \mathscr{L}$.

- $\mathscr{A} \cap \mathscr{L}$: Walks with no inflating steps are empty or a sequence of south-east and south-west steps; their generating function is

$$\frac{1}{1-t^2u}.$$

Walks with last inflating step east are formed by appending steps to walks in $\mathscr{A} \cap \mathscr{R}$. Their generating function is

$$\sum_{n,i\geq0} ar_{n,i}t^n \cdot t\left(\sum_{k=0}^{i} t^{2k}u^{i-k} + \sum_{l=1}^{\infty} t^{2l}u^{i+l}\right)$$

(The sum over k considers walks with north-east and north-west steps following the east step; the sum over l is for walks with south-east and south-west steps following the east step.)

$$= t\sum_{n,i\geq0} ar_{n,i}t^n \cdot \left(\frac{u^{i+1}-t^{2i+2}}{u-t^2} + \frac{t^2u^{i+1}}{1-t^2u}\right)$$
$$= \frac{t}{u-t^2}[uA_R(t;u) - t^2A_R(t;t^2)] + \frac{t^3u}{1-t^2u}A_R(t;u).$$

Walks with last inflating step north-east are formed by appending steps to walks in $\mathscr{B} \cap \mathscr{R}$. These steps must be a sequence of east and south-east steps, unless the walk in $\mathscr{B} \cap \mathscr{R}$ was also in $\mathscr{A}$: in that case, we can only append a single north-west step. The generating function is thus

$$B_L(t;t^2) - B_L(t;0) + t^2B_L(t;0).$$

Adding the above three generating functions together gives $A_L(t;u)$, and rearranging what results gives (3.167).

133

- $\mathscr{A} \cap \mathscr{R}$: The same three options are applicable here – for the most part, the only difference is in the parity of the number of steps we append after the inflating step. The problem of the $\mathscr{B}$ walks which are also in $\mathscr{A}$ is not an issue here.

- $\mathscr{B} \cap \mathscr{L}$: Walks with no inflating steps are empty or a sequence of west and north-west steps; their generating function is

$$\frac{1}{1 - t^2 u}.$$

Walks with last inflating step north-east are formed by appending steps to walks in $\mathscr{B} \cap \mathscr{L}$. Their generating function is

$$\sum_{n,i \geq 0} bl_{n,i} t^n \cdot t \left(\sum_{k=1}^{i} t^{2k-1} u^{i-k} + \sum_{l=1}^{\infty} t^{2l-1} u^{i+l-1} \right)$$

$$= t \sum_{n,i \geq 0} bl_{n,i} t^n \cdot \left(\frac{t(u_i - t^{2i})}{u - t^2} + \frac{t u^i}{1 - t^2 u} \right)$$

$$= \frac{t^2}{u - t^2} [B_L(t; u) - B_L(t; t^2)] + \frac{t^2}{1 - t^2 u} B_L(t; u)$$

Walks with last inflating step east are formed by appending steps to walks in $\mathscr{A} \cap \mathscr{R}$. These steps must be a sequence of north-east and north-west steps. The generating function is thus

$$t A_R(t; t^2).$$

Adding together the above three generating functions gives $B_L(t; u)$, and (3.169) follows.

- $\mathscr{B} \cap \mathscr{R}$: The same three options are applicable here – for the most part, the only difference is in the parity of the number of steps we append after the inflating step. We do have to deal with the problem of walks in $\mathscr{A}$ which are also in $\mathscr{B}$. Also, a single south-east step will also be counted by B_R. ∎

Lemma 3.60. *Define*

$$J \equiv J(t) = \frac{1 - t^2 + 2t^4 - \sqrt{1 - 2t^2 - 3t^4 + 4t^8}}{2t^2}.$$

Then the generating function $H^{(2)}(t)$ of 2-sided prudent walks is

$$H^{(1)}(t) = \frac{(1+t)(J - t - 3t^2 - 2t^3 + t^5 + 2tJ + 3t^4 J + 2t^5 J - t^2 J^2 - 2t^3 J^2)}{(1 - 3t^2)(1 - t^2 J)(J - t - t^2)} \tag{3.171}$$

$$= \frac{(1+t)\left(2 + t - 5t^2 - 5t^3 - 3t^4 + 4t^5 + 2t^6 + t(1-t)\sqrt{1 - 2t^2 - 3t^4 + 4t^8}\right)}{2(1 - 3t^2)(1 - t - t^2 - t^3 + t^5)}$$

$$= 1 + 3t + 5t^2 + 9t^3 + 16t^4 + 29t^5 + 51t^6 + 92t^7 + 163t^8 + O(t^9).$$

Proof. The kernels in (3.168) and (3.169) are cancelled by the substitution $u = J$. Upon performing this substitution we have two equations in $A_R(t;t^2)$ and $B_L(t;t^2)$; these can be solved to give

$$A_R(t;t^2) = \frac{J(J - t^2)}{t(1 - t^2 J)(t^2 - t^4 + 2t^2 J - J^2)}$$

and

$$B_L(t;t^2) = \frac{(J - t^2)(1 - t^2 + J)}{(1 - t^2 J)(t^2 - t^4 + 2t^2 J - J^2)}.$$

These can be substituted into (3.168) and (3.169) to give solutions for $A_R(t;1)$, $A_R(t;0)$, $B_L(t;1)$ and $B_L(t;0)$; these in turn can be substituted into (3.167) and (3.170) to give $A_L(t;1)$ and $B_R(t;1)$.

The overall solution is then

$$H^{(1)}(t) = A_L(t;1) + A_R(t;1) + B_L(t;1) + B_R(t;1) - A_R(t;0) - B_L(t;0). \qquad \blacksquare$$

As usual with algebraic functions, the dominant singularity of $H^{(1)}(t)$ is relatively easy to locate: it is a simple pole at a root of $J - t - t^2 = 0$.

Corollary 3.61. *The number $h_n^{(2)}$ of 2-sided prudent walks of length n is asymptotically*

$$h_n^{(2)} \sim \varkappa \lambda^n,$$

where $\lambda = 1.77848\ldots$, a root of $1 - \lambda^2 - \lambda^3 - \lambda^4 + \lambda^5 = 0$, and $\varkappa = 1.69314\ldots$.

Recall that the number of SAWs on the honeycomb lattice is expected to behave like

$$c_n \sim A n^{\gamma - 1} \mu^n,$$

where $\gamma = 43/32$ and $\mu = \sqrt{2 + \sqrt{2}} = 1.84776\ldots$. Thus, we see that even the relatively simple model of 2-sided prudent walks on the honeycomb lattice has a growth rate remarkably close to that of general SAWs.

135

Chapter 4

Interacting polymer models

Recall from Chapter 1 that self-avoiding walks have long been considered as a canonical model for long-chain polymers in solution [93]. The standard model associates a weight (or fugacity) x with each step, or monomer.[1] If c_n is the number of n-step SAWs (equivalent up to translation), the length generating function is

$$Z(x) = \sum_{n \geq 0} c_n x^n.$$

As we saw in Chapter 1, it has long been known that the limit

$$\log \mu = \lim_{n \to \infty} \frac{1}{n} \log c_n$$

exists, where the *growth constant* μ depends on the lattice being considered.

To model the behaviour of a polymer interacting with an impenetrable surface, we consider SAWs restricted to a half-space. Although this restriction limits the number of walk configurations, it has no effect on the exponential growth of the number of walks and hence the growth constant [118]. To model the monomer-surface interactions of an adsorbing polymer, we associate a fugacity e^α with each visit of the walk (edge or vertex) to the boundary of the half-space. If we define $c_n^+(m)$ as the number of n-step half-space walks beginning on the surface with m contacts on the surface, then the partition function for walks of length n is

$$Z_n(\alpha) = \sum_{m=0}^{n} c_n^+(m) e^{m\alpha},$$

where $\alpha = -\epsilon/k_B T$, where ϵ is the energy associated with a surface visit, T is the absolute temperature and k_B is Boltzmann's constant. For sufficiently large values of α, configurations with many contacts dominate the partition function and the walk is said to be in an *adsorbed* state; otherwise, the loss in configurational entropy dominates, and the walk is repelled by the surface and said to be in a *desorbed* state.

---

[1]In keeping with the notation of Chapters 2 and 3 and much of the wider literature, we will use the fugacity x when studying general SAWs and the fugacity t for solvable subclasses.

The free energy of the system is

$$\widehat{\varkappa}(\alpha) = \lim_{n \to \infty} \frac{1}{n} \log Z_n(\alpha), \tag{4.1}$$

which has been shown [62, 103] to exist for the hypercubic lattice $\mathbb{Z}^d$ for all α, for both the edge- and vertex-weighted models. (Later in this chapter we will adapt existing arguments for the hypercubic lattice to the honeycomb lattice, where we are able to conjecture and prove exact results.) The free energy $\widehat{\varkappa}(\alpha)$ is a continuous and convex function of α. When $\alpha < 0$ the surface fugacity has no effect on the free energy, and $\widehat{\varkappa}(\alpha) = \log \mu$ [118]. For $\alpha \geq 0$, it has been shown that

$$\widehat{\varkappa}(\alpha) \geq \max\{\log \mu, \alpha\}.$$

This implies the existence of a critical value α_c with $0 \leq \alpha_c \leq \log \mu$. At this critical point the free energy $\widehat{\varkappa}(\alpha)$ is non-analytic, and so we say that α_c is the location of the *adsorption phase transition:* for $\alpha < \alpha_c$, the polymer is desorbed, while for $\alpha > \alpha_c$ the polymer is adsorbed.[2]

Closely related to the free energy is the mean density of steps in the surface; for walks of length n, this is given by

$$\delta_n(\alpha) = \frac{1}{n} \frac{\sum_m m c_n^+(m) \mathrm{e}^{m\alpha}}{\sum_m c_n^+(m) \mathrm{e}^{m\alpha}} = \frac{1}{n} \frac{\partial \log Z_n(\alpha)}{\partial \alpha}.$$

In the limit of infinitely long polymers, this density tends to[3]

$$\delta(\alpha) = \frac{\partial \widehat{\varkappa}(\alpha)}{\partial \alpha}.$$

(We note here that the symbol ρ, rather than δ, is frequently used in the literature to denote the density of steps in the surface. However, in this chapter we will sometimes use ρ to denote the radius of convergence of a power series (another common convention).)

The function δ acts as an order parameter for the system, and signals the onset of a phase transition. If $\alpha < \alpha_c$ then the density is 0, while the density is non-zero for $\alpha > \alpha_c$. If the adsorption transition is *second-order* then δ is continuous for all α, while a *first-order* transition results in a jump discontinuity at α_c.

The behaviour of the free energy at the critical point α_c is thought to be characterised by the *crossover exponent* ϕ, satisfying

$$\widehat{\varkappa}(\alpha) - \widehat{\varkappa}(\alpha_c) \sim c(\alpha - \alpha_c)^{1/\phi} \quad \text{as } \alpha \to \alpha_c^+$$

for some constant c.

---

[2] The existence of one point of non-analyticity is guaranteed; it is possible that there may be others in the region $\alpha > \alpha_c$. If that were the case, then α_c would still be the location of the adsorption transition; the other critical points might then signal other types of transitions.

[3] The exchange of the limit and the derivative is possible thanks to the convexity of $\widehat{\varkappa}(\alpha)$, see for instance [112, Theorem B7].

Instead of directly working with the partition functions Z_n for the models in question, we will consider the generating function $Z(x; \alpha) = \sum_n Z_n(\alpha) x^n$. As we saw in Chapter 3, a central tenet of analytic combinatorics [44] is that the *dominant singularities* (the points of non-analyticity closest to the origin), and thus the radius of convergence, of a generating function determine the asymptotic behaviour of its coefficients. Specifically, if for a given α the radius of convergence of $Z(x; \alpha)$ occurs at $x = x_c(\alpha)$, then

$$\widehat{x}(\alpha) = \lim_{n \to \infty} \frac{1}{n} \log Z_n(\alpha) = -\log x_c(\alpha). \tag{4.2}$$

For many models (such as those considered in Section 4.3), it is much easier to determine the generating function than to compute the individual partition functions Z_n for each n. But since (4.2) enables us to determine the free energy $\widehat{x}(\alpha)$ of a model directly from its generating function, we need never consider the Z_n anyway.

For ease of notation we will henceforth use the surface fugacity $y = e^\alpha$, and consider the free energy, etc. as functions of y. The above arguments are still valid, with the critical fugacity occuring at $y_c = e^{\alpha_c}$. We will write $x(y) = \widehat{x}(\log y)$.

This chapter is comprised of three sections. In Section 4.1 we develop a generalisation of the method of Duminil-Copin and Smirnov [106, 34] (see also Section 2.1), and use it to study adsorbing self-avoiding walks on the honeycomb lattice. In particular, we obtain predictions for the critical surface fugacity y_c for two orientations of the honeycomb lattice, which agree with conjectures of Batchelor and Yung [7] and Batchelor, Bennett-Wood and Owczarek [6]. For one of these orientations, we present a proof of that critical value. This proof depends on the generating function of self-avoiding bridges which span a strip of width T disappearing in the limit $T \to \infty$. A proof of that fact, due primarily to Hugo Duminil-Copin, is given in Appendix A. For the other orientation, we expect the proof to be very similar, and intend to complete and publish it in the near future.

In Section 4.2 we investigate how the identities used in Section 4.1 can be adapted to the square and triangular lattices, in a similar manner to the methodology of Section 2.2. We are able to calculate numerical estimates of several critical surface fugacities, and find that they not only agree well with existing estimates but exceed their numerical precision by several orders of magnitude.

Finally, in Section 4.3 we consider *solvable* models of polymer adsorption. We briefly review existing models, which are traditionally based on *directed* walks. We then introduce new solvable models based on *prudent* walks, which do not have a directedness restriction, and investigate their critical behaviour.

4.1   Exact results for the honeycomb lattice

Recall from Section 2.1 that in 1982 Nienhuis [89] showed that, for $n \in [-2, 2]$, the $O(n)$ loop model on the honeycomb lattice could be mapped to a solid-on-solid model, from which he

was able to derive the critical points and exponents (subject to some plausible assumptions). In particular, he was able to predict the value of the growth constant for SAWs on the honeycomb lattice; this result was proved in 2010 by Duminil-Copin and Smirnov [34] (the proof was outlined in Section 2.1).

The results of Nienhuis and Duminil-Copin and Smirnov were concerned with bulk systems. Interesting surface phenomena can also be studied if one considers the n-vector model in a half-space, with vertices in the surface having an associated fugacity. If this fugacity is made repulsive, adsorption onto the surface will be energetically unfavourable; if the fugacity is made attractive, adsorption becomes increasingly favoured. The adsorption transition is an example of a *special* surface transition [14].

In 1995 Batchelor and Yung [7] extended Nienhuis's work to the adsorption problem described above, and making similar assumptions to Nienhuis conjectured the value of the critical surface fugacity for the two-dimensional honeycomb lattice n-vector model using the integrability of an underlying lattice model. In this section we show that the key identity proved by Smirnov [106] for the $O(n)$ model with $n \in [-2, 2]$ can be generalised to a semi-infinite system with a surface fugacity. We use this to prove a generalisation of the identity of Duminil-Copin and Smirnov linking certain generating functions in finite domains to include a surface fugacity. The contribution of one of these generating functions vanishes at a particular value of the surface fugacity, and this critical value coincides with the conjectured value of the critical fugacity by Batchelor and Yung. That is, we have an independent prediction of the value of the critical surface fugacity:

Conjecture 4.1. *For the $O(n)$ loop model on a hexagonal lattice half-plane with $n \in [-2, 2]$, associate a fugacity $x_c(n) = 1/\sqrt{2 + \sqrt{2 - n}}$ with occupied vertices and an additional fugacity y with occupied vertices on the boundary. Then the model undergoes a special surface transition at*

$$y = y_c(n) = 1 + \frac{2}{\sqrt{2 - n}}.$$

At $n = 0$ this model becomes that of a self-avoiding walk interacting with an impenetrable surface. For this model, it is known that the critical value of the (attractive) fugacity exists, and at this point a macroscopic fraction of steps in an arbitrarily long walk becomes adsorbed onto the surface. (Below this critical value the expected fraction of steps adsorbed onto the surface is zero). In general we do not even have conjectures for the exact values of the critical fugacities on various lattices; instead, numerical estimates using series analysis and Monte Carlo methods are the best current results.

As indicated above (and indeed by the title of this section), however, the honeycomb lattice is special. In that case, the critical value of the bulk fugacity x_c has been proven in [34], and the critical surface fugacity was conjectured in [7]. In this thesis we go further, and prove the exact value:
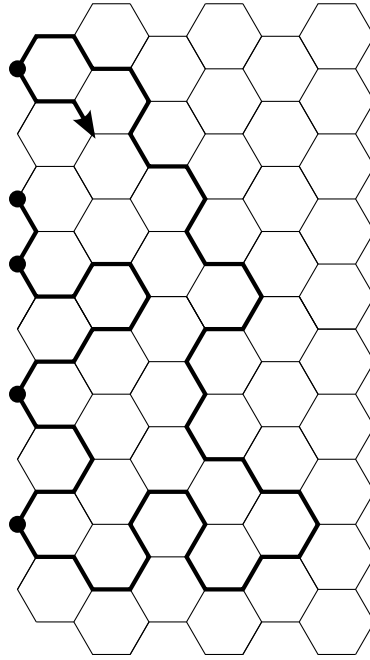
Figure 4.1: A SAW on the honeycomb lattice interacting with an impenetrable surface (the left hand side of the picture). This orientation of the surface (i.e. with lattice edges perpendicular to the surface) is the one considered in Theorem 4.2.

Theorem 4.2. *The critical surface fugacity for self-avoiding walks on the hexagonal lattice, oriented so that the lattice contains edges perpendicular to the surface, is*

$$y_c = 1 + \sqrt{2}.$$

See Figure 4.1 for an example of a SAW interacting with the impenetrable surface in this orientation.

In order to prove this result we require that two generating functions, originally defined in [34] and generalised here to accommodate a surface fugacity, tend to 0 as the length and width of a lattice segment grows large. In Section 2.1 we proved that at $y = 1$, one of these functions does indeed disappear in the appropriate limit, subject to some simple results being adapted from the square to the honeycomb lattice.[4] In Subsection 4.1.2 we demonstrate that this adaptation can indeed be done, and furthermore we extend the results to general y. In Appendix A we present the proof, due to Hugo Duminil-Copin, that the other generating function also disappears in the appropriate limit.

These results, once adapted to the square and triangular lattices, also play a key role in the methodology of Sections 2.2 and 4.2.

Figure 4.2: A configuration γ on a finite domain, with the weighted vertices on the right hand boundary indicated. The contribution of γ to $F(z)$ is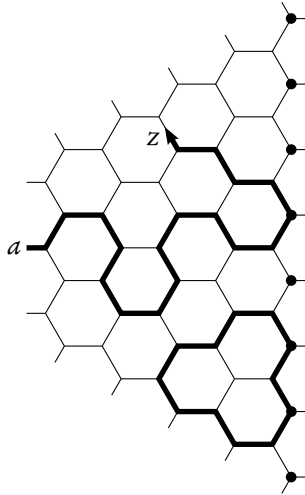 $e^{-2i\sigma\pi/3}x^{30}y^3n$. Note that in this picture, the entire lattice has been rotated, so that the surface is still oriented in the same way as Figure 4.1.

4.1.1   Identity in the presence of a boundary

Let H be the set of mid-edges on a half-plane of the honeycomb lattice, and define the *surface* to be the set of vertices in the half-plane which are adjacent to only two other vertices. Recall from Section 2.1 the definition of a *domain* Ω and a *configuration* γ. In this section we insist that at least one of the vertices of $V(\Omega)$ is in the surface (otherwise adsorption onto the surface would of course be impossible). Recall also that we denote by $|\gamma|$ the number of vertices in γ and by $c(\gamma)$ the number of closed loops. We introduce here a new measurement, $v(\gamma)$, which is the number of vertices of γ lying in the surface.

Define the following observable: for $a \in \partial\Omega, z \in \Omega$, set

$$F(\Omega, a, z; x, y, n, \sigma) := F(z) = \sum_{\gamma \subset \Omega : a \to z} e^{-i\sigma\, W(\gamma : a \to z)} x^{|\gamma|} y^{v(\gamma)} n^{c(\gamma)},$$

where the sum is over all configurations $\gamma \subset \Omega$ for which the SAW component runs from the mid-edge a to a mid-edge z (we say that γ *ends* at z). We denote by $W(\gamma : a \to z)$ the winding angle of that self-avoiding walk. See Figure 4.2 for an example.

Recall Lemma 2.1, where Smirnov [106] showed that, at $y = 1$ and for special values of x and σ, the observable F satisfies an identity of the form

$$(p - v)F(p) + (q - v)F(q) + (r - v)F(r) = 0,$$

where p, q, r are the three mid-edges adjacent to an arbitrary $v \in V(\Omega)$.

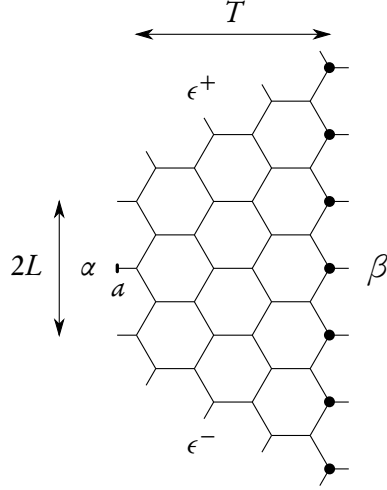[4]Specifically, we proved that $E_{T,L}(x_c, 1) \to 0$ as $L \to \infty$.

Figure 4.3: Finite patch $S_{3,1}$ of the hexagonal lattice with a boundary. Contours, possibly closed, of the $O(n)$ model run from mid-edge to mid-edge acquiring a weight x for each step, and a weight y for each contact (shown as a black disc) with the right hand side boundary. The SAW component of a configuration starts on the central mid-edge of the left boundary (shown as a).

In Section 2.1 we outlined the proof by Duminil-Copin and Smirnov [34] that the growth constant of the self-avoiding walk is equal to $x_c^{-1} = (2\cos(\pi/8)) = \sqrt{2 + \sqrt{2}}$. Recall that the proof involved a special domain $S_{T,L}$ (see Figure 2.3) and generating functions of SAWs ending on the different sides of this domain.

Here we generalise their construction to include a boundary weight. As shown in Figure 4.3, we will identify the surface with the β boundary of $S_{T,L}$.[5]

Let us define the following generating functions:

$$A_{T,L}(x,y) := \sum_{\substack{\gamma \subset S_{T,L}: \\ a \to \alpha \setminus \{a\}}} x^{|\gamma|} y^{\nu(\gamma)} n^{c(\gamma)},$$

$$B_{T,L}(x,y) := \sum_{\substack{\gamma \subset S_{T,L}: \\ a \to \beta}} x^{|\gamma|} y^{\nu(\gamma)} n^{c(\gamma)},$$

$$E_{T,L}(x,y) := \sum_{\substack{\gamma \subset S_{T,L}: \\ a \to \epsilon^+ \cup \epsilon^-}} x^{|\gamma|} y^{\nu(\gamma)} n^{c(\gamma)},$$

where the sums are over all configurations for which the SAW component runs from a to the α,

---

[5]It may seem strange to put the surface weights on the β boundary instead of the α boundary. While it is indeed possible to derive an identity with the weights on the α boundary, at $n = 0$ and for some values of y in the region we need to consider, some of the coefficients in that identity are negative. As we saw in Section 2.1, it is of crucial importance that the coefficients are non-negative, so that we can make statements about upper bounds, limits, etc. of the generating functions.

β or ϵ^+, ϵ^- boundaries respectively. Furthermore define the special generating function

$$D_{T,L}(x,y) := \sum_{\substack{\gamma \subset S_{T,L}: \\ a \to a}} x^{|\gamma|} y^{\nu(\gamma)} n^{c(\gamma)}$$

which sums over configurations comprising *only* closed loops inside $S_{T,L}$; that is, configurations whose self-avoiding walk component is the empty walk $a \to a$.

Proposition 4.3. *Let $n = 2\cos\theta$ with $\theta \in [0, \pi]$ and define*

$$x_c = \frac{1}{2}\sec\left(\frac{\pi \pm \theta}{4}\right) = (2 \mp \sqrt{2-n})^{-1/2}.$$

Then

$$D_{T,L}(x_c, y) = \cos\left(\frac{3(\pi \pm \theta)}{4}\right) A_{T,L}(x_c, y) + \cos\left(\frac{\pi \pm \theta}{2}\right) E_{T,L}(x_c, y) + \frac{y^* - y}{y(y^* - 1)} B_{T,L}(x_c, y),$$

(4.3)

where

$$y^* = \frac{1}{1 - 2x_c^2} = 1 + \frac{1}{\cos((\pi \pm \theta)/2)}, \quad \text{or equivalently,} \quad y^* x_c^2 = \mp(2-n)^{-1/2}.$$

Proof. Let p_v, q_v, r_v be the mid-edges adjacent to a vertex v. We compute the sum

$$\sum_{v \in V(S_{T,L})} (p_v - v)F(p_v) + (q_v - v)F(q_v) + (r_v - v)F(r_v)$$

(4.4)

in two ways. (As in Lemma 2.1, we calculate $(p_v - v)$, etc. by considering p_v, q_v, r_v and v as points in the complex plane.) Firstly, any internal mid-edge (i.e. any mid-edge not in $\partial S_{T,L}$) will contribute to two terms in the sum, and these two terms will cancel. Thus we are left with precisely the contributions of those mid-edges in $\partial S_{T,L}$:

$$-\sum_{z \in \alpha} F(z) + \bar{j}\sum_{z \in \epsilon^-} F(z) + j\sum_{z \in \epsilon^+} F(z) + \sum_{z \in \beta} F(z).$$

(4.5)

On the other hand, if we set

$$\sigma = \frac{\pi \mp 3\theta}{4\pi}$$

then the identity (2.3) actually holds for any non-weighted vertex v, irrespective of the value of y. (The proof is exactly the same as for Lemma 2.1.) So the only vertices in $V(S_{T,L})$ which make a non-zero contribution to (4.4) are the weighted vertices on the β boundary. Since it is impossible for a configuration to occupy all three mid-edges adjacent to a weighted vertex, we need only consider the cases illustrated in Figure 4.4. If the leftmost configurations in Figure 4.4 are denoted γ_1 and γ_2 respectively, then the contributions from each triplet are

$$x_c^{|\gamma_1|} y^{\nu(\gamma_1)} n^{c(\gamma_1)}(\bar{j}\lambda + x_c y j \lambda^2 + x_c y) = -x_c^{|\gamma_1|} y^{\nu(\gamma_1)} n^{c(\gamma_1)}(y-1)\bar{j}\lambda,$$

(4.6)

$$x_c^{|\gamma_2|} y^{\nu(\gamma_2)} n^{c(\gamma_2)}(j\bar{\lambda} + x_c y \bar{j}\bar{\lambda}^2 + x_c y) = -x_c^{|\gamma_2|} y^{\nu(\gamma_2)} n^{c(\gamma_2)}(y-1)j\bar{\lambda}.$$

(4.7)
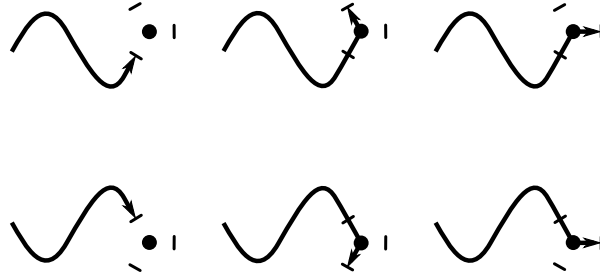
Figure 4.4: The two groupings of walks ending adjacent to a boundary vertex. The top three lead to (4.6), and the bottom three lead to (4.7).

(As in the proof of Lemma 2.1, we define $\lambda = e^{-i\sigma\pi/3}$ and $j = e^{2i\pi/3}$.) Now every configuration ending at a mid-edge $z \in \beta$ can be obtained uniquely by adding a step to a γ_1 or γ_2 configuration. Moreover every γ_1 configuration can be reflected in the horizontal axis to obtain a γ_2 configuration (and vice versa). Using these facts together with (4.6) and (4.7), we find that the contribution to (4.4) from weighted vertices is

$$-\frac{y-1}{2x_c y}(j\bar{\lambda} + \bar{j}\lambda)\sum_{z\in\beta}F(z).\tag{4.8}$$

Equating (4.5) and (4.8) gives

$$0 = -\sum_{z\in\alpha}F(z) + \bar{j}\sum_{z\in\epsilon^-}F(z) + j\sum_{z\in\epsilon^+}F(z) + \left(1 + \frac{y-1}{2x_c y}(j\bar{\lambda} + \bar{j}\lambda)\right)\sum_{z\in\beta}F(z).\tag{4.9}$$

As with the $y = 1$ case, we can write $\sum_{z\in\beta}F(z) = B_{T,L}(x,y)$. We also have that

$$\sum_{z\in\alpha}F(z) = D_{T,L}(x,y) + \frac{1}{2}(\lambda^3 + \bar{\lambda}^3)A_{T,L}(x,y),\tag{4.10}$$

where the $D_{T,L}$ term arises because the empty walk is included in the sum on the left. Also,

$$\bar{j}\sum_{z\in\epsilon^-}F(z) + j\sum_{z\in\epsilon^+}F(z) = \frac{1}{2}(\bar{j}\bar{\lambda}^2 + j\lambda^2)E_{T,L}(x,y).\tag{4.11}$$

With these definitions it follows from (4.9) that

$$D_{T,L}(x,y) = \cos\left(\frac{3(\pi\pm\theta)}{4}\right)A_{T,L}(x_c,y) + \cos\left(\frac{\pi\pm\theta}{2}\right)E_{T,L}(x_c,y)$$
$$+ \left(1 - \frac{1}{2}(1-y^{-1})x_c^{-2}\right)B_{T,L}(x_c,y),$$

from which the proposition follows. ∎

The identity (4.3) can be rewritten in the form

$$1 = \cos\left(\frac{3(\pi\pm\theta)}{4}\right)A_{T,L}^*(x_c,y) + \cos\left(\frac{\pi\pm\theta}{2}\right)E_{T,L}^*(x_c,y) + \frac{y^* - y}{y(y^* - 1)}B_{T,L}^*(x_c,y)\tag{4.12}$$

where

$$A^*_{T,L}(x,y) := \frac{A_{T,L}(x,y)}{D_{T,L}(x,y)}, \ B^*_{T,L}(x,y) := \frac{B_{T,L}(x,y)}{D_{T,L}(x,y)}, \text{ and } E^*_{T,L}(x,y) := \frac{E_{T,L}(x,y)}{D_{T,L}(x,y)}$$

can be considered as the "normalised" versions of A, B and E. Note that for the SAW case, $n = 0$ and $D_{T,L}(x,y) = 1$.

4.1.2 Self-avoiding walks in a strip

In this subsection and the next, we specialise to $n = 0$, corresponding to self-avoiding walks. Here we will discuss some results for SAWs on strips of the honeycomb lattice, which can be adapted in a very straightforward manner from existing results for the square lattice.

The usual model of surface-interacting walks considers walks originating in a surface and interacting with monomers or edges in that surface. One way to study such systems is to consider interacting walks in a strip, and then to take the limit as the strip width becomes infinite. Clearly, if one studies walks in a strip, it is possible to consider interactions with one or both boundaries.

To remain consistent with the previous subsection, we will consider walks which begin and end on mid-edges of the lattice. We take a vertical strip of the honeycomb lattice, with walks 'entering' with a horizontal half-step to the left-hand boundary (in exactly the same manner as the previous subsection – see for example Figure 4.2). We will associate a fugacity y with vertices on the left-hand boundary and a fugacity z with those on the right. We consider three types of walks in the strip of width T: those which start and end on the left boundary (loops), those which start on the left and end on the right (bridges), and those which start on the left but may end anywhere in the strip. (Clearly there is an analogy between the first two types and the walks counted by the A and B generating functions defined in the previous subsection.)

Define then the quantities $a_{T,n}(i,j)$, $b_{T,n}(i,j)$ and $c_{T,n}(i,j)$ which count the number of n-step walks in a strip of width T which start on the left boundary, visit i vertices on the left boundary and j vertices on the right, and end, respectively, on the left, right or anywhere in the strip. (Defined, of course, up to vertical translation.) The associated partition function is

$$A_{T,n}(y,z) = \sum_{i,j} a_{T,n}(i,j) y^i z^j$$

for loops, and we likewise define the partition functions $B_{T,n}(y,z)$ and $C_{T,n}(y,z)$ for bridges and general walks.

Proposition 4.4. *The free energy for general walks in a strip of width T, for $y, z > 0$, is defined to be*

$$\varkappa_T(y,z) = \lim_{n \to \infty} \frac{1}{n} \log C_{T,n}(y,z).$$

It is finite, continuous, and non-decreasing in y and z, and is a convex function of $\log y$ *and* $\log z$. *Moreover,* $\varkappa_T$ *is also given by*

$$\varkappa_T(y,z) = \lim_{n\to\infty} \frac{1}{n}\log A_{T,n}(y,z) = \lim_{n\to\infty} \frac{1}{n}B_{T,n}(y,z).$$

Since $B_{T,n}(y,z) = B_{T,n}(z,y)$, *it follows that* $\varkappa_T(y,z) = \varkappa_T(z,y)$.

Proof. The equivalent results for the square lattice are presented in [115, Ch. 4 and 6]. There, the authors demonstrate a relationship (Lemma 4.4) between the partition functions of walks in a strip and *unfolded* walks in a strip. Unfolding a walk ω involves reflecting parts of ω in horizontal lines, in such a way as to obtain a new walk ω' with the property that the first and last points of ω' have minimal and maximal y-coordinates respectively.[6] Unfolded walks can thus be easily concatenated without creating self-intersections, and such concatenations can result in loops or bridges if desired.

Then, in Lemma 4.5, the authors show that the concatenation of unfolded walks leads to a submultiplicative sequence involving their partition functions, and using the same reasoning as that behind Lemma 1.1, show that the free energy of unfolded walks exists. Finally, in Corollary 4.7, the relationship between walks and unfolded walks is used to show that the free energy of walks in a strip is the same as that of unfolded walks, and also that of loops and bridges.

All of these arguments apply equally well to walks on the honeycomb lattice. In particular, they can be unfolded and concatenated in the same way as square lattice walks.

The convexity and continuity of $\varkappa_T$ on the square lattice is proved in [115, Lemma 6.6], and that result also transfers over to the honeycomb lattice without difficulty. ∎

The identity (4.3) we derived in the previous subsection was concerned with a domain $S_{T,L}$ with walks starting on the left side and surface weights on the right side. If we send $L \to \infty$ then the domain becomes a strip. We thus no longer need to consider surface weights on both sides of the strip, and so we now define the generating function

$$A_T(x,y) = \sum_n A_{T,n}(1,y)x^n$$

for loops in a strip of width T, and likewise the generating functions $B_T(x,y)$ and $C_T(x,y)$ for bridges and general walks.

In this case the usual relationship between the free energy and radius of convergence holds, and we see that for a given $y > 0$, the radius of convergence of $A_T(x,y)$ (and $B_T(x,y), C_T(x,y)$) is

$$\rho_T(y) = \exp(-\varkappa_T(1,y)).$$

Proposition 4.5. *For* $y > 0$,

$$\varkappa_T(1,y) > \varkappa_{T-1}(1,y).$$

---

[6]This is similar to the arguments behind Lemma 1.4.

Moreover, as $T \to \infty$,

$$\varkappa_T(1, y) \to \varkappa(y),$$

where $\varkappa(y) = \widehat{\varkappa}(\log y)$ and $\widehat{\varkappa}(\alpha)$ is the free energy of adsorbing SAWs in a half-plane of the honeycomb lattice, as defined in (4.1).

Proof. Since $\varkappa_T(1, y) = \varkappa_T(y, 1)$, in this proof we switch to loops in a strip of width T with surface interactions on the *left-hand* side of the strip. We say that a loop from mid-edge a to mid-edge b is *unfolded* if the ordinate $y(v)$ of every non-final vertex v of the walk satisfies $y(a) \leq y(v) < y(b)$. Two unfolded loops can be concatenated (after deleting the last half-edge of the first loop and the first half-edge of the second loop) to form a new unfolded loop. We say an unfolded loop is *irreducible* if it is not the concatenation of two (or more) unfolded loops. Define $\vec{A}_T(x, y)$ to be the generating function of unfolded loops in the strip of width T with interactions on the left-hand boundary of the strip.

As mentioned above, in [115, Corollary 4.7] it is shown that the free energy (and thus the radius of convergence) of unfolded loops is the same as that of all loops. Moreover, any unfolded loop can be uniquely decomposed into a sequence of irreducible loops, so we see that

$$\vec{A}_T(x, y) = \frac{P_T(x, y)}{1 - P_T(x, y)/(xy)},$$

where $P_T(x, y)$ is the generating function of irreducible loops in a strip of width T. (The $(xy)^{-1}$ term accounts for the half-edges which need to be removed when concatenating two loops and the surface contact which occurs in both parts.)

Because any generating function for SAWs in a finite-width strip is rational (this follows from transfer matrix methodology – see Subsection B.1.2 and e.g. [1]), and the radius of convergence of $\vec{A}_T(x, y)$ is $\rho_T(y)$, we must have that

$$\frac{P_T(\rho_T(y), y)}{y \rho_T(y)} = 1.$$

Now consider an irreducible unfolded loop w in a strip of width T which does not fit in a strip of width $T - 1$. For example, the shortest such loop has length $4T + 3$, and thus contributes a term $x^{4T+3} y^2$ in the series $P_T(x, y)$ (see Figure 4.5). Let $\tilde{P}_T(x, y) = P_T(x, y) - x^{4T+3} y^2$. The generating function of unfolded loops which do not contain such a loop as a factor is just

$$\tilde{A}_T(x, y) = \frac{\tilde{P}_T(x, y)}{1 - \tilde{P}_T(x, y)/(xy)},$$

and if its radius of convergence is given by $\tilde{\rho}_T(y)$ then we have

$$\frac{\tilde{P}_T(\tilde{\rho}_T(y), y)}{y \tilde{\rho}_T(y)} = 1.$$

But since $\tilde{P}_T(x, y) < P_T(x, y)$ for $x, y > 0$, it must be that $\rho_T(y) < \tilde{\rho}_T(y)$.
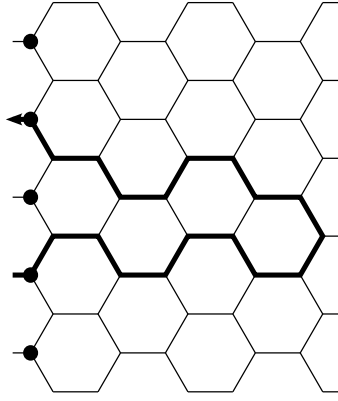
Figure 4.5: The shortest irreducible unfolded loop spanning a strip of width $T = 4$.

Finally, because w does not fit in a strip of width $T - 1$, no unfolded loop counted by $\vec{A}_{T-1}(x, y)$ can contain w as a factor. Hence, every object counted by $\vec{A}_{T-1}(x, y)$ is also counted by $\tilde{A}_T(x, y)$. So $\rho_{T-1}(y) \geq \tilde{\rho}_T(y)$, and thus $\rho_{T-1}(y) > \rho_T(y)$, or equivalently $\varkappa_T(1, y) > \varkappa_{T-1}(1, y)$.

The result that $\varkappa_T(1, y) \to \varkappa(y)$ on the square lattice is proved in [115, Thm. 6.5]; the proof for the honeycomb lattice is analogous. ∎

Finally, we prove a result which relates the critical surface fugacity y_c, which demarcates the adsorbed and desorbed phases of infinitely long polymers, to the radius of convergence $\rho_T(y)$ of walks in a strip.

Corollary 4.6. *There exists a unique $y_T > 0$ such that $\rho_T(y_T) = x_c = 1/\mu$, where $\mu = \sqrt{2 + \sqrt{2}}$ is the growth constant of the honeycomb lattice. The series (in y) $A_T(x_c, y)$, $B_T(x_c, y)$ and $C_T(x_c, y)$ have radius of convergence y_T, and $y_T \to y_c$ as $T \to \infty$.*

Proof. The existence of y_T follows from the intermediate value theorem: ρ_T is continuous, $\rho_T(1) > x_c$ and $\rho_T(y) \to 0$ as $y \to \infty$. The uniqueness of y_T follows from the convexity of $\varkappa_T(1, y)$ as a function of $\log y$: there cannot exist two distinct points y, y' with $\rho_T(y) = \rho_T(y') = x_c$. This also naturally implies that

$$\rho_T(y) < \rho_T(y_T) \iff y > y_T \quad \text{and} \quad \rho_T(y) > \rho_T(y_T) \iff y < y_T.$$

Since $\rho_T(y_T) = x_c$ by definition of y_T, it follows that y_T must be the radius of convergence of $A_T(x_c, y), B_T(x_c, y)$ and $C_T(x_c, y)$.

Now we prove that y_T decreases to y_c. Since $\rho_T(y) > \rho_{T+1}(y)$, we have $y_{T+1} < y_T$. Let $\bar{y} = \lim_{T \to \infty} y_T$. For $y \leq y_c$, we have $\rho_T(y) > x_c$, so it must be that $y_T > y_c$ for all T, and hence $\bar{y} \geq y_c$. Since $\bar{y} < y_T$, we have $\rho_T(\bar{y}) > x_c$, and thus $\rho(\bar{y}) \geq x_c$, where $\rho(y) = \exp(-\varkappa(y))$ is the radius of convergence of walks in the half-plane. But $\rho(y) < x_c$ for $y > y_c$ by definition of y_c, and thus $\bar{y} \leq y_c$. ∎

After we take $L \to \infty$ in our $S_{T,L}$ domain, the generating functions in our identity (4.12) (specifically, at $n = 0$ in the dilute regime) will be exactly $A_T(x_c, y)$ and $B_T(x_c, y)$ (with surface weights associated with vertices on the right boundary). It is Corollary 4.6 that relates these functions to the critical surface fugacity y_c, and thus enables us to complete the proof that $y_c = y^*$.

4.1.3 The critical surface fugacity for SAWs is $1 + \sqrt{2}$

At $n = 0$, the identity (4.12) becomes, for the dense and dilute regimes respectively,

$$1 = \cos\left(\frac{3\pi(2 \pm 1)}{8}\right) A_{T,L}^*(x_c, y) + \cos\left(\frac{\pi(2 \pm 1)}{4}\right) E_{T,L}^*(x_c, y) + \frac{y^* - y}{y(y^* - 1)} B_{T,L}^*(x_c, y)$$

$$= \mp \frac{\sqrt{2 \pm \sqrt{2}}}{2} A_{T,L}(x_c, y) \mp \frac{1}{\sqrt{2}} E_{T,L}(x_c, y) + \frac{y^* - y}{y(y^* - 1)} B_{T,L}(x_c, y), \tag{4.13}$$

where

$$x_c = \sqrt{2 \mp \sqrt{2}}, \qquad y^* = 1 \mp \sqrt{2}.$$

Note that at $n = 0$ we have $D_{T,L}(x_c, y) = 1$, so the generating functions are equal to their normalised counterparts.

As seen in [34], the identity (4.13) provides easy bounds, existence of limits, etc. if all coefficients are positive, so we now consider only the dilute regime. We denote

$$c_\alpha = \cos\left(\frac{3\pi}{8}\right) = \frac{\sqrt{2 - \sqrt{2}}}{2}, \qquad c_\epsilon = \cos\left(\frac{\pi}{4}\right) = \frac{1}{\sqrt{2}}$$

$$c_\beta(y) = \frac{y^* - y}{y(y^* - 1)} = \frac{1 + \sqrt{2} - y}{\sqrt{2} y},$$

so that the identity of interest is

$$1 = c_\alpha A_{T,L}(x_c, y) + c_\epsilon E_{T,L}(x_c, y) + c_\beta(y) B_{T,L}(x_c, y). \tag{4.14}$$

For $y < y^*$, the coefficients c_α, c_ϵ and $c_\beta(y)$ are positive. Since $A_{T,L}(x_c, y)$ and $B_{T,L}(x_c, y)$ are clearly non-decreasing in L (for non-negative y), we can take the limit $L \to \infty$ in (4.14) to obtain

$$1 = c_\alpha A_T(x_c, y) + c_\epsilon E_T(x_c, y) + c_\beta(y) B_T(x_c, y), \tag{4.15}$$

where $A_T(x, y)$ and $B_T(x, y)$ were defined in the previous subsection (they are, respectively, the generating functions of loops and bridges in a strip of width T) and $E_T(x, y)$ is defined to be

$$E_T(x, y) = \lim_{L \to \infty} E_{T,L}(x, y).$$

Lemma 4.7. *The critical surface fugacity y_c satisfies*

$$y_c \geq y^*.$$

Proof. The identity (4.15) shows that $A_T(x_c, y)$ and $B_T(x_c, y)$ are bounded, and thus convergent, for $y < y^*$. Hence $\rho_T(y) > x_c$ for $y < y^*$, and then $\rho(y) = \lim_{T \to \infty} \rho_T(y) \geq x_c$. But by definition of y_c, $\rho(y) < x_c$ for $y > y_c$, so we must have $y_c \geq y^*$. ■

Corollary 4.8. *For* $0 \leq y < y^*$,

$$\lim_{L \to \infty} E_{T,L}(x_c, y) \equiv E_T(x_c, y) = 0,$$

and hence

$$1 = c_\alpha A_T(x_c, y) + c_\beta(y) B_T(x_c, y). \tag{4.16}$$

Proof. In the proof of Lemma 4.7 we showed $\rho_T(y) > x_c$ for $y < y^*$. Since $\rho_T(y)$ is also the radius of convergence of $C_T(x, y)$, we must then have

$$\sum_L E_{T,L}(x_c, y) \leq C_T(x_c, y) < \infty,$$

and thus $\lim_{L \to \infty} E_{T,L}(x_c, y) = 0$. ■

Now consider the identity (4.16) at $y = 1$:

$$1 = c_\alpha A_T(x_c, 1) + B_T(x_c, 1).$$

Since $A_T(x_c, 1)$ is non-decreasing with T and bounded above by $1/c_\alpha$, it must have a finite limit as $T \to \infty$. Hence, so too must $B_T(x_c, 1)$. We define

$$\delta = \lim_{T \to \infty} B_T(x_c, 1) = 1 - \lim_{T \to \infty} c_\alpha A_T(x_c, 1).$$

Proposition 4.9. *If* $\delta = 0$, *then* $y^* \geq y_c$, *and hence* $y^* = y_c$.

Proof. As in Corollary 4.6, define y_T to be the radius of convergence of $C_T(x_c, y)$, so that $\lim_{T \to \infty} y_T = y_c$, and $A_T(x_c, y)$ and $B_T(x_c, y)$ are convergent on $0 \leq y < y_T$.

Walks counted by $A_{T+1}(x, y)$ and not $A_T(x, y)$ must at some point touch the β boundary of the strip of width $T + 1$; they can thus be decomposed into a pair of bridges (see Figure 4.6). This leads to the inequality

$$A_{T+1}(x_c, y) - A_T(x_c, 1) \leq x_c B_T(x_c, 1) B_{T+1}(x_c, y), \tag{4.17}$$

valid for $0 \leq y < y_{T+1}$. Using (4.16) to eliminate the two A terms, we obtain

$$B_T(x_c, 1) - c_\beta(y) B_{T+1}(x_c, y) \leq c_\alpha x_c B_T(x_c, 1) B_{T+1}(x_c, y),$$

and hence

$$0 \leq \frac{1}{B_{T+1}(x_c, y)} \leq c_\alpha x_c + \frac{c_\beta(y)}{B_T(x_c, 1)} = c_\alpha x_c + \frac{1}{B_T(x_c, 1)} \frac{y^* - y}{y(y^* - 1)}.$$
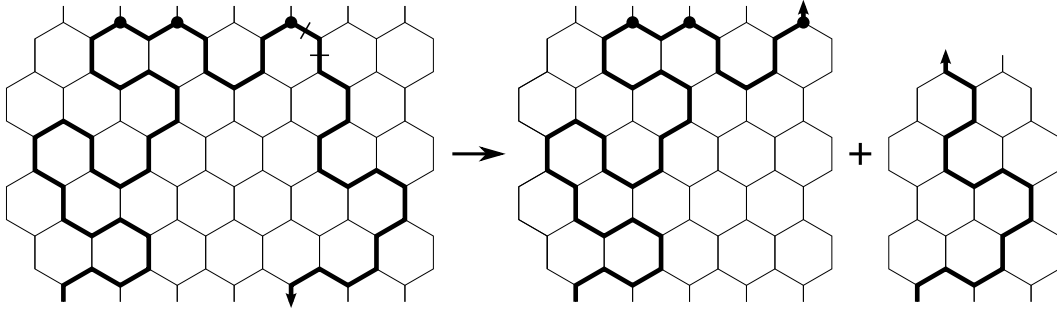
Figure 4.6: Factorisation of a loop of height $T+1$ into two bridges, of height $T+1$ and T respectively. (In this figure the strip is oriented horizontally, rather than vertically.)

In particular, for $y < y_c = \lim_{T \to \infty} y_T$,

$$0 \leq c_\alpha x_c + \frac{1}{B_T(x_c, 1)} \frac{y^* - y}{y(y^* - 1)}. \tag{4.18}$$

Now recall that, by assumption, $B_T(x_c, 1) \to \delta = 0$. If $y_c > y^*$, then for any y with $y^* < y < y_c$, the right side of (4.18) will become arbitrarily large (in modulus) and negative as $T \to \infty$. Thus, we are forced to conclude that $y_c \leq y^*$. ∎

It is clear then that the proof that $y_c = 1 + \sqrt{2}$ depends entirely on $B_T(x_c, 1) \to 0$ as $T \to \infty$. The proof of this fact is presented in Appendix A; it is due primarily to Hugo Duminil-Copin, and is of a markedly different style to the majority of the work in the rest of this thesis.

4.1.4 Rotated honeycomb lattice

If we rotate the honeycomb lattice through 90° (or alternatively, rotate the surface – either way, we end up with lattice edges *parallel* to the surface, rather than perpendicular), we can now put weights on the surface vertices and again ask where the adsorption transition occurs (see Figure 4.8). See Figure 4.7 for a SAW interacting with an impenetrable surface in this new orientation. Batchelor, Bennett-Wood and Owczarek conjecture [6] that the critical surface fugacity is

$$y_c = \sqrt{\frac{2 + \sqrt{2}}{1 + \sqrt{2} - \sqrt{2 + \sqrt{2}}}} = 2.455\ldots,$$

based on the Bethe Ansatz solution to the $O(n)$ loop model on the honeycomb lattice.

Our method for proving the critical fugacity in the original orientation (i.e. the presentation of the previous subsections) can be adapted to this situation, and the proof is only slightly more complicated than before. In the same way as the previous proof, it depends on us being able to prove that $B_T(x)$, the generating function of bridges which span a strip of width T, vanishes as $T \to \infty$ when $x = x_c = 1/\sqrt{2 + \sqrt{2}}$. We expect that the proof of this fact will follow in much
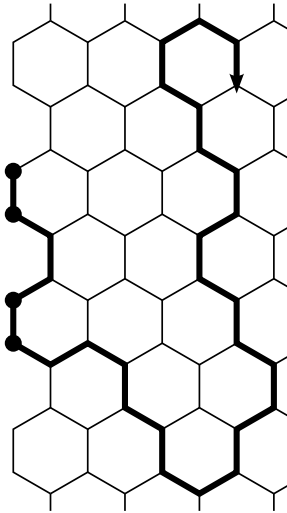
Figure 4.7: A SAW interacting with an impenetrable surface (the left hand side of the picture) in the rotated orientation. (Note that the lattice now contains edges parallel to the surface.) This orientation is the focus of Subsection 4.1.4.
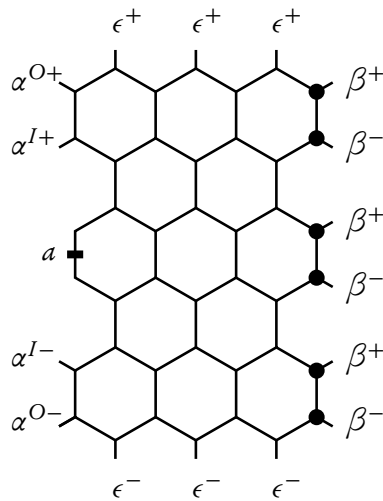
Figure 4.8: The domain we will use in the proof, with the weighted vertices on the β boundary indicated. Walks start at mid-edge a. The labels on the external mid-edges indicate the set containing the walks which end there. (The external mid-edges above and below a will not play a part in our identity, and thus are not shown in the figure.) This domain has height $2L + 1 = 5$ and width $T = 5$; in the weighted case we will require $L + T \equiv 1 \pmod 2$, so as to be sure we can pair the B^+ and B^- mid-edges appropriately.

152

the same way as that in Appendix A, but have not yet had the chance to complete it. In this subsection, then, we prove a result equivalent to Proposition 4.9.

First we consider the unweighted case, as there are already a number of factors which complicate the situation compared to the geometry used by Duminil-Copin and Smirnov. As before, we use the parafermionic observable

$$F(z) = \sum_{\gamma \subset \Omega : a \to z} e^{-i\sigma W(\gamma : a \to z)} x^{|\gamma|},$$

where a is a mid-edge on the boundary of the domain and z is any mid-edge in the domain. Then as always, Lemma 2.1 holds for any vertex $v \in V(\Omega)$. For now we will consider only the $n = 0$ case in the dilute regime, so that $x = x_c = 1/\sqrt{2 + \sqrt{2}}$ and $\sigma = 5/8$. (Later in this section we consider the more general model.)

We will work in a domain as illustrated in Figure 4.8, with width T and height $2L + 1$. (We are initially ignoring the weights on the β boundary.) SAWs start at the mid-edge a, in the middle of the left-hand side of the domain, and end on external mid-edges, which fall into one of 8 classes (shown in the figure). Because, for reasons of symmetry, we cannot take a to be an external mid-edge, we end up having to count self-avoiding polygons (i.e. walks which start and end at a) as well. Our identity in this geometry is the following. Define

$$A^O_{T,L}(x) = \sum_{\gamma : a \to \alpha^{O+} \bigcup \alpha^{O-}} x^{|\gamma|} \qquad\qquad A^I_{T,L}(x) = \sum_{\gamma : a \to \alpha^{I+} \bigcup \alpha^{I-}} x^{|\gamma|}$$

$$E_{T,L}(x) = \sum_{\gamma : a \to \epsilon^{+} \bigcup \epsilon^{-}} x^{|\gamma|} \qquad\qquad B_{T,L}(x) = \sum_{\gamma : a \to \beta^{+} \bigcup \beta^{-}} x^{|\gamma|}$$

$$P_{T,L}(x) = \sum_{\rho \ni a} x^{|\rho|}$$

where the last sum is over all *undirected* (non-empty) self-avoiding polygons in the domain which contain the mid-edge a.

Proposition 4.10. *The generating functions $A^O_{T,L}, A^I_{T,L}, E_{T,L}, B_{T,L}$ and $P_{T,L}$, evaluated at $x = x_c$, satisfy the identity*

$$\frac{1}{2}\sqrt{2 - \sqrt{2 - \sqrt{2}}}\, A^O_{T,L}(x_c) + \frac{1}{2}\sqrt{2 - \sqrt{2 + \sqrt{2}}}\, A^I_{T,L}(x_c) + \frac{1}{2}\sqrt{2 + \sqrt{2 - \sqrt{2}}}\, E_{T,L}(x_c)$$

$$+ \frac{1}{2}\sqrt{2 + \sqrt{2 + \sqrt{2}}}\, B_{T,L}(x_c) + \sqrt{4 + 2\sqrt{2} - \sqrt{2(10 + 7\sqrt{2})}}\, P_{T,L}(x_c)$$

$$= \sqrt{2 - \sqrt{2} + \sqrt{\frac{1}{2}\left(2 - \sqrt{2}\right)}} \quad (4.19)$$

Proof. We would like winding angle of the reflection (through the horizontal axis) of a walk to be the negative of that of the original walk. We also require that all the walks ending on a given external mid-edge have the same winding angle. This results in a strange situation: the empty
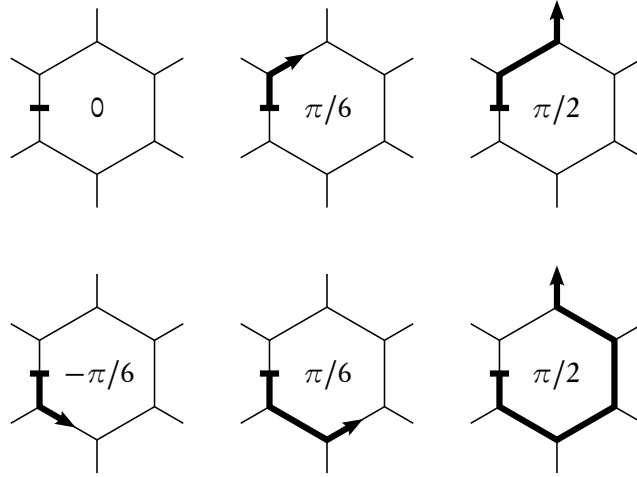
Figure 4.9: Examples of some walks on the rotated honeycomb domain, with their winding angles indicated. Non-empty walks "begin" with a winding angle $\pm\pi/2$, depending on whether they start in the up or down direction.

walk must have winding angle 0 (it is its own reflection), but a non-empty walk must "begin" with a winding angle of $\pm\pi/2$ (depending on whether it starts in the up or down direction from a).[7] We provide some small examples for illustrative purposes in Figure 4.9.

Now the identity (2.3) will hold for any vertex v which is not adjacent to the origin a. (The proof is completely identical.) It will, however, fail at the two vertices above and below a, for the reasons described above: the empty walk has winding angle 0, but non-empty walks begin with non-zero winding. Thus, we sum (2.3) over all vertices in the domain *except* the ones immediately above and below a. Clearly this sum adds to 0.

On the other hand, we can compute the sum by noting that any mid-edge adjacent to two of the vertices summed over will contribute 0 to the sum. The remaining mid-edges are then the external ones marked as in Figure 4.8, as well as the two adjacent to the vertices above and below a, which we will denote by ζ and $\bar{\zeta}$ respectively. If we define $j = \exp(5\pi i/6)$ and $\theta = \exp(-i\sigma) = \exp(-5i/8)$, then the coefficients of the walks ending on external mid-edges are

$$\alpha^{O+} : j\theta^{5\pi/6} = \exp(5\pi i/16) \qquad \alpha^{O-} : \bar{j}\theta^{-5\pi/6} = \exp(-5\pi i/16)$$
$$\alpha^{I+} : \bar{j}\theta^{7\pi/6} = \exp(7\pi i/16) \qquad \alpha^{I-} : j\theta^{-7\pi/6} = \exp(-7\pi i/16)$$
$$\epsilon^{+} : i\theta^{\pi/2} = \exp(3\pi i/16) \qquad \epsilon^{-} : -i\theta^{-\pi/2} = \exp(-3\pi i/16)$$
$$\beta^{+} : -\bar{j}\theta^{\pi/6} = \exp(\pi i/16) \qquad \beta^{-} : -j\theta^{-\pi/6} = \exp(-\pi i/16)$$

There are two types of walks ending at ζ and $\bar{\zeta}$: those with winding $\pm\pi/6$ and those with winding $\mp7\pi/6$. The first type comprises only one walk for each of ζ and $\bar{\zeta}$: a single step

---

[7]An alternative way to view this situation is to consider the empty walk to be comprised of a "right step of length 0". Then non-empty walks actually begin with a turn of angle $\pm\pi/2$.

through the vertex above/below a. The contribution of these walks is thus

$$\bar{j}\theta^{\pi/6}x_c + j\theta^{-\pi/6}x_c = -\sqrt{2 - \sqrt{2} + \sqrt{\frac{1}{2}\left(2 - \sqrt{2}\right)}}.$$

The second type of walks loop around almost all the way back to a: it seems sensible to then just add a step and be left with self-avoiding polygons containing the mid-edge a. If $P_{T,L}(x)$ is the generating function for *undirected* polygons, then the contribution of these walks is

$$\left(\frac{\bar{j}\theta^{7\pi/6}}{x_c} + \frac{j\theta^{-7\pi/6}}{x_c}\right)P_{T,L}(x_c) = \sqrt{4 + 2\sqrt{2} - \sqrt{2\left(10 + 7\sqrt{2}\right)}}P_{T,L}(x_c).$$

For the walks ending on external mid-edges, just note that we can pair walks (via reflection through the horizontal axis) ending in τ^+ and τ^-, where τ is any of $\alpha^O, \alpha^I, \epsilon, \beta$. So the contributions of these walks are

$$\alpha^{O+} \cup \alpha^{O-} : \left(\frac{j\theta^{5\pi/6} + \bar{j}\theta^{-5\pi/6}}{2}\right)A_{T,L}^O(x_c) = \frac{1}{2}\sqrt{2 - \sqrt{2 - \sqrt{2}}}A_{T,L}^O(x_c)$$

$$\alpha^{I+} \cup \alpha^{I-} : \left(\frac{\bar{j}\theta^{7\pi/6} + j\theta^{-7\pi/6}}{2}\right)A_{T,L}^I(x_c) = \frac{1}{2}\sqrt{2 - \sqrt{2 + \sqrt{2}}}A_{T,L}^I(x_c)$$

$$\epsilon^+ \cup \epsilon^- : \left(\frac{i\theta^{\pi/2} - i\theta^{-pi/2}}{2}\right)E_{T,L}(x_c) = \frac{1}{2}\sqrt{2 + \sqrt{2 - \sqrt{2}}}E_{T,L}(x_c)$$

$$\beta^+ \cup \beta^- : \left(\frac{-\bar{j}\theta^{\pi/6} - j\theta^{-\pi/6}}{2}\right)B_{T,L}(x_c) = \frac{1}{2}\sqrt{2 + \sqrt{2 + \sqrt{2}}}B_{T,L}(x_c)$$

Adding all the above contributions and equating with 0 gives the proposition. ∎

We now introduce the the surface weights on the β boundary (as shown in Figure 4.8). Define the same generating functions as before, but now with a variable y keeping track of the number of surface contacts:

$$A_{T,L}^O(x,y) = \sum_{\gamma: a \to \alpha^{O+}\bigcup \alpha^{O-}} x^{|\gamma|}y^{\nu(\gamma)}, \qquad A_{T,L}^I(x,y) = \sum_{\gamma: a \to \alpha^{I+}\bigcup \alpha^{I-}} x^{|\gamma|}y^{\nu(\gamma)},$$

$$E_{T,L}(x,y) = \sum_{\gamma: a \to \epsilon^+\bigcup \epsilon^-} x^{|\gamma|}y^{\nu(\gamma)}, \qquad B_{T,L}(x,y) = \sum_{\gamma: a \to \beta^+\bigcup \beta^-} x^{|\gamma|}y^{\nu(\gamma)},$$

$$P_{T,L}(x,y) = \sum_{\rho \ni a} x^{|\rho|}y^{\nu(\rho)}.$$

Proposition 4.11. *The functions $A_{T,L}^O(x,y)$, $A_{T,L}^I(x,y)$, $E_{T,L}(x,y)$, $B_{T,L}(x,y)$ and $P_{T,L}(x,y)$, evaluated at $x = x_c$, satisfy the functional equation*

$$\frac{1}{2}\sqrt{2 - \sqrt{2 - \sqrt{2}}}A_{T,L}^O(x_c,y) + \frac{1}{2}\sqrt{2 - \sqrt{2 + \sqrt{2}}}A_{T,L}^I(x_c,y) + \frac{1}{2}\sqrt{2 + \sqrt{2 - \sqrt{2}}}E_{T,L}(x_c,y)$$

$$+ c_\beta(y)B_{T,L}(x_c,y) + \sqrt{4 + 2\sqrt{2} - \sqrt{2(10 + 7\sqrt{2})}}P_{T,L}(x_c,y) = \sqrt{2 - \sqrt{2} + \sqrt{\frac{1}{2}\left(2 - \sqrt{2}\right)}},$$

$$(4.20)$$

where

$$c_\beta(y) = \frac{c_\beta}{x_c y(1 + x_c y)} - \frac{x_c y \sqrt{2 - \sqrt{2 - \sqrt{2}}}}{2(1 + x_c y)}$$

and

$$c_\beta = c_\beta(1) = \frac{1}{2}\sqrt{2 + \sqrt{2 + \sqrt{2}}}.$$

In particular, $c_\beta(y)$ is a positive, non-increasing function of y on $(0, y^\dagger)$ where

$$y^\dagger = \sqrt{\frac{2 + \sqrt{2}}{1 + \sqrt{2} - \sqrt{2 + \sqrt{2}}}}$$

and $c_\beta(y^\dagger) = 0$.

Proof. We again sum the LHS of (2.3) over all vertices except the two adjacent to a. When $y \neq 1$ the contribution of β vertices will not be 0, but can instead be written as a multiple of the $B_{T,L}$ generating function.

A β^+ walk must approach its final vertex either from the north-east or from the south. Let γ_1 be a walk approaching a β^+ vertex from the north-east, and say γ_1^l and γ_1^r are the walks obtained by appending a left or right turn to γ_1 respectively. Then the sum of the contributions of γ_1, γ_1^l and γ_1^r is

$$x_c^{|\gamma_1|} y^{\nu(\gamma_1)} (j\theta^{-\pi/6} + x_c y(-\bar{j})\theta^{\pi/6} + x_c y(-\mathrm{i})\theta^{-\pi/2}).$$

Similarly, let γ_2 be a walk approaching a β^+ vertex from the south, and γ_2^l and γ_2^r its two extensions. Then the contribution of these three walks is

$$x_c^{|\gamma_2|} y^{\nu(\gamma_2)} (-\mathrm{i}\theta^{\pi/2} + x_c y j\theta^{5\pi/6} + x_c y(-\bar{j})\theta^{\pi/6})$$

Now any walk finishing adjacent to a β^+ vertex must be described by exactly one of $\gamma_1, \gamma_1^l, \gamma_1^r, \gamma_2, \gamma_2^l, \gamma_2^r$. So if $G_{T,L}^1(x,y)$ and $G_{T,L}^2(x,y)$ are the generating functions for γ_1 and γ_2 walks respectively, the contribution of all β^+ vertices is

$$(j\theta^{-\pi/6} + x_c y(-\bar{j})\theta^{\pi/6} + x_c y(-\mathrm{i})\theta^{-\pi/2})G_{T,L}^1(x_c, y)$$
$$+ (-\mathrm{i}\theta^{\pi/2} + x_c y j\theta^{5\pi/6} + x_c y(-\bar{j})\theta^{\pi/6})G_{T,L}^2(x_c, y). \quad (4.21)$$

But now it's easy to see that any reflected γ_1 walk can be extended to a unique γ_2 walk, and any γ_2 walk is an extension of a unique reflected γ_1 walk. So in fact

$$G_{T,L}^2(x,y) = xy G_{T,L}^1(x,y).$$

So (4.21) becomes

$$(j\theta^{-\pi/6} + x_c y(-\bar{j})\theta^{\pi/6} + x_c y(-\mathrm{i})\theta^{-\pi/2} + x_c y(-\mathrm{i})\theta^{\pi/2} + x_c^2 y^2 j\theta^{5\pi/6} + x_c^2 y^2(-\bar{j})\theta^{\pi/6})G_{T,L}^1(x_c, y)$$
$$(4.22)$$

Since any β^+ vertex can be reflected to give a β^- vertex, the contribution of β^- vertices is

$$(\bar{j}\theta^{\pi/6} + x_c y(-j)\theta^{-\pi/6} + x_c y i \theta^{\pi/2} + x_c y i \theta^{-\pi/2} + x_c^2 y^2 \bar{j}\theta^{-5\pi/6} + x_c^2 y^2(-j)\theta^{-\pi/6})G_{T,L}^1(x_c,y)$$

(4.23)

Now any walk counted by $B_{T,L}$ can be obtained by extending a unique $G_{T,L}^1$ walk (or a reflected one) by either a single step or by two steps. Similarly, any $G_{T,L}^1$ walk (or a reflected one) can be extended by one or two steps to give a $B_{T,L}$ walk. So we have

$$B_{T,L}(x,y) = 2(xy + x^2 y^2)G_{T,L}^1(x,y)$$

(4.24)

So by adding (4.22) and (4.23) and substituting (4.24), we find that the contribution of all β^+ and β^- vertices is

$$\frac{n_0 + n_1 y + n_2 y^2}{d_1 y + d_2 y^2}B_{T,L}(x_c,y)$$

with n_0, n_1, n_2, d_1, d_2 as defined in the proposition.

Now nothing has changed from the unweighted case for the other external mid-edges (and the two special mid-edges ζ and $\bar{\zeta}$), so we obtain

$$\frac{1}{2}\sqrt{2 - \sqrt{2 - \sqrt{2}}}A_{T,L}^O(x_c,y) + \frac{1}{2}\sqrt{2 - \sqrt{2 + \sqrt{2}}}A_{T,L}^I(x_c,y) + \frac{1}{2}\sqrt{2 + \sqrt{2 - \sqrt{2}}}E_{T,L}(x_c,y)$$

$$+ \frac{1}{2}\sqrt{2 + \sqrt{2 + \sqrt{2}}}B_{T,L}(x_c,y) + \sqrt{4 + 2\sqrt{2} - \sqrt{2(10 + 7\sqrt{2})}}P_{T,L}(x_c,y)$$

$$= \sqrt{2 - \sqrt{2} + \sqrt{\frac{1}{2}\left(2 - \sqrt{2}\right)}} + \frac{n_0 + n_1 y + n_2 y^2}{d_1 y + d_2 y^2}B_{T,L}(x_c,y)$$

from the which the proposition immediately follows. ∎

Here we will omit discussion of results analogous to those of Subsection 4.1.2, and will merely point out tha the same arguments apply to the rotated honeycomb lattice.[8] We will use the same notation as before, and define the free energy $\varkappa_T(1,y)$ for walks[9] in a strip of width T with fugacity y associated with vertices on the right-hand side of the strip. Then as before, the free energy $\varkappa(y)$ of walks in a half-plane satisfies

$$\lim_{T\to\infty}\varkappa_T(1,y) = \varkappa(y).$$

The radius of convergence of all the generating functions of walks in the strip is given by $\rho_T(y) = \exp(-\varkappa_T(1,y))$, and again we define $\rho(y) = \lim_{T\to\infty}\rho_T(y)$. Finally, y_T is defined to be the

---

[8] Actually, the process of unfolding is a little more complicated in this situation, because the lattice is not invariant under reflection through a horizontal line passing through a vertex. This means that an edge must be inserted when concatenating two pieces of a walk after one has been reflected. It is straightforward to show that when unfolding a walk of length n, one needs to use only $O(\sqrt{n})$ reflections, and so in the end this does not invalidate any of the arguments. Since this is currently work in progress (and in this thesis we will not actually complete the proof of the critical surface fugacity), we will include further details in a later publication.

[9] Regardless of whether they end on the left, right or anywhere in the strip.

unique value of y satisfying $\rho_T(y_T) = x_c = 1/\sqrt{2+\sqrt{2}}$, and then it can be shown that y_c, the critical fugacity of walks in a half-plane, satisfies $y_c = \lim_{T \to \infty} y_T$.

Proposition 4.12. *If it can be shown that*

$$\lim_{T \to \infty} B_T(x_c, 1) \equiv B(x_c, 1) = 0$$

then $y_c = y^\dagger$.

This proof is essentially the same as that of Proposition 4.9, except for the added complication of two different A generating functions and the P generating function.

For $y < y^\dagger$, every term in (4.20) is non-negative, and since every generating function except $E_{T,L}$ is clearly non-decreasing with L, it follows that $E_{T,L}$ must be non-increasing. Hence everything has a limit in L, so we can take that limit and get

$$\frac{1}{2}\sqrt{2-\sqrt{2-\sqrt{2}}}A_T^O(x_c,y) + \frac{1}{2}\sqrt{2-\sqrt{2+\sqrt{2}}}A_T^I(x_c,y) + \frac{1}{2}\sqrt{2+\sqrt{2-\sqrt{2}}}E_T(x_c,y)$$

$$+ c_\beta(y)B_T(x_c,y) + \sqrt{4+2\sqrt{2}-\sqrt{2(10+7\sqrt{2})}}P_T(x_c,y) = \sqrt{2-\sqrt{2}+\sqrt{\frac{1}{2}\left(2-\sqrt{2}\right)}} \quad (4.25)$$

In an obvious shorthand we rewrite (4.25) as

$$c_\alpha^O A_T^O(x_c,y) + c_\alpha^I A_T^I(x_c,y) + c_\epsilon E_T(x_c,y) + c_\pi P_T(x_c,y) + c_\beta(y)B_T(x_c,y) = c_{\text{RHS}}. \quad (4.26)$$

Lemma 4.13. *The critical surface fugacity y_c satisfies*

$$y_c \geq y^\dagger.$$

Proof. The identity (4.26) shows that, for $y < y^\dagger$, the generating function $A_T^O(x_c,y)$ (among others) is bounded, and thus convergent. Hence $\rho_T(y) > x_c$ for $y < y^\dagger$, and then $\lim_{T \to \infty} \rho_T(y) = \rho(y) \geq x_c$. But by definition of y_c, $\rho(y) < x_c$ for $y > y_c$, and thus we must have $y_c \geq y^\dagger$. ∎

Corollary 4.14. *For $0 \leq y < y^\dagger$,*

$$\lim_{L \to \infty} E_{T,L}(x_c,y) \equiv E_T(x_c,y) = 0,$$

so that

$$c_\alpha^O A_T^O(x_c,y) + c_\alpha^I A_T^I(x_c,y) + c_\pi P_T(x_c,y) + c_\beta(y)B_T(x_c,y) = c_{\text{RHS}}. \quad (4.27)$$

Proof. The proof is the same as for Corollary 4.8. The radius of convergence of $C_T(x,y)$, the generating function of all walks in the strip starting at the point a, is $\rho_T(y)$, and since $\rho_T(y) > x_c$ for $0 \leq y < y^\dagger$, we have

$$\sum_L E_{T,L}(x_c,y) \leq C_T(x_c,y) < \infty,$$

which gives the result of the lemma. ∎
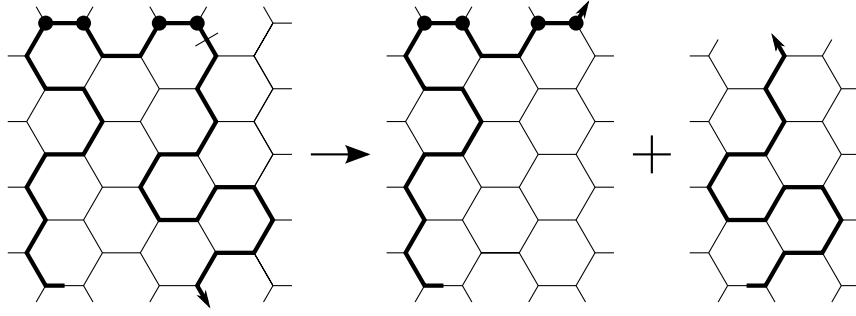
Figure 4.10: Factorisation of a walk counted by A^O_{T+1} into two bridges. In this figure the strip is oriented horizontally, rather than vertically.

Now consider the identity (4.27) at $y = 1$:

$$c^O_\alpha A^O_T(x_c, 1) + c^I_\alpha A^I_T(x_c, 1) + c_\pi P_T(x_c, 1) + c_\beta B_T(x_c, 1) = c_{\text{RHS}}. \tag{4.28}$$

Since $A^O_T(x_c, 1)$, $A^I_T(x_c, 1)$ and $P_T(x_c, 1)$ are all non-decreasing with T and bounded above by (4.28), they must have limits in T. Hence, so too must $B_T(x_c, 1)$. We will denote

$$\delta = \lim_{T \to \infty} c_\beta B_T(x_c, 1).$$

Proof of Proposition 4.12. Recall that y_T is the radius of convergence of $C_T(x_c, y)$ as well as all functions in (4.27).

In the same way that we decomposed walks counted by $A_{T+1}(x_c, y) - A_T(x_c, 1)$ in the proof of Proposition 4.9, we can decompose walks counted by

$$A^O_{T+1}(x_c, y) - A^O_T(x_c, 1),$$
$$A^I_{T+1}(x_c, y) - A^I_T(x_c, 1), \text{ and}$$
$$P_{T+1}(x_c, y) - P_T(x_c, 1)$$

into pairs of bridges. For example, we have

$$A^O_{T+1}(x_c, y) - A^O_T(x_c, 1) \leq B_{T+1}(x_c, y) B_T(x_c, 1).$$

See Figure 4.10 for an illustration.

Combining this decomposition for A^O, A^I and P, we find for $0 \leq y < y_{T+1}$,

$$c^O_\alpha [A^O_{T+1}(x_c, y) - A^O_T(x_c, 1)] + c^I_\alpha [A^I_{T+1}(x_c, y) - A^I_T(x_c, 1)] + c_\pi [P_{T+1}(x_c, y) - P_T(x_c, 1)]$$
$$\leq (c^O_\alpha + c^I_\alpha + c_\pi) B_{T+1}(x_c, y) B_T(x_c, 1). \tag{4.29}$$

Using (4.27) to eliminate the A^O, A^I and P terms, we obtain

$$c_\beta B_T(x_c, 1) - c_\beta(y) B_{T+1}(x_c, y) \leq (c^O_\alpha + c^I_\alpha + c_\pi) B_{T+1}(x_c, y) B_T(x_c, 1),$$

159

and hence

$$0 \le \frac{1}{B_{T+1}(x_c, y)} \le \frac{c_\alpha^O + c_\alpha^I + c_\pi}{c_\beta} + \frac{c_\beta(y)}{c_\beta B_T(x_c, 1)}. \tag{4.30}$$

In particular, for $0 \le y < y_c = \lim_{T \to \infty} y_T$ and for any T,

$$0 \le \frac{c_\alpha^O + c_\alpha^I + c_\pi}{c_\beta} + \frac{c_\beta(y)}{c_\beta B_T(x_c, 1)}.$$

Taking $T \to \infty$, we find

$$0 \le \frac{c_\alpha^O + c_\alpha^I + c_\pi}{c_\beta} + \frac{c_\beta(y)}{\delta}. \tag{4.31}$$

Now by assumption, $\delta = 0$. Suppose (for a contradiction) that $y_c > y^\dagger$. Then for any $y^\dagger < y < y_c$, the RHS of (4.31) will be arbitrarily large in modulus and negative, contradicting the inequality. Thus, we are forced to conclude $y_c \le y^\dagger$, and hence $y_c = y^\dagger$. ∎

General n

We conclude this section by generalising the identities for the rotated honeycomb lattice to the $O(n)$ loop model with $n \in [-2, 2]$. As with the original orientation, we let $n = 2\cos\theta$ with $\theta \in [0, \pi]$. Then the local cancellation identity (2.3) holds for

$$\sigma = \frac{\pi - 3\theta}{4\pi}, \qquad x_c^{-1} = 2\cos\left(\frac{\pi + \theta}{4}\right) = \sqrt{2 - \sqrt{2 - n}} \qquad \text{or}$$

$$\sigma = \frac{\pi + 3\theta}{4\pi}, \qquad x_c^{-1} = 2\cos\left(\frac{\pi - \theta}{4}\right) = \sqrt{2 + \sqrt{2 - n}}$$

Following the same method as for SAWs, we obtain the same identity as before, with a few differences:

- Obviously all generating functions now have an additional variable n.

- The coefficients of each generating function now depend on n (via θ, via σ, via θ) but are calculated in exactly the same way as before.

- The closed loop counted by P which passes through a will not naturally pick up a weight of n (since it's really a SAW which finishes just above or below a), but for consistency we can artifically insert an extra factor of n and then divide the coefficient c_π by n.

- The objects that contributed to the RHS of (4.19) and (4.20) (single steps up or down) now become generating functions of these objects plus closed loops. If we absorb a factor of x_c into the coefficient then this is the generating function of the empty walk plus closed loops which *do not* contain the mid-edge a.

(It would be nice if we could combine the generating functions of closed loops which do or do not contain a, but unfortunately the coefficients just don't match.)

Following all this through, we end up with

Proposition 4.15. *If $n = 2\cos\theta$ with $\theta \in [0, \pi]$ and $x_c^{-1} = 2\cos((\pi \mp \theta)/4)$ then*

$$\cos\left(\frac{5(\pi \mp \theta)}{8}\right) A_{T,L}^O(x_c, y) + \cos\left(\frac{9\pi \pm 7\theta}{8}\right) A_{T,L}^I(x_c, y)$$

$$+ \cos\left(\frac{3(\pi \mp \theta)}{8}\right) E_{T,L}(x_c, y) + \frac{2}{x_c \cos\theta} \cos\left(\frac{7(\pi \mp \theta)}{8}\right) P_{T,L}(x_c, y)$$

$$+ \left[\cos\left(\frac{\pi \mp \theta}{8}\right) - \frac{\cos\left(\frac{9\pi \mp \theta}{8}\right)(1 - x_c y - x_c^2 y^2) + \cos\left(\frac{5(\pi \mp \theta)}{8}\right) x_c^2 y^2}{x_c y (1 + x_c y)}\right] B_{T,L}(x_c, y)$$

$$= 2x_c \cos\left(\frac{\pi \mp \theta}{8}\right) D_{T,L}(x_c, y) \quad (4.32)$$

where in each generating function n is conjugate to the number of closed loops, and $D_{T,L}$ is the generating function of the empty walk and the configurations containing only *closed loops which do not pass through the mid-edge a.*

The first equation in (4.32) corresponds to the smaller value of x_c and thus describes the so-called dilute regime (where fewer closed loops are favoured), while the second describes the dense regime. When $\theta \to \pi/2$ (so $n \to 0$), the dilute equation exactly reduces to (4.20).

We note here that, while we know of no existing conjectures for the critical surface fugacity for the $O(n)$ loop model in this geometry, we might expect that, as with the original orientation, this critical value $y_c(n)$ is the function satisfying $c_\beta(y_c(n)) = 0$. This leads us to formulate the following conjecture.

Conjecture 4.16. *For the $O(n)$ loop model with $n \in [-2, 2]$ on the honeycomb lattice with an impenetrable surface, oriented so that the lattice contains edges parallel to the surface, associate a fugacity $x = x_c = 1/\sqrt{2 + \sqrt{2 - n}}$ with occupied vertices, a fugacity n with closed loops and a fugacity y with occupied vertices on the surface. Then the model undergoes a surface transition at*

$$y = y_c(n) = \sqrt{\frac{2 + \sqrt{2 - n}}{1 + \sqrt{2 - n} - \sqrt{2 + \sqrt{2 - n}}}}.$$

4.2 Numerical estimates

In Section 2.2, we showed that the identity of Duminil-Copin and Smirnov [34] on the honeycomb lattice,

$$c_\alpha A_T(x_c) + B_T(x_c) = 1,$$

could be used to obtain estimates for the growth constants (and other quantities) of the square and triangular lattices. It seems reasonable then to suppose that identities like (4.16) and (4.27) which incorporate a surface fugacity y, and which display special behaviour at the critical value of that surface fugacity, might be useful for computing estimates on other lattices. In this section

we show that this is indeed the case, and we compute estimates for the critical fugacities of a number of adsorption models on the honeycomb, square and triangular lattices.

Recall the identity (4.15),

$$1 = c_\alpha A_T(x_c, y) + c_\epsilon E_T(x_c, y) + c_\beta(y) B_T(x_c, y),$$

where

$$c_\alpha = \cos\left(\frac{3\pi}{8}\right) = \frac{\sqrt{2-\sqrt{2}}}{2}, \qquad c_\epsilon = \cos\left(\frac{\pi}{4}\right) = \frac{1}{\sqrt{2}}$$

$$c_\beta(y) = \frac{y_c - y}{y(y_c - 1)} = \frac{1 + \sqrt{2} - y}{\sqrt{2}y}.$$

This identity is valid for all y, and in particular for $0 \leq y \leq y_c = 1 + \sqrt{2}$. It was proved in Section 4.1 that $E_T(x_c, y) = 0$ for $0 \leq y < y_c$. In fact we are able to compute the exact generating functions A_T, B_T and E_T for $T = 0, 1, 2$, and for those values of T we observe $E_T(x_c, y_c) = 0$. We thus make the following conjecture.

Conjecture 4.17. *On the honeycomb lattice,*

$$\lim_{L \to \infty} E_{T,L}(x_c, y_c) \equiv E_T(x_c, y_c) = 0 \qquad (4.33)$$

for all $T \geq 0$.

For the rest of this section we will take Conjecture 4.17 as given. (Since this section is devoted to *numerical estimates*, such an assumption is quite reasonable.)

In light of Conjecture 4.17, at $y = y_c = 1 + \sqrt{2}$, the identity (4.15) reduces to

$$c_\alpha A_T(x_c, y_c) = 1. \qquad (4.34)$$

For other lattices we do not have an equivalent identity. However if one plots $A_T(x_c, y)$ versus y in these cases, one might be forgiven for thinking that such an identity exists. In Figure 4.11 we show a plot of $A_T(x_c, y)$ versus y on the square lattice for a range of strip widths T (with the weight y corresponding to vertices in the surface, and with x evaluated at the current best estimate for x_c, namely 0.37905227776). To graphical accuracy it appears that there is a unique point of intersection for plots corresponding to higher values of T. Even finer resolution (see inset) suggests that this is the case. The actual small deviation can be seen from the data given in Table 4.3.

We denote by $y^*(T)$ the point of intersection of $A_T(x_c, y)$ and $A_{T+1}(x_c, y)$. We observe that the sequence $\{y^*(T)\}$ is a monotone function of T. We argue, as in Section 2.2, that in the scaling limit all two dimensional SAW models are given by the same conformal field theory. (Supported also by [22].) Since it is known for one of these models (i.e. honeycomb lattice SAWs) that the critical point can be found by locating the point y^* which satisfies (4.34), it follows that in the
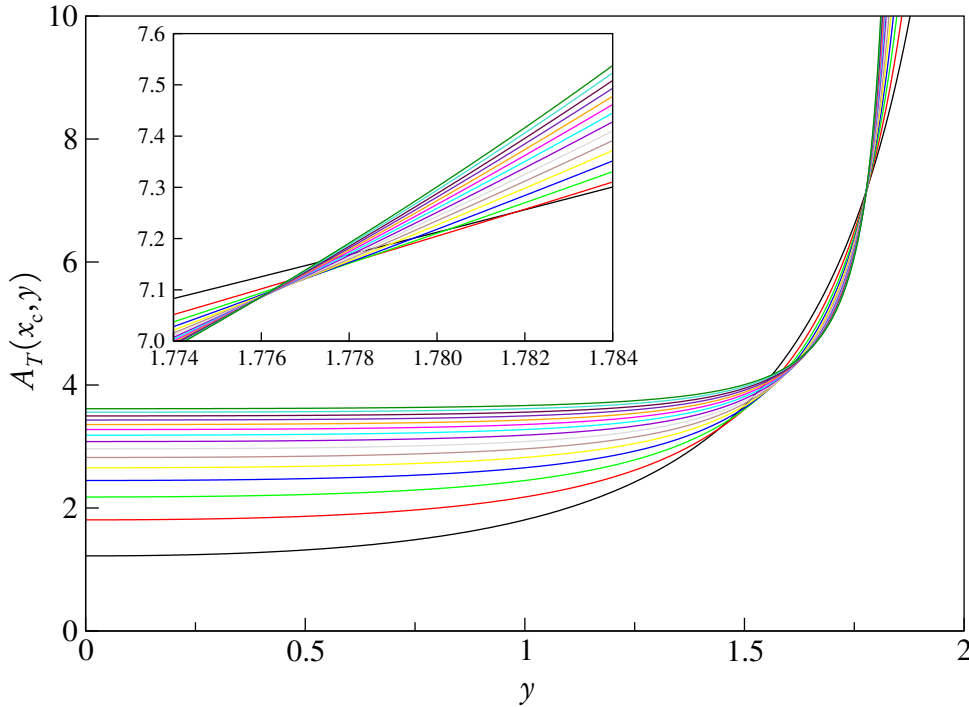
Figure 4.11: Square lattice with surface vertex interactions. $A_T(x_c, y)$ versus y for $T = 1 \ldots 15$. Inset shows the intersection region in finer scale.

scaling limit the same should be true for all two-dimensional SAWs. This is entirely consistent with our observations, and suggests that $\lim_{T \to \infty} y^*(T) = y_c$.

This then suggests a potentially powerful new numerical approach to estimating y_c. One calculates the generating functions $A_T(x_c, y)$, for all strip widths $T = 0, 1, 2, \ldots T_{max}$, uses these to calculate $y^*(T)$ for $T = 0, 1, 2, \ldots T_{max-1}$ as defined above, and then extrapolates this monotone sequence by a variety of standard sequence extrapolation methods. This is of course essentially the same idea we employed in Section 2.2.

In Subsection B.1.2 we describe the derivation of the generating functions $A_T(x_c, y)$ by the finite-lattice method for a range of strip widths T that are needed in this study. For the value of the critical step fugacity x_c, we use the exact result $x_c = 1/\sqrt{2 + \sqrt{2}}$ for the honeycomb lattice, and the best available series estimates in the case of the square and triangular lattices. These are $x_c(\text{sq}) = 0.37905227776$ [77, 72], with uncertainty in the last digit, and $x_c(\text{tr}) = 0.2409175745$ [75], with similar uncertainty. We performed a sensitivity analysis of our critical surface fugacity estimates in order to determine how sensitive they are to uncertainties in our estimates of x_c. The estimates of x_c are sufficiently precise that a change in our estimate of x_c by a factor of 10 times the estimated uncertainty will not change our estimates of the surface fugacity y_c in even the least significant digit.

In Subsections 4.2.1–4.2.3 we estimate the critical fugacities by extrapolating $y^*(T)$ using a range of standard extrapolation algorithms. These are Levin's u-transform, Brezinskii's θ al-
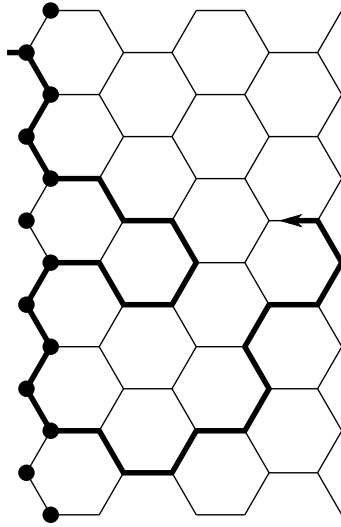
Figure 4.12: The double-vertex model on the honeycomb lattice, with the weighted vertices on the surface indicated.

gorithm, Neville tables, Wynn's ϵ algorithm and the Barber-Hamer algorithm. Descriptions of these algorithms, and codes for their implementation, can be found in [52]. However, we find the most precise estimates are given by the Bulirsch-Stoer algorithm [21]. This algorithm requires a parameter w, which can be thought of as a correction-to-scaling exponent. For the purpose of the current exercise, we have set this parameter to 1, corresponding to a T^{-2} correction term as observed. Our implementation of the algorithm is precisely as described by Monroe [88], and we retained 50 digit precision throughout.

We used this method to estimate the critical fugacity for a number of cases of interest for two-dimensional SAWs. We note that, as described in the previous section, there are two different orientations of the honeycomb lattice; we restrict ourselves here to the orientation considered in Subsections 4.1.1–4.1.3 (see, for example, Figure 4.3). For the vertex weighted model, we have already proved (recall Section 4.1) that $y_c = 1 + \sqrt{2}$, as conjectured by Batchelor and Yung [7]. It is a straightforward consequence of this result – the argument is given in Subsection 4.2.1 below – that for the edge weighted model, the critical fugacity is $\sqrt{1 + \sqrt{2}}$. We introduce here another model on the honeycomb lattice, which we call the *double-vertex* weighted model – see Figure 4.12. For that model, we estimate the critical fugacity to be $y_c = 1.46767$ where the error in this estimate (and all such estimates given below), is expected to be confined to a few parts in the last quoted digit. We know of no other estimate of this quantity in the literature.

In Subsection 4.2.2 we discuss the critical fugacity for vertex and edge weighted adsorption on the square lattice. The only previous estimates for the vertex weighted case can be found in [117], where Monte Carlo methods were used to obtain the estimate $y_c(\text{vertex}) = 1.76 \pm 0.02$. Our estimate, $y_c(\text{vertex}) = 1.77564$ is three orders of magnitude more precise than this. For the edge weighted case, a transfer matrix estimate is given in [50], and is $y_c(\text{edge}) = 2.041 \pm 0.002$.

In [49] a Monte Carlo estimate of comparable precision is given, $y_c(\text{edge}) = 2.038 \pm 0.002$. Our estimate is $y_c(\text{edge}) = 2.040135$, again some three orders of magnitude more precise.

For the triangular lattice, discussed in Subsection 4.2.3 we are unaware of any previous investigations of the critical fugacity. We find, in Subsection 4.2.3, that $y_c(\text{vertex}) = 2.144181$ and $y_c(\text{edge}) = 2.950026$. We repeat that errors in our quoted estimates are expected to be confined to a few parts in the last quoted digit.

We note here that, as with Section 2.2, the estimates computed in this section are *biased*, as they rely on unproven assumptions regarding the limiting behaviour of the generating functions $A_T(x_c, y_c), B_T(x_c, y_c)$ and $E_T(x_c, y_c)$. While, as discussed above, we have reasons to believe that these assumptions are valid, it should be remembered that there are no guarantees that these estimate are accurate.

4.2.1 Honeycomb lattice

In Section 4.1 we proved that the critical fugacity for the case of interacting vertices on the honeycomb lattice is $y_c = 1 + \sqrt{2}$. It is a straightforward consequence of the proof given in Section 4.1 that $y_c = \sqrt{1 + \sqrt{2}}$ in the edge weighted case. The proof of this result, in outline, is the following: We denote the generating functions A and B, as defined in Section 4.1 for the vertex case, by subscript v (for vertex). We denote the corresponding generating functions for the case with edge weighting with the subscript e. Then it is clear by inspection that $A_e(x_c, y) = A_v(x_c, y^2)$, as every time a walk contributing to the A generating function passes through n surface vertices, whether adjacent or not, it must pass through $2n$ surface edges.

By the same argument, every time a walk contributing to the B generating function passes through n surface vertices, whether adjacent or not, it must pass through $2n - 1$ surface edges. This then gives rise to $B_e(x_c, y) = \frac{1}{y} B_v(x_c, y^2)$. From either of these two equations it follows that $y_c(\text{vertex}) = (y_c(\text{edge}))^2$, hence $y_c(\text{edge}) = \sqrt{1 + \sqrt{2}}$.

We now consider the double-vertex weighted model. We generated data for $A_T(x_c, y)$ for $T \leq 14$ as described in Subsection B.1.2, and found the intersection points where $A_T(x_c, y) = A_{T+1}(x_c, y)$, which defines $y^*(T)$. These data are tabulated in Table 4.1. Extrapolating $y^*(T)$ as described above, we estimate

$$y_c(\text{double}) = 1.46767.$$

We also find, by an identical method of extrapolation, that $A(x_c, y_c) = 2.613$, which is probably exactly $\sec(3\pi/8)$, as is the case when considering interactions with surface vertices, see (4.34).

4.2.2 Square lattice

We next consider data for the square lattice, with every surface vertex carrying a fugacity y. We generated data for $A_T(x_c, y)$ for $T \leq 15$ as described in Subsection B.1.2, and found the intersection points where $A_T(x_c, y) = A_{T+1}(x_c, y)$, which defines $y^*(T)$. These data are tabulated

Table 4.1: Estimates of $y^*(T)$ and $A_T(x_c, y^*(T))$ for the honeycomb lattice double-vertex model.

T	$y^*(T)$	$A_T(x_c, y^*(T))$
1	1.474342684974343	2.758023465753132
2	1.471231066324457	2.699581979117133
3	1.469859145369675	2.671309655463187
4	1.469144651946551	2.655387366045945
5	1.468728339703417	2.645467247042683
6	1.468465540675101	2.638829094236329
7	1.468289428840316	2.634145423791235
8	1.468122140755486	2.629489693948282
9	1.468008309717543	2.626054066036805
10	1.467956382495343	2.624432487387554
11	1.467915603443970	2.623117304368586
12	1.467883002922926	2.622033892173660
13	1.467856536243392	2.621129346334020

in Table 4.2. Extrapolating $y^*(T)$ as described above, we estimate

$$y_c(\text{vertex}) = 1.77564.$$

We have also found, by an identical method of extrapolation, that $A(x_c, y_c) = 2.678405$, which is $1.024981/\cos(3\pi/8)$. In Subsection 2.2.2 we found, for the non-interacting case (corresponding to $y = 1$), $A(x_c, 1) = 2.678365 = 1.024966/\cos(3\pi/8)$. Thus there appears to be a very weak y dependence. (In the normalisation of the generating function $A_T(x_c, y)$ used here, two extra half-steps are included, giving an extra factor of the step fugacity x_c, compared to the value that would be quoted if contributing walks started and ended *on* the surface. This explains the difference between the values quoted in Table 4.2 and the ordinates in Figure 4.11.)

Table 4.3 shows the corresponding data for the edge weighted case. Extrapolating $y^*(T)$ as described above, we estimate

$$y_c(\text{edge}) = 2.040135.$$

We also find that $A(x_c, y_c) = 2.678405$, which is $1.024981/\cos(3\pi/8)$. In [12] we found, for the non-interacting case (corresponding to $y = 1$), $A(x_c, 1) = 2.6783 = 1.0249/\cos(3\pi/8)$. This is too imprecise to see any evidence of y dependence.

4.2.3 Triangular lattice

We next consider data for the triangular lattice, with every surface vertex carrying a fugacity y. We generated data for $A_T(x_c, y)$ for $T \leq 11$ as described in Subsection B.1.2, and found the

Table 4.2: Estimates of $y^*(T)$ and $A_T(x_c, y^*(T))$ for the square lattice surface vertex model.

T	$y^*(T)$	$A_T(x_c, y^*(T))$
1	1.781782909906119	2.748677355944862
2	1.778386591113354	2.715115253913871
3	1.777378005442640	2.704018907440273
4	1.776850407093364	2.697681121136133
5	1.776527700942633	2.693512738663579
6	1.776316359764735	2.690608915840792
7	1.776170974231462	2.688500944397294
8	1.776066934443028	2.686918847615982
9	1.775990033953699	2.685698355993929
10	1.775931645420429	2.684735010917280
11	1.775886299456907	2.683959815456866
12	1.775850398954429	2.683325675630414
13	1.775821502307431	2.682799521958416
14	1.775797906369155	2.682357553489197

Table 4.3: Estimates of $y^*(T)$ and $A_T(x_c, y^*(T))$ for the square lattice surface edge model.

T	$y^*(T)$	$A_T(x_c, y^*(T))$
1	2.023317607727152	2.519464246890523
2	2.031649211433080	2.585125356952430
3	2.035085448834840	2.616332757155513
4	2.036771224259312	2.633293109539552
5	2.037723730407517	2.643677266387231
6	2.038317002192238	2.650588857893349
7	2.038712823877066	2.655469267857106
8	2.038990695898482	2.659069610531442
9	2.039193569770578	2.661816780067225
10	2.039346383471084	2.663969985883853
11	2.039464457297598	2.665695001241074
12	2.039557641399558	2.667102372510593
13	2.039632511102958	2.668268404182947
14	2.039693596208206	2.669247312794744

Table 4.4: Estimates of $y^*(T)$ and $A_T(x_c, y^*(T))$ for the triangular lattice surface vertex model.

T	$y^*(T)$	$A_T(x_c, y^*(T))$
1	2.169017975620833	5.299883162257977
2	2.152124186067447	5.089804987842667
3	2.147952081330057	5.033100087535114
4	2.146325209334416	5.009022287728647
5	2.145537862947824	4.996485228732837
6	2.145102964455591	4.989109337635192
7	2.144840361941141	4.984402909686655
8	2.144671215263562	4.981219362650799
9	2.144556764080381	4.978968525942606
10	2.144476246964690	4.977320728801566

intersection points where $A_T(x_c, y) = A_{T+1}(x_c, y)$, which defines $y^*(T)$. These data are tabulated in Table 4.4. Extrapolating $y^*(T)$ as described above, we estimate

$$y_c(\text{vertex}) = 2.144181.$$

We have also found, by an identical method of extrapolation, that $A(x_c, y_c) = 4.97002$, which is $1.901944/\cos(3\pi/8)$. In Subsection 2.2.3 we found, for the non-interacting case (corresponding to $y = 1$), $A(x_c, 1) = 4.970111 = 1.901979/\cos(3\pi/8)$. Thus there again appears to be a very weak y dependence.

Table 4.5: Estimates of $y^*(T)$ and $A_T(x_c, y^*(T))$ for the triangular lattice surface edge model.

T	$y^*(T)$	$A_T(x_c, y^*(T))$
1	2.933665548671216	4.793416679321919
2	2.939352607034002	4.841229819027843
3	2.942788011875285	4.873934294210283
4	2.944814166604381	4.895179517868169
5	2.946090146548846	4.909648090189844
6	2.946944189466541	4.919989731979732
7	2.947544335340955	4.927679988442194
8	2.947982663246637	4.933582932189477
9	2.948312910101248	4.938231892866670
10	2.948568146735367	4.941971526310544

Table 4.5 shows the corresponding data for the edge weighted case. Extrapolating $y^*(T)$ as described above, we estimate

$$y_c(\text{edge}) = 2.950026.$$

We also find that $A(x_c, y_c) = 4.9696$, which is $1.90178/\cos(3\pi/8)$. In Subsection 2.2.3 we found, for the non-interacting case (corresponding to $y = 1$), $A(x_c, 1) = 4.970111 = 1.901979/\cos(3\pi/8)$. Again, there is evidence of weak y dependence.

4.3 Solvable models

In Chapter 3 we considered a number of subclasses of SAWs and SAPs. Our primary motivation in doing so is that many such models are *solvable* – we can rigorously prove explicit expressions for their generating functions or their series coefficients. This allows for a level of scrutiny which may not be possible for general SAWs or SAPs; moreover, the solutions to such models may help to shed light on the general models.

Such an approach has also been used extensively in the study of polymer models, and in particular, models of polymer adsorption. For relatively simple models we are able to determine exact expressions for quantities of interest (in particular, generating functions and the free energy), and these allow for detailed analyses of phase transitions and other phenomena.

In this section we return to the methodology of Chapter 3, and consider subclasses of SAWs for which we can derive functional equations and, hopefully, exact solutions to generating functions. Here, however, these SAW models will be interacting with an impenetrable surface, and thus their generating functions will include a fugacity y associated with visits to the surface. We first briefly review some existing solvable models of polymer adsorption, which all have a directedness restriction. We then introduce some new models, based on the prudent walks discussed in Chapter 3, which do not have a directedness restriction, and show that in some cases these models are exactly solvable. We believe these are the only known solvable models of polymer adsorption which are not directed; moreover, they display interesting phase transition properties not seen in the simpler directed models.

Throughout this section the impenetrable surface will always be the x-axis, and the walks will be confined to the upper half-plane of the lattice.

4.3.1 Directed walks

Recall from Chapter 3 that a SAW is *directed* if it is forbidden from stepping in at least one direction on its lattice. More specifically, if S is a set of possible step directions on a lattice, then a walk is S-directed if it only takes steps from S. Below we summarise some of the most commonly-used directed models of polymer adsorption.
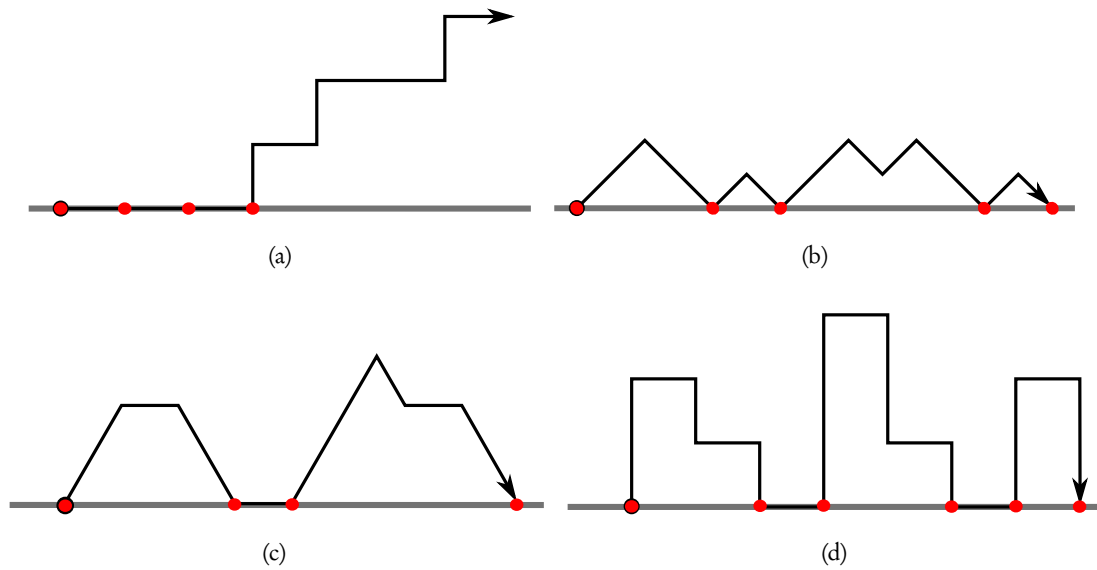
Figure 4.13: Examples of directed polymer adsorption models, with vertex weights indicated: (a) a fully directed walk, (b) a Dyck path, (c) a Motzkin path and (d) a partially directed walk.

Fully directed walks

This is essentially the simplest two-dimensional model of polymer adsorption, and we will use it as an example to demonstrate the techniques employed later in this section. We consider walks on the square lattice, and only allow north or east steps. (So our walks are NE-directed – see Figure 4.13 for an example.) Clearly no surface interactions can occur after a walk has taken a north step, so we can split walks into two independent pieces: the 'interacting' phase (when the walk remains on the surface) and the 'bulk' phase (when the walk has left the surface).

If $f_{n,v}$ is the number of NE-directed walks of length n with v *edges* along the x-axis, then the generating function

$$F(t;y) = \sum_{n,v} f_{n,v} t^n y^v$$

is given by

$$F(t;y) = \frac{1}{1 - ty} \cdot \frac{1 - t}{1 - 2t}. \tag{4.35}$$

(The first factor is the generating function for the interacting phase and the second is the generating function of the bulk phase.)

If we restrict y to be real and non-negative, then it is clear by inspection that the dominant singularity $t_c(y)$ of $F(t;y)$ for a given y is

$$t_c(y) = \begin{cases} 1/2 & y \leq 2 \\ 1/y & y > 2. \end{cases}$$

Recalling the relationship (4.2) between the dominant singularity and the free energy, we have

that

$$x(y) = -\log t_c(y) = \begin{cases} \log 2 & y \le 2 \\ \log y & y > 2. \end{cases}$$

So for fully directed walks the adsorption phase transition occurs at $y = y_c = 2$. Moreover, this transition is *first-order*, as the first derivative $x'(y)$ is discontinuous at $y = 2$. The crossover exponent (defined at the beginning of this chapter) is $\phi = 1$.

In the limit of infinitely long polymers, the mean density $\delta(y)$ of steps in the surface is given by

$$\delta(y) = y \frac{dx}{dy} = \begin{cases} 0 & y < 2 \\ 1 & y > 2. \end{cases}$$

Note that a fully directed walk with n *edges* in the x-axis will have precisely $n + 1$ *vertices* in the surface, so the generating function for the vertex-weighted model is simply $yF(t; y)$. This doesn't change the singular behaviour of the generating function, so the free energy etc. remain unchanged.

Dyck/ballot paths

These are fully directed walks on the (45° rotated) square lattice, formed by allowing only north-east and south-east steps. A *Dyck path* must start and end on the surface, while a *ballot path* need only start on the surface. See [110, 113] for further details, and Figure 4.13 for an example. Since there are no steps parallel to the surface, only the vertex-weighted model can be considered.

If we define

$$S(t) \equiv S = \frac{1 - \sqrt{1 - 4t^2}}{2t}$$
$$= t + t^3 + 2t^5 + 5t^7 + 14t^9 + O(t^{10}),$$

then the two-variable generating function of Dyck paths (with t conjugate to length and y conjugate to surface contacts) is

$$D(t; y) = \frac{y}{1 - tSy} \tag{4.36}$$
$$= y + t^2 y^2 + t^4(y^2 + y^3) + t^6(2y^2 + 2y^3 + y^4) + O(t^8),$$

while the generating function of ballot paths is

$$B(t; y) = \frac{y(1 - t - tS)}{(1 - 2t)(1 - tSy)} \tag{4.37}$$
$$= y + ty + t^2(y + y^2) + t^3(2y + y^2) + t^4(3y + 2y^2 + y^3) + O(t^5).$$

The dominant singularity for both generating functions is

$$t_c(y) = \begin{cases} 1/2 & y \leq 2 \\ \dfrac{\sqrt{y-1}}{y} & y > 2, \end{cases}$$

and thus the free energy is given by

$$\varkappa(y) = -\log t_c(y) = \begin{cases} \log 2 & y \leq 2 \\ \log\left(\dfrac{y}{\sqrt{y-1}}\right) & y > 2. \end{cases}$$

So as with fully directed walks, the adsorption phase transition occurs at $y = y_c = 2$. This time, however, it is a *second-order* transition, as the derivative $\varkappa'(y)$ is continuous at $y = 2$. The crossover exponent here is $\phi = 1/2$, since

$$\log\left(\frac{y}{\sqrt{y-1}}\right) - \log 2 \sim \frac{1}{2}(\log y - \log 2)^2 \quad \text{as } y \to 2^+.$$

Finally, the mean density of monomers in the surface (in the limit of infinitely long walks) is given by

$$\delta(y) = y\frac{d\varkappa}{dy} = \begin{cases} 0 & y \leq 2 \\ \dfrac{2-y}{2(1-y)} & y > 2. \end{cases}$$

Note here that $\delta(y) \to 1/2$ as $y \to \infty$; since Dyck paths must always have even length, it is common to compute the generating function for half-length instead of length. In that case we would observe $\delta(y) \to 1$ in the limit.

Motzkin paths

These are generated on the triangular lattice by allowing north-east, east and south-east steps. See [114] for further details, and Figure 4.13 for an example. Both edge- and vertex-weightings are possible, and unlike fully directed walks, the relationship between the two is non-trivial.

For the edge-weighted model, the free energy is given by

$$\varkappa(y) = \begin{cases} \log 3 & y \leq 2 \\ \log\left(\dfrac{1-y+y^2}{y-1}\right) & y > 2, \end{cases}$$

and so the phase transition occurs at $y = y_c = 2$ and is second-order, and the crossover exponent is $\phi = 1/2$. The density of monomers in the surface is then

$$\delta(y) = \begin{cases} 0 & y \leq 2 \\ \dfrac{y^2(2-y)}{(1-y)(1-y+y^2)} & y > 2. \end{cases}$$

For the vertex-weighted model, the free energy is given by

$$\varkappa(y) = \begin{cases} \log 3 & y \leq 3/2 \\ \log\left(\dfrac{2y}{1-y+\sqrt{(y-1)(y+3)}}\right) & y > 3/2, \end{cases}$$

and so the phase transition occurs at $y = y_c = 3/2$ and is second-order. The crossover exponent is $\phi = 1/2$. The density of monomers in the surface is then

$$\delta(y) = \begin{cases} 0 & y \leq 3/2 \\ \dfrac{-6+y+y^2+y\sqrt{(y-1)(y+3)}}{2(y-1)(y+3)} & y > 3/2. \end{cases}$$

Partially directed walks

These are generated on the square lattice by allowing north, east and south steps. See [98, 46, 119] for further details, and Figure 4.13 for an example.

For the edge-weighted model, the free energy is given by

$$\varkappa(y) = \begin{cases} \log(1+\sqrt{2}) & y \leq (2+\sqrt{2})/2 \\ -\log f(y) & y > (2+\sqrt{2})/2, \end{cases}$$

where $f(y)$ is a root of

$$1 - y - y(1-y)f(y) + yf(y)^2 + y(1-y)f(y)^3 = 0.$$

(We are unable to find a neat explicit expression for $f(y)$.) The phase transition is second-order and occurs at $y = y_c = (2+\sqrt{2})/2$, and the crossover exponent is $\phi = 1/2$. In the limit of infinitely long polymers, the density of edges in the surface is given by

$$\delta(y) = \begin{cases} 0 & y \leq (2+\sqrt{2})/2 \\ \dfrac{-yf'(y)}{f(y)} & y > (2+\sqrt{2})/2. \end{cases}$$

For the vertex-weighted model, the free energy is given by

$$\varkappa(y) = \begin{cases} \log(1+\sqrt{2}) & y \leq (1+\sqrt{2})(\sqrt{5}-1)/2 \\ -\log g(y) & y > (1+\sqrt{2})(\sqrt{5}-1)/2, \end{cases}$$

where $g(y)$ is a root of

$$1 - y^2 - y(1-y)(2+y)g(y) + y^2(2-y)g(y)^2 + y^2(1-y)g(y)^3 - y^3(1-y)g(y)^4 = 0.$$

(Again, we are unable to find a neat explicit expression for $g(y)$.) The phase transition is second-order and occurs at $y = y_c = (1 + \sqrt{2})(\sqrt{5} - 1)/2$, and the crossover exponent is $\phi = 1/2$. In the limit of infinitely long polymers, the density of vertices in the surface is given by

$$\delta(y) = \begin{cases} 0 & y \leq (1 + \sqrt{2})(\sqrt{5} - 1)/2 \\ \frac{-y g'(y)}{g(y)} & y > (1 + \sqrt{2})(\sqrt{5} - 1)/2. \end{cases}$$

4.3.2 Square lattice prudent walks

In this subsection and the next, we introduce several new models of polymer adsorption, some of which have the pleasing property of being able to step in all directions (i.e. four on the square lattice and six on the triangular) on their respective lattices. This means that, in a sense, they interpolate between the directed walks discussed in the previous subsection and the general model of adsorbing SAWs. On both lattices we will find solvable non-directed models; it turns out, however, that those on the triangular lattice are somewhat easier to analyse. We will restrict our investigation to the edge-weighted models; most of the methodology should be the same for the vertex-weighted models, and we hope to consider those cases in a future study.

Recall from Section 3.2 that a SAW is *prudent* if it never takes a step towards a previously-visited vertex. On the square lattice, this forces a prudent walk to end on the boundary of its *bounding box* – the smallest lattice rectangle enclosing the entire walk. We thus form a sub-classification of prudent walks according to which sides of their box they end on: *1-sided* prudent walks always end on the east side of their box, *2-sided* walks always end on the east or north sides, and *3-sided* walks always end on the east, north or west sides.

When no surface is present, we also define *4-sided* or *unrestricted* prudent walks to be those walks which can end anywhere on their box. The imposition of the impenetrable surface in the x-axis, however, means that the south side of the box always lies in the surface, and thus every prudent walk which starts at the surface will be at most 3-sided. See Figure 4.14 for an illustration.

The simplest type of prudent walks, 1-sided, coincides exactly with partially directed walks, and the critical behaviour of that model was discussed in the previous subsection.

2-sided prudent walks

Our method for solving the generating function of 2-sided prudent walks follows along the same lines as that used by Bousquet-Mélou [17] to solve the model without a surface. She was able to exploit the symmetry of walks in the line $y = x$; we are unable to do so here, so we instead have

Figure 4.14: A 3-sided prudent walk above an impenetrable surface. In order to become "4-sided", the walk would have to step towards the east from the south-west corner of its box, but such a step can never be prudent.

(a)   (b)

Figure 4.15: 2-sided prudent walks above an impenetrable surface, (a) ending on the right of the box and (b) ending on the top, with the distances i and j indicated.

to use two generating functions. Define the generating functions

$$R(t;u,v;y) = \sum_{n,i,j,\nu} r_{n,i,j,\nu} t^n u^i v^j y^\nu$$

$$T(t;u,v;y) = \sum_{n,i,j,\nu} t_{n,i,j,\nu} t^n u^i v^j y^\nu$$

where $r_{n,i,j,\nu}$ (resp. $t_{n,i,j,\nu}$) is the number of n-step 2-sided prudent walks which start on a horizontal impenetrable surface and end on the east (resp. north) side of their bounding box, with distance i from the endpoint to the north (resp. east) of the box, distance j from the endpoint to the surface, and ν steps along the surface. See Figure 4.15 for examples of the two types of 2-sided prudent walks, with the distances measured by the catalytic variables indicated.

Lemma 4.18. *The generating functions $R(t;u,v;y)$ and $T(t;u,v;y)$ satisfy the functional relations*

$$L(t;u,v)T(t;u,v;y) = \frac{1}{1-tuy} - \frac{t^2v}{u-t}T(t;t,v;y) + tR(t;tv,v;y) - t(1-y)R(t;tv,0;y)$$

$$(4.38)$$

and

$$M(t;u,v)R(t;u,v;y) = 1 + tvT(t;t,v;y) - \frac{t^2v}{u-tv}R(t;tv,v;y) - \frac{t^2u}{v-tu}R(t;u,tu;y)$$
$$- \frac{tu(1-y)}{u-tv}R(t;u,0;y) + \frac{t^2v(1-y)}{u-tv}R(t;tv,0;y), \quad (4.39)$$

where

$$L(t;u,v) = 1 - \frac{tuv(1-t^2)}{(u-t)(1-tu)}$$
$$M(t;u,v) = 1 - \frac{tuv(1-t^2)}{(v-tu)(u-tv)}.$$

Proof. We will go through the derivation of (4.38); much of the same reasoning applies for (4.39). We recursively construct walks counted by T by considering their last inflating step - the last step which moved either the north or east boundaries of their box.

- Walks with no inflating steps must be empty or contain only west steps; the generating function of such walks is

$$\frac{1}{1-tuy}.$$

- Walks whose last inflating step was north can be split into three parts: the section which came before the inflating step (which could be any walk counted by T), the inflating step itself, and the steps which came after the inflating step (which must be east or west). The generating function of these walks is thus

$$\sum_{n,i,j,\nu} t_{n,i,j,\nu} t^n y^\nu \cdot tv^{j+1}\left(\sum_{l=0}^{i} t^l u^{i-l} + \sum_{m=1}^{\infty} t^m u^{i+m}\right)$$

(The sum over l is for walks with east steps following the inflating step; the sum over m is for those with west steps following the inflating step.)

$$= tv\sum_{n,i,j,\nu} t^n y^\nu v^j\left(\frac{u^{i+1} - t^{i+1}}{u-t} + \frac{tu^{i+1}}{1-tu}\right)$$
$$= \frac{tv}{u-t}(uT(t;u,v;y) - tT(t;t,v;y)) + \frac{t^2uv}{1-tu}T(t;u,v;y).$$

- Walks whose last inflating step was east can also be split into three parts: the part of the walk which came before the inflating step (which could be any walk counted by R), the inflating step itself, and the steps which came after the inflating step (which must be north). The generating function of these walks is thus

$$\sum_{n,i,j,\nu} r_{n,i,j,\nu} t^n y^\nu \cdot t^{i+1}v^{i+j}$$
$$= tR(t;tv,v;y).$$

176

However, the situation is slightly different if that inflating step was along the surface – in that case, the walk should have acquired an extra weight of y. We can correct this by subtracting those walks from the above sum and adding them back with the extra weight:

$$-tR(t;tv,0;y)+tyR(t;tv,0;y).$$

All the above contributions must add exactly to $T(t;u,v;y)$; rearranging the resulting equation gives (4.38). ∎

Lemma 4.19. *The generating functions $T(t;u,v;y)$ and $R(t;u,v;y)$ satisfy the functional equation*

$$M(t;u,v)R(t;u,v;y)=A(t;u,v;y)+B(t;u,v;y)R(t;u,0;y)+C(t;u,v)T(t;t,v;y) \quad (4.40)$$

where

$$\Lambda(t;v)=\frac{1+t^2-tv+t^3v-\sqrt{(1+t^2-tv+t^3v)^2-4t^2}}{2t}$$

$$A(t;u,v;y)=\frac{v(u-t^2u-tu\Lambda(t;v)y+t^2v\Lambda(t;v)y)}{(u-tv)(v-tu)(1-t\Lambda(t;v)y)}$$

$$B(t;u,v;y)=-\frac{tu(u+v-tv-t^2v-vy+t^2vy)}{(u-tv)(v-tu)}$$

$$C(t;u,v)=-\frac{tv(tu-u\Lambda(t;v)+tv\Lambda(t;v))}{(\Lambda(t;v)-t)(u-tv)}$$

Proof. Setting $v=0$ in (4.39) yields

$$(1+t-tx)R(t;u,0;y)=1+tR(t;u,tu;y),$$

which can be used to eliminate $R(t;u,tu;y)$ from (4.39). Meanwhile, $\Lambda(t;v)$ satisfies the equation $L(t;\Lambda(t;v),v)=0$ (as does another function of v, but Λ is the only root which is a power series in t and thus the only one which will eventually give a well-defined solution), so substituting $u=\Lambda(t;v)$ into (4.38) cancels the LHS. The resulting equation can be written as

$$-\frac{t^2v}{u-tv}R(t;tv,v;y)+\frac{t^2v(1-y)}{u-tv}R(t;tv,0;y)$$

$$=\frac{tv}{(u-tv)(1-t\Lambda(t;v)y)}-\frac{t^3v^2}{(u-tv)(\Lambda(t;v)-t)}T(t;t,v;y), \quad (4.41)$$

which can then be used to eliminate $R(t;tv,v;y)$ and $R(t;tv,0;y)$ from (4.39). Simple manipulation gives (4.40) in the stated form. ∎

Proposition 4.20. *The form of $R(t;u,0;y)$ is given by*

$$R(t;u,0;y)=\sum_{n=0}^{\infty}\mathscr{H}(t;u\bar{\Lambda}(t)^{2n};y)\prod_{k=0}^{n-1}\mathscr{I}(t;u\bar{\Lambda}(t)^{2k};y) \quad (4.42)$$

177

when considered as a formal power series in t, u, y, where

$$\bar{\Lambda} = \Lambda(t;1) = \frac{1 - t + t^2 + t^3 - \sqrt{1 - 2t - t^2 - t^4 + 2t^5 + t^6}}{2t}$$

$$\mathcal{H}(t;z;y) = \frac{t(1 - \bar{\Lambda}^2)(\bar{\Lambda} - t^2\bar{\Lambda} + ty(t - 2\bar{\Lambda} + t\bar{\Lambda}^2)\Lambda(t;z\bar{\Lambda}) + y(\bar{\Lambda} - t)(1 - t\bar{\Lambda})\Lambda(t;z\bar{\Lambda})^2)}{(1 - t\bar{\Lambda})(1 - t\Lambda(t;z\bar{\Lambda})y)(1 + \bar{\Lambda} - t\bar{\Lambda} - t^2\bar{\Lambda} - \bar{\Lambda}y + t^2\bar{\Lambda}y)(t\bar{\Lambda} - (\bar{\Lambda} - t)\Lambda(t;z\bar{\Lambda}))}$$

$$\mathcal{I}(t;z;y) = \frac{\bar{\Lambda}(\bar{\Lambda} - t)(1 - t - t^2 - y + t^2 y + \bar{\Lambda})(t - (1 - t\bar{\Lambda})\Lambda(t;z\bar{\Lambda}))}{(1 - t\bar{\Lambda})(1 + \bar{\Lambda} - t\bar{\Lambda} - t^2\bar{\Lambda} - \bar{\Lambda}y + t^2\bar{\Lambda}y)(t\bar{\Lambda} - (\bar{\Lambda} - t)\Lambda(t;z\bar{\Lambda}))}$$

Proof. Equation (4.19) is susceptible to the iterative kernel method [116]. The kernel M is cancelled at $(u,v) = (u, u\bar{\Lambda})$ (and since $M(t;u,v) = M(t;v,u)$, it also disappears at $(u,v) = (v\bar{\Lambda}, v)$), and so by substituting the pairs

$$(u,v) = (u, u\bar{\Lambda}) \quad \text{and} \quad (u,v) = (u\bar{\Lambda}^2, u\bar{\Lambda}),$$

we obtain two equations in $R(t;u,0;y), R(t;u\bar{\Lambda}^2, 0;y)$ and $T(t;t,u\bar{\Lambda};y)$. One can then eliminate $T(t;t,u\bar{\Lambda};y)$ to obtain

$$R(t;u,0;y) = \mathcal{H}(t;u;y) + \mathcal{I}(t;u;y)R(t;u\bar{\Lambda}^2, 0;y) \tag{4.43}$$

where

$$\mathcal{H}(t;u;y) = -\frac{A(t;u,u\bar{\Lambda};y)}{B(t;u,u\bar{\Lambda};y)} + \frac{C(t;u,u\bar{\Lambda})A(t;u\bar{\Lambda}^2, u\bar{\Lambda};y)}{B(t;u,u\bar{\Lambda};y)C(t;u\bar{\Lambda}^2, u\bar{\Lambda})}$$

$$\mathcal{I}(t;u;y) = \frac{C(t;u,u\bar{\Lambda})B(t;u\bar{\Lambda}^2, u\bar{\Lambda};y)}{B(t;u,u\bar{\Lambda};y)C(t;u\bar{\Lambda}^2, u\bar{\Lambda})}$$

After simplification $\mathcal{H}$ and $\mathcal{I}$ take the form given in the proposition.

Iterating (4.43) gives (4.42), as long as everything converges as a formal power series. But now $\mathcal{H}(t;u;y)$ is a power series in t with coefficients in $\mathbb{Z}[u,y]$ of the form $1 + ty + O(t^2)$, and likewise $\mathcal{I}(t;u;y)$ is a power series of the form $t^4(1 - y - u + uy) + O(t^5)$. So the sum converges as a formal power series in t with coefficients in $\mathbb{Z}[u,y]$. ∎

Theorem 4.21. *The generating functions $R(t;u,v;y)$ and $T(t;u,v;y)$ have the solutions*

$$R(t;u,v;y) = \frac{1}{M(t;u,v)}\left(A(t;u,v;y) + B(t;u,v;y)R(t;u,0;y) + C(t;u,v)T(t;t,v;y)\right)$$

$$T(t;u,v;y) = \frac{1 - tu}{L(t;u,v)}\left((t - u)\left(\frac{1}{1 - tuy} - \frac{1}{1 - t\Lambda(t;v)y}\right)\right.$$

$$\left. + t^2 v\left(1 + \frac{t - u}{\Lambda(t;v) - t}\right)T(t;t,v;y)\right)$$

where

$$T(t;t,v;y) = \frac{-1}{C(t;v\bar{\Lambda}, v)}(A(t;v\bar{\Lambda}, v;y) + B(t;v\bar{\Lambda}, v;y)R(t;v\bar{\Lambda}, 0;y))$$

and $R(t; u, 0; y), R(t; v\bar{\Lambda}, 0; y)$ are given by (4.42). The overall generating function for 2-sided pru-dent walks above an impenetrable surface is then

$$W_\square(t; u, v; y) = \sum_{n,i,j,\nu} w^\square_{n,i,j,\nu} t^n u^i v^j y^\nu$$

$$= R(t; u, v; y) + T(t; u, v; y) - R(t; u, 0; y)$$

where $w^\square_{n,i,j,\nu}$ is the number of n-step 2-sided prudent walks which end a distance i from the north-east corner of their box and a distance j above the surface, with ν steps along the surface.

Proof. Substituting $(u, v) = (v\bar{\Lambda}, v)$ into (4.40) cancels the LHS, and rearranging gives the stated form of $T(t; t, v; y)$. Since $v\bar{\Lambda}$ is a power series in t with coefficients in $\mathbb{Z}[v]$, $R(t; v\bar{\Lambda}, 0; y)$ is well-defined as a formal power series. Rearranging (4.40) then gives $R(t; u, v; y)$.

Equation (4.41) can be rewritten as

$$t R(t; tv, v; y) - t(1-y)R(t; tv, 0; y) = \frac{-1}{1 - t\Lambda(t; v)y} + \frac{t^2 v}{\Lambda(t; v) - t} T(t; t, v; y),$$

and this can be used to eliminate $R(t; tv, v; y)$ and $R(t; tv, 0; y)$ from (4.38). Rearranging the resulting equation gives the stated expression for $T(t; u, v; y)$.

The overall generating function $W_\square(t; u, v; y)$ is then found by adding the generating functions for walks ending on the right and the top of the box, and we subtract $R(t; u, 0; y)$ because walks ending at the north-east corner have been counted twice. ∎

While we are quite certain of the location of the dominant singularity of $W_\square(t; 1, 1; y)$ for all $y \geq 0$, we have been unable to complete a rigorous proof. We will thus present the result as just a conjecture. (Fortunately we will be able to go further for the triangular lattice.)

Conjecture 4.22. *For a given $y \geq 0$, the dominant singularity of $W_\square(t; 1, 1; y)$ is located at*

$$t_c(y) = \begin{cases} \alpha \approx 0.403032 & y \leq 2 \\ h(y) & y > 2, \end{cases}$$

where α is a root of $1 - 2\alpha - 2\alpha^2 + 2\alpha^3 = 0$, and $h(y)$ is a root of

$$1 - y - y(1-y)h(y) + yh(y)^2 + y(1-y)h(y)^3 = 0.$$

The free energy of 2-sided prudent walks adsorbing onto an impenetrable surface is thus given by $\varkappa(y) = -\log t_c(y)$, and the adsorption phase transition occurs at $y = y_c = 2$ and is first-order. The crossover exponent is $\phi = 1$. In the limit of infinitely long polymers, the mean density of monomers (i.e. edges) in the surface is

$$\delta(y) = \begin{cases} 0 & y < 2 \\ \frac{-yh'(y)}{h(y)} & y > 2. \end{cases}$$

179

(See Figure 4.16 for plots of the dominant singularity and surface density.) The constant part of the dominant singularity (i.e. the radius of convergence in the desorbed phase) appears to be a pole in the $T(t;t,1;y)$ term, which appears in both $R(t;1,1;y)$ and $T(t;1,1;y)$. More specifically, α is a root of $C(t;\bar{\Lambda},1) = 0$. The dominant singularity in the adsorbed phase appears as a pole of $R(t;1,0;y)$ and $R(t;\bar{\Lambda},0;y)$. In fact, $h(y)$ is a root (in the variable t) of

$$1 + \bar{\Lambda} - t\bar{\Lambda} - t^2\bar{\Lambda} - \bar{\Lambda}y + t^2\bar{\Lambda}y = 0,$$

which in turn makes it a root of $B(t;1,\bar{\Lambda};y) = 0$ (and in fact a root of $B(t;\bar{\Lambda}^n,\bar{\Lambda}^{n+1};y) = 0$ for $n \geq 0$), and thus a pole of $\mathscr{H}(t;\bar{\Lambda}^n;y)$ and $\mathscr{I}(t;\bar{\Lambda}^n;y)$ for all $n \geq 0$.

Series analysis of $W_\square(t;1,1;y)$ for a variety of y values confirms the validity of this conjecture. We point out that the dominant singularity in the desorbed phase, $\alpha \approx 0.403032$, is the same as for 2-sided prudent walks *without* a boundary (see Corollary 3.7).

The fact that this model undergoes a *first-order* adsorption transition is perplexing. On the square lattice, we have the scheme

partially directed walks $\subset$ *2-sided prudent walks* $\subset$ *general SAWs.*

But now as we saw in the previous subsection, partially directed walks undergo a second-order transition, and general SAWs are expected [117, 104] to do the same. It seems natural, then, to expect that a model like 2-sided prudent walks, which 'interpolates' between PDWs and SAWs, would exhibit the same behaviour.

At first, our intuition was that the phase transition for prudent walks might be first-order because the 'average' 2-sided prudent walk, in the desorbed phase, moves away from the surface at a speed which is linearly proportional to its length. That is, if we denote by $\langle S_n \rangle$ the mean distance between the endpoint of a 2-sided prudent walk of length n and the surface, then we expect $\langle S_n \rangle \sim cn$ for some positive constant c. This conclusion arises via [44, Theorem IX.9], the conditions of which we believe to be satisfied by $W_\square(t;1,v;1)$.

That theorem essentially states that if a bivariate generating function like $F(z,u)$ satisfies a certain set of conditions (a *meromorphic schema*), then the sequence of random variables X_n with probability generating functions

$$p_n(u) = \frac{[z^n]F(z,u)}{[z^n]F(z,1)}$$

converges to a Gaussian random variable, and the mean and standard deviation of X_n are asymptotically linear in n. If we use $F(z,u) = W_\square(z;1,u;1)$ (here the variable u is conjugate to the distance between the endpoint of a walk and the surface) and find that the meromorphic schema is satisfied, then this would precisely imply that $\langle S_n \rangle \sim cn$. This would demonstrate a similarity between 2-sided prudent walks and the fully directed walks discussed in the previous subsection – for that model, it is clear that $\langle S_n \rangle \sim n/2$. Recall that fully directed walks were the only model considered which displayed a first-order phase transition.
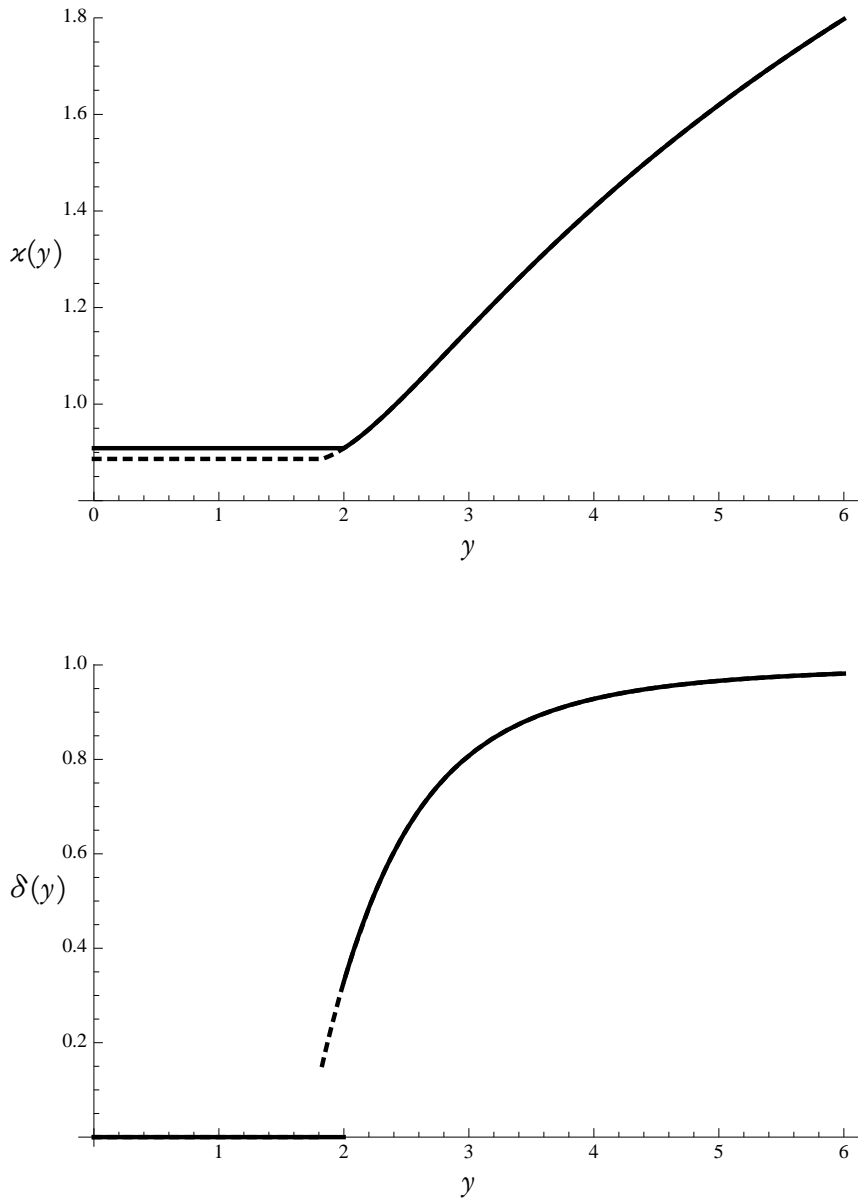
Figure 4.16: Top: The (conjectured) free energy of 2-sided prudent walks (solid) and loops (dashed) as function of the surface fugacity y. Bottom: The (conjectured) density of edges in the surface for 2-sided prudent walks (solid) and loops (dashed). Note the discontinuity in the surface density for both models, corresponding to a first-order adsorption transition.

On the other hand, it is straightforward to show that for partially directed walks, $\langle S_n \rangle \sim$ $cn^{1/2}$. Moreover, for SAWs we would expect that $\langle S_n \rangle \sim cn^{\nu}$, where 2ν is the exponent which characterises the mean squared end-to-end distance (among other measurements) of two-dimensional SAWs. Recall from Chapter 1 that ν is expected to be 3/4 in two dimensions.[10]

So while 2-sided prudent walks do indeed interpolate between the smaller class of partially directed walks and the larger class of general SAWs, we expect them to move away from the surface faster than either of the other two models. In this respect, 2-sided prudent walks more closely resemble fully directed walks, a model which we know undergoes a first-order adsorption transition.

However, on closer inspection this reasoning seems to break down. If it was indeed the case that 2-sided prudent walks undergo a first-order transition because they drift away from the surface faster than PDWs or general SAWs, then it would be worthwhile to investigate a subset of prudent walks which remain close to the surface. In particular, (4.42) gives us the form of $R(t; u, 0; y)$, and then

$$W_{\square}(t; 1, 0; y) = R(t; 1, 0; y) + \frac{ty}{1 - ty}.$$

(The second term accounts for walks comprising a sequence of west steps, but will not affect the dominant singular behaviour of the model.) We refer to walks which both start and end on the surface as *loops*. In this case, we can perform a similar study to that of $W_{\square}(t; 1, 1; y)$, and we find the following result. (Again, we are unable to complete a rigorous proof.)

Conjecture 4.23. *For a given $y \geq 0$, the dominant singularity of $W_{\square}(t; 1, 0; y)$ is located at*

$$t_c(y) = \begin{cases} \tau \approx 0.412095 & y \leq \sigma \approx 1.82476 \\ h(y) & y > \sigma, \end{cases}$$

where τ is a root of

$$1 - 3\tau - \tau^2 + 6\tau^3 - 7\tau^7 - \tau^8 + 3\tau^9 + \tau^{10} = 0,$$

σ is a root of

$$1 - 7\sigma + 45\sigma^2 - 143\sigma^3 + 277\sigma^4 - 346\sigma^5 + 285\sigma^6 - 155\sigma^7 + 54\sigma^8 - 11\sigma^9 + \sigma^{10} = 0,$$

and $h(y)$ is as defined in Conjecture 4.22. The free energy of 2-sided prudent loops is thus given by $\varkappa(y) = -\log t_c(y)$, and the adsorption transition occurs at $y = y_c = \sigma$ and is first-order, and the crossover exponent is $\phi = 1$. In the limit of infinitely long polymers, the mean density of monomers

---

[10]To be precise, the constant c for partially directed walks is

$$\frac{3 + 2\sqrt{2}}{2}\sqrt{\pi(5\sqrt{2} - 7)} = 1.377\ldots.$$
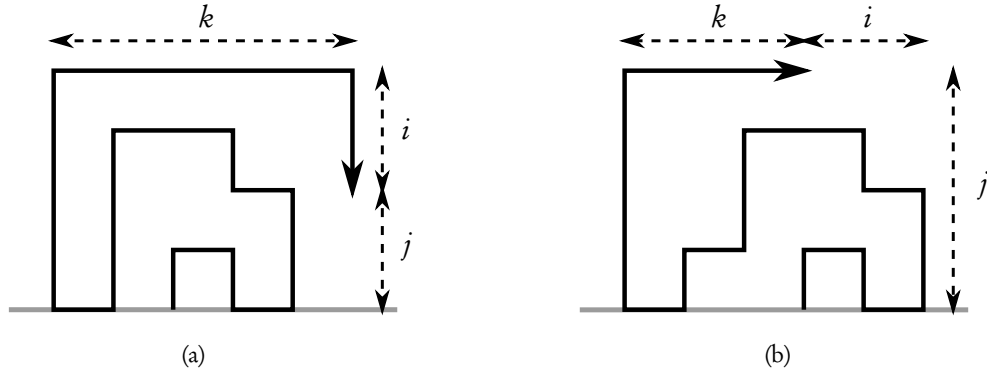
Figure 4.17: 3-sided prudent walks above an impenetrable surface, (a) ending on the right of the box and (b) ending on the top, with the distances i, j and k indicated.

in the surface is

$$\delta(y) = \begin{cases} 0 & y < \sigma \\ \dfrac{-y\,b'(y)}{b(y)} & y > \sigma. \end{cases}$$

(See Figure 4.16 for plots of the dominant singularity and surface density.) The dominant singularity in the adsorbed phase is the same as for $W_\square(t;1,1;y)$, and arises from the same terms. The dominant singularity in the desorbed phase appears to be a root of $C(t;\bar\Lambda^2,\bar\Lambda) = 0$, which would make it a pole of $\mathscr{H}(t;1;y)$ and $\mathscr{I}(t;1;y)$.

So we see that even forcing 2-sided prudent walks to end on the surface does not induce a second-order phase transition.

As we will see in Subsection 4.3.3, the picture is complicated even further when we start looking at prudent walks on the triangular lattice. For that model, it *does* make a difference whether the endpoint of a walk is forced to lie on the surface or not.

3-sided prudent walks

We present here the functional equations for 3-sided prudent walks on the square lattice, though we are unable to solve them. Define the generating functions

$$R^*(t;u,v,w;y) = \sum_{n,i,j,k,\nu} r^*_{n,i,j,k,\nu}\, t^n u^i v^j w^k y^\nu$$

$$T^*(t;u,v,w;y) = \sum_{n,i,j,k,\nu} t^*_{n,i,j,k,\nu}\, t^n u^i v^j w^k y^\nu$$

where $r^*_{n,i,j,k,\nu}$ counts n-step walks ending on the right of their box and $t^*_{n,i,j,k,\nu}$ counts those ending at the top of their box. In both cases i is the distance from the endpoint to the north-east corner of the box, j is the distance from the endpoint to the surface, k is the distance from the endpoint to the west side of the box, and ν is the number of occupied edges along the surface. See Figure 4.17 for an illustration of these two types of walks, and the distances measured by the catalytic variables.

Lemma 4.24. *The generating functions $R^*(t; u, v, w; y)$ and $T^*(t; u, v, w; y)$ satisfy the equations*

$$L^*(t; u, v, w) T^*(t; u, v, w; y) = 1 + t w R^*(t; t v, v, w; y)$$
$$+ t w (y - 1) R^*(t; t v, 0, w; y) + t u R^*(t; t v, v, u; y) + t u (y - 1) R^*(t; t v, 0, u; y)$$
$$- \frac{t^2 v w}{u - t w} T^*(t; t w, v, w; y) - \frac{t^2 u v}{w - t u} T^*(t; u, v, t u; y) \quad (4.44)$$

$$L^*(t; u, w, v) R^*(t; u, v, w; y) = 1 - \frac{t^2 v w}{u - t v} R^*(t; t v, v, w; y) - \frac{t^2 u w}{v - t u} R^*(t; u, t u, w; y)$$
$$+ \frac{t u w (y - 1)}{u - t v} R^*(t; u, 0, w; y) - \frac{t^2 v w (y - 1)}{u - t v} R^*(t; t v, 0, w; y) + t v T^*(t; t w, v, w; y) \quad (4.45)$$

where

$$L^*(t; u, v, w) = 1 - \frac{t u v w (1 - t^2)}{(u - t w)(w - t u)}.$$

We omit the proof, as the details follow in much the same way as in Lemma 4.18. Here, inflating steps can be north, west or east; inflating steps to the north are attached to walks counted by T^*, and inflating steps to the west and east are attached to walks counted by R^* (symmetry means R^* also counts walks ending on the *west* side of the bounding box).

Though the equations are somewhat more complicated, the problem we face here is similar to the difficulty of solving general prudent walks without a boundary [17] – the presence of three catalytic variables seems to render the kernel method useless.

4.3.3 Triangular lattice prudent walks

We now turn our attention to prudent walks on the triangular lattice. (Recall the definitions from Section 3.4.) As in the previous subsections, we place the impenetrable surface in the x-axis, and consider walks which start on the surface. Walks will accumulate a weight y with each edge along the surface. We will restrict our investigation to just two solvable subclasses of prudent walks: *1-sided* prudent walks, which are comprised of only north-west, north-east, east and south-east steps, and *equilateral* prudent walks, which must always end on the boundary of their bounding triangle. (Recall that the model of equilateral walks, without a boundary, was the one solved by Bousquet-Mélou [17].) While the latter model is much harder to solve, it turns out that these two models have essentially the same singular behaviour.

1-sided prudent walks

This model is very similar to that of partially directed walks on the square lattice, and might be considered the analogue of that model on the triangular lattice. We define the generating function

$$P(t; v; y) = \sum_{n, i, \nu} p_{n, i, \nu} t^n v^i y^\nu,$$
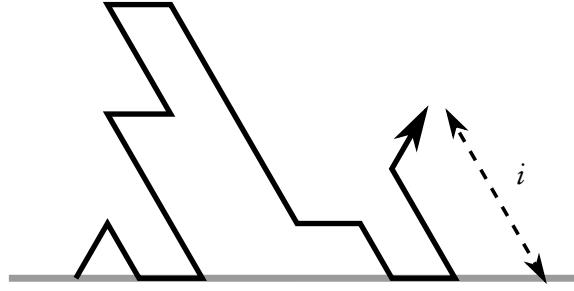
Figure 4.18: A 1-sided triangular prudent walk with the distance i indicated.

where $p_{n,i,v}$ is the number of n-step 1-sided prudent walks which end a distance i above the surface and contain v steps along the surface. (See Figure 4.18.)

Lemma 4.25. *The generating function $P(t;v;y)$ satisfies the functional equation*

$$\left(1 - \frac{tv(1-t^2)(1+v)}{(v-t)(1-tv)}\right) P(t;v;y) = \frac{1}{1-tv} - \frac{t^2(1+t)}{v-t} P(t;t;y) - \frac{t(1-y)}{1-tv} P(t;0;y). \quad (4.46)$$

Proof. As usual, we count walks by conditioning on the last inflating step, which for this model could be north-east or east.

- A walk with no inflating steps must be empty or contain only north-west steps. The generating function of such walks is

$$\frac{1}{1-tv}.$$

- Walks whose last inflating step was north-east can be broken into three parts: the section which came before the inflating step (which could be any walk), the inflating step itself, and then a sequence of north-west or south-east steps. Note that the inflating step increases the distance from the endpoint to the surface by one. The generating function for these walks is

$$\sum_{n,i,v} p_{n,i,v} t^n y^v \cdot t \left(\sum_{l=0}^{\infty} t^l v^{i+l+1} + \sum_{m=1}^{i+1} t^m v^{i-m+1} \right)$$

(The sum over l is for walks with north-west steps following the inflating step, and the sum over m is for walks with south-east steps following the inflating step.)

$$= t \sum_{n,i,v} p_{n,i,v} t^n y^v \left(\frac{v^{i+1}}{1-tv} + \frac{t(v^{i+1} - t^{i+1})}{v-t} \right)$$

$$= \frac{tv}{1-tv} P(t;v;y) + \frac{t^2}{v-t} (vP(t;v;y) - tP(t;t;y)).$$

185

- Walks whose last inflating step was east can be broken into three parts in the same way as the above case, with two differences: here the height above the surface does not change with the inflating step, and if the inflating step was along the surface then we need to account for the walk picking up an extra weight of y. We at first ignore the second point, and obtain the generating function

$$\sum_{n,i,v} p_{n,i,v} t^n y^v \cdot t \left(\sum_{l=0}^{\infty} t^l v^{i+l} + \sum_{m=1}^{i} t^m v^{i-m} \right)$$

$$= t \sum_{n,i,v} p_{n,i,v} t^n y^v \left(\frac{v^i}{1-tv} + \frac{t(v^i - t^i)}{v-t} \right)$$

$$= \frac{t}{1-tv} P(t;v;y) + \frac{t^2}{v-t}(P(t;v;y) - P(t;t;y)).$$

We then subtract off the contribution of those walks whose inflating step was along the surface, and add them back in with an extra factor of y:

$$-\frac{t}{1-tv} P(t;0;y) + \frac{ty}{1-tv} P(t;0;y).$$

The contributions of all the above must add to give precisely $P(t;v;y)$; rearranging the resulting equation gives (4.46). ∎

A simple application of the kernel method suffices to derive the solution to P.

Theorem 4.26. *The generating functions $P(t;1;y)$ and $P(t;0;y)$, of 1-sided prudent walks and prudent loops respectively, are given by*

$$P(t;1;y) = \frac{(1+t)(1-F)}{(1-3t-2t^2)(1+F-tF-t^2F-Fy+t^2Fy)} \tag{4.47}$$

$$P(t;0;y) = \frac{F(1-t^2)}{t(1+F-tF-t^2F-Fy+t^2Fy)}, \tag{4.48}$$

where

$$F \equiv F(t) = \frac{1-t+t^2+t^3-(1+t)\sqrt{1-4t+2t^2+t^4}}{2t(2-t^2)}.$$

Proof. Setting $v = 0$ in (4.46) gives

$$(1+t-ty)P(t;0;y) = 1 + t(1+t)P(t;t;y), \tag{4.49}$$

while setting $v = F$ cancels the kernel of (4.46) and gives

$$P(t;t;y) = \frac{F-t}{t^2(1+t)(1-tF)}(1 - t(1-y)P(t;0;y)). \tag{4.50}$$

Solving (4.49) and (4.50) simultaneously yields $P(t;0;y)$ as stated in the theorem, and

$$P(t;t;y) = \frac{F-t}{t^2(1+t)(1+F-tF-t^2F-Fy+t^2Fy)}.$$

Substitution into (4.46) finally gives $P(t;1;y)$. ∎

It is straightforward to determine the dominant singular behaviour of $P(t; 1; y)$ and $P(t; 0; y)$ for all y, and from there to obtain the free energy and surface density. The y-dependent factor is the same for the two generating functions.

Corollary 4.27. *The dominant singularity for 1-sided prudent walks, for $y \geq 0$, is located at*

$$t_c = \begin{cases} \rho = \frac{\sqrt{17}-3}{4} \approx 0.281 & y \leq \chi = \frac{7+\sqrt{17}}{4} \approx 2.78 \\ \frac{y^2 - \sqrt{y(-4+8y-4y^2+y^3)}}{2y(y-1)} & y > \chi. \end{cases}$$

The singularity is a simple pole for all $y \geq 0$ except at $y = \chi$, where it is a double pole. The free energy for 1-sided prudent walks is thus $x(y) = -\log t_c(y)$, and the adsorption transition occurs at $y = \chi$ and is first-order. The crossover exponent is $\phi = 1$. In the limit of infinitely long polymers, the density of steps in the surface is

$$\delta(y) = \begin{cases} 0 & y < \chi \\ \frac{4-12y+12y^2-5y^3+y^4+y(y-3)\sqrt{y(-4+8y-4y^2+y^3)}}{2(y-1)(-4+8y-4y^2+y^3)} & y > \chi. \end{cases}$$

The dominant singularity for 1-sided prudent loops, for $y \geq 0$, is located at

$$t_c(y) = \begin{cases} \beta \approx 0.296 & y \leq \gamma \approx 2.191 \\ \frac{y^2 - \sqrt{y(-4+8y-4y^2+y^3)}}{2y(y-1)} & y > \gamma, \end{cases}$$

where β is a root of $1 - 3\beta - \beta^2 - \beta^3 = 0$ and γ is a root of $1 - 4\gamma + 6\gamma^2 - 2\gamma^3 = 0$. The singularity is a square root singularity for $y < \gamma$, a pole of order $1/2$ at $y = \gamma$, and a simple pole for $y > \gamma$. The free energy for 1-sided prudent loops is thus $x(y) = -\log t_c(y)$, and the adsorption transition occurs at $y = \gamma$ and is second-order. The crossover exponent is $\phi = 1/2$. In the limit of infinitely long polymers, the density of monomers in the surface is

$$\delta(y) = \begin{cases} 0 & y \leq \gamma \\ \frac{4-12y+12y^2-5y^3+y^4+y(y-3)\sqrt{y(-4+8y-4y^2+y^3)}}{2(y-1)(-4+8y-4y^2+y^3)} & y > \gamma. \end{cases}$$

So we see that for this model, unlike the case of 2-sided prudent walks on the square lattice, the phase transition *is* second-order when we force walks to end on the surface. Even though the results on the square lattice seemed to suggest that our intuition regarding the speed with which walks move away from the surface was incorrect, we will take this opportunity to further complete the picture on the triangular lattice.

We also point out that the dominant singularity in the desorbed phase, $(\sqrt{17} - 3)/4$, is precisely the same as for 1-sided prudent walks on the triangular lattice *without* a boundary (see Lemma 3.53).

Lemma 4.28. *For 1-sided prudent walks on the triangular lattice in the desorbed phase, the mean distance $\langle S_n \rangle$ from the endpoint of a walk of length n to the surface satisfies*

$$\langle S_n \rangle \sim \left(\frac{17 + \sqrt{17}}{68} \right) n.$$

Proof. The probability generating function for the end-to-surface distance of walks of length n is

$$p_n(v) = \frac{[t^n] P(t; v; 1)}{[t^n] P(t; 1; 1)},$$

thus the mean is given by

$$\langle S_n \rangle = \left. \frac{[t^n] \frac{\partial}{\partial v} P(t; v; 1)}{[t^n] P(t; v; 1)} \right|_{v=1}.$$

Calculating $P(t; v; 1)$, and hence $\frac{\partial}{\partial v} P(t; v; 1)$, from (4.46) is straightforward, as is extracting the exact asymptotic form of the coefficient of t^n in each. After simplification the ratio is given by the result of the lemma. ∎

Equilateral prudent walks

The model of 1-sided prudent walks considered above is pleasing, as we saw that the nature of the adsorption transition can be changed simply by forcing walks to end on the surface. It is, however, still a directed model – there are two directions on the lattice in which 1-sided walks cannot step. We thus now turn our attention to a solvable model which is not characterised by a directedness restriction.

As mentioned in Section 3.4, the model of equilateral prudent walks was introduced by Bousquet-Mélou [17]. (The 'equilateral' name is ours.) These are prudent walks on the triangular lattice with the extra condition that the endpoint of a walk must lie on the boundary of the smallest (north-pointing) lattice triangle which contains the entire walk. Bousquet-Mélou solved the generating function of these walks without a surface, and found it to be non-D-finite. The iterative construction used here is modelled on the one used by her, though it is somewhat complicated by the presence of the surface.

We define the generating function

$$E(t; u, v; y) = \sum_{n, i, j, v} e_{n, i, j, v} t^n u^i v^j y^v$$

where $e_{n, i, j, v}$ is the number of n-step equilateral prudent walks starting on a horizontal impenetrable surface and ending on the east side of the bounding triangle, with a distance i from the endpoint to the top of the triangle, a distance j from the endpoint to the surface, and v steps along the surface. (See Figure 4.19 for an illustration of an equilateral prudent walk, together with the distances measured by the catalytic variables.)

Figure 4.19: An equilateral prudent walk above an impenetrable surface, with the distances i and j indicated.

Lemma 4.29. *The generating function $E(t; u, v; y)$ satisfies the functional relation*

$$K(t; u, v)E(t; u, v; y) = 1 - \frac{t^2 u^2 (1 + t)}{v - tu} E(t; u, tu; y) + \frac{tv(1 + t)(u - 2tv)}{u - tv} E(t; tv, v; y)$$
$$- \frac{tu^2(1 - y)}{u - tv} E(t; u, 0; y) - \frac{t^2 v(u - 2tv)(1 - y)}{u - tv} E(t; tv, 0; y) \quad (4.51)$$

where

$$K(t; u, v) = 1 - \frac{tuv(1 - t^2)(u + v)}{(u - tv)(v - tu)}.$$

Proof. Walks are constructed by considering the location and direction of their last inflating step, which moved either the left or right boundaries of the bounding triangle.

- The only walk with no inflating steps is the empty walk, which has generating function 1.

- A walk whose last inflating step was east can be split into three parts: the section which came before the inflating step (any walk ending on the east side of its triangle), the inflating step itself, and then a sequence of north-west or south-east steps. The generating function of these walks is thus

$$\sum e_{n,i,j,\nu} t^n y^\nu \cdot t \left(\sum_{l=1}^{j} t^l u^{i+l+1} v^{j-l} + \sum_{m=0}^{i+1} t^m u^{i-m+1} v^{j+m} \right)$$

(The sum over l is for walks with south-east steps following the inflating step, and the sum over m is for walks with north-west steps following the inflating step. See Figure 4.20 for an illustration.)

$$= t \sum e_{n,i,j,\nu} t^n y^\nu \left(\frac{tu^{i+2}(v^j - (tu)^j)}{v - tu} + \frac{v^j(u^{i+2} - (tv)^{i+2})}{u - tv} \right)$$
$$= \frac{t^2 u^2}{v - tu}(E(t; u, v; y) - E(t; u, tu; y)) + \frac{t}{u - tv}(u^2 E(t; u, v; y) - t^2 v^2 E(t; tv, v; y)).$$

However, if the east step was along the surface, the walk should have picked up an extra factor of y. We correct this by subtracting those terms and then adding them back with
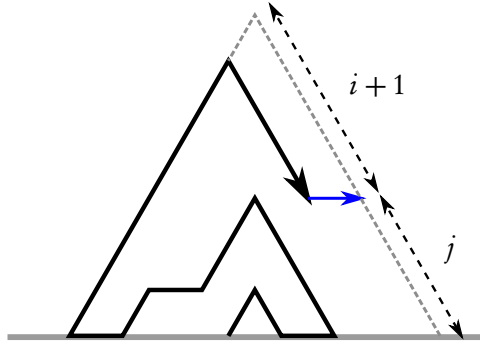
189

Figure 4.20: After an east inflating step, the walk can take up to $i+1$ steps north-west or up to j steps south-east without inflating again.

the extra factor:
$$-\frac{t(1-y)}{u-tv}(u^2 E(t;u,0;y) - t^2 v^2 E(t;tv,0;y)).$$

- If the last inflating step was north-east we have essentially the same considerations as the previous case, without the issue of a step along the surface. The generating function of these walks is
$$\frac{t^2 u}{v-tu}(vE(t;u,v;y) - tuE(t;u,tu;y)) + \frac{tv}{u-tv}(uE(t;u,v;y) - tvE(t;tv,v;y)).$$

- A walk whose last inflating step was north-west must have followed the inflating step with enough north-east steps to reach the top corner of the triangle. Since E also counts walks ending on the left of their triangle (by reflective symmetry), the generating function of these walks is
$$tvE(t;tv,v;y).$$

- Walks with last inflating step west are similar to the north-west case, except we must consider separately the case when the inflating step was along the surface:
$$t^2 vE(t;tv,v;y) - t^2 v(1-y)E(t;tv,0;y).$$

Since the five contributions above are mutually disjoint and cover all possible inflating steps, adding them together gives $E(t;u,v;y)$, and (4.51) follows. $\blacksquare$

Despite the complexity of (4.51), some careful manipulation eventually leads to a solution.

Corollary 4.30. *The generating function $E(t;u,0;y)$ of walks which end on the surface satisfies the functional relation*

$$E(t;u,0;y) = 1 + tu(1+t)E(t;u,tu;y) - tu(1-y)E(t;u,0;y) \qquad (4.52)$$

Proof. This result follows immediately from setting $v=0$ in (4.51). Equivalently, the construction used in the proof of Lemma 4.29 can be modified so as to only generate walks ending on the surface. $\blacksquare$

Equation (4.52) can be used to eliminate one of $E(t;u,0;y)$ and $E(t;u,tu;y)$ from (4.51), giving

$$K(t;u,v)E(t;u,v;y)=X(t;u,v;y)+Y(t;u,v;y)E(t;u,tu;y)+Z(t;u,v)Q(t;v;y), \quad (4.53)$$

where

$$Q(t;v;y)=(1+t)E(t;tv,v;y)-t(1-y)E(t;tv,0;y),$$

$$X(t;u,v;y)=1-\frac{tu^2(1-y)}{(u-tv)(1+tu-tuy)},$$

$$Y(t;u,v;y)=-\left(\frac{t^2u^2(1+t)}{v-tu}+\frac{t^2u^3(1+t)(1-y)}{(u-tv)(1+tu-tuy)}\right),$$

$$Z(t;u,v)=\frac{tv(u-2tv)}{u-tv}.$$

Proposition 4.31. *The form of $Q(t;v;y)$ is given by*

$$Q(t;v;y)=\sum_{n=0}^{\infty}\mathscr{F}(t;t^{2n}S(t;v);y)\prod_{k=0}^{n-1}\mathscr{G}(t;t^{2k}S(t;v);y) \qquad (4.54)$$

when considered as a formal power series in t,v,y, where

$$\mathscr{F}(t;z;y)=\frac{(1+ty)(1+tz)(1+t^2z)(1+t^3z)(1-t^5z^2)}{(1+t^4z)(1-tz(1-2t^2))(1+t^2z(1-y(1-t)))}$$

$$\mathscr{G}(t;z;y)=-\frac{t^6z^2(1+tz)(1-y(1-t)+t^3z)(1-2t^2-t^4z)}{(1+t^4z)^2(1-tz(1-2t^2))(1+t^2z(1-y(1-t)))}$$

$$S(t;v)=\frac{1-t-tv-t^2v-\sqrt{(1-t-tv-t^2v)^2-4t^3v^2}}{2t^3v}.$$

Proof. Equation (4.53) is susceptible to the iterative kernel method [116]. The root $D(t;v)$ which gives $K(t;D(t;v),v)=0$ (since K is quadratic in u and v, there are two such roots, but D is the only one which is a power series in t) is

$$D(t;v)=v\cdot\frac{1-tv+t^2+t^3v-(1+t)\sqrt{(1-t)(1-t-2tv-2t^2v+t^2v^2-t^3v^2)}}{2t(1+v-t^2v)}.$$

Since $K(t;u,v)=K(t;v,u)$, we also have $K(t;u,D(u))=0$.

A naive application of the iterative kernel method to (4.53) would lead to an infinite sequence of equations, obtained by substituting

$$(u,v)\mapsto(D(v),v),(D(v),D(D(v))),(D(D(D(v))),D(D(v))),\dots$$

Without considerable simplification, any solution obtained in this way would be unwieldy and would make the extraction of series coefficients, singularities, etc. very difficult.[11]

---

[11]This is the method we had to use to solve the model of 2-sided perimeter walks (see Lemma 3.31). Recall that for that model we were unable to obtain a proof of the dominant singularity, or even determine its precise value.

Instead, note that the kernel K is the same as that which features in [17] as part of the functional equation for equilateral prudent walks. For that problem, Bousquet-Mélou finds a *rational parametrisation* for the curve $K(t; u, v) = 0$, namely

$$K(t; U(t; z), U(t; tz)) = 0 \quad \text{for} \quad U(t; z) = \frac{z(1 - t)}{(1 + tz)(1 + t^2 z)}$$

for any $z \neq -t^{-1}, -t^{-2}$.

In the non-boundary problem, the two terms on the RHS of the functional equation are $F(u)$ and $F(v)$ for a certain generating function F, and this symmetry allows Bousquet-Mélou to obtain a solution by substituting $(u, v) = (U(t; z), U(t; tz))$ and then iterating $z \mapsto tz$. This method will not work here because $E(t; u, tu; y) \neq Q(t; u; y)$.

Instead, we can exploit the symmetry of $K(t; u, v)$ and substitute the pairs

$$(u, v) = (U(tz), U(z)) \quad \text{and} \quad (u, v) = (U(tz), U(t^2 z)),$$

both of which cancel K. Doing so and eliminating the common term $E(t; U(tz), t U(tz); y)$ gives the equation

$$Q(t; U(z); y) = \mathscr{F}(t; z; y) + \mathscr{G}(t; z; y) Q(t; U(t^2 z); y), \tag{4.55}$$

where

$$\mathscr{F}(t; z; y) = -\frac{X(t; U(tz), U(z); y)}{Z(t; U(tz), U(z))} + \frac{Y(t; U(tz), U(z); y)}{Z(t; U(tz), U(z))} \cdot \frac{X(t; U(tz), U(t^2 z); y)}{Y(t; U(tz), U(t^2 z); y)}$$

$$\mathscr{G}(t; z; y) = \frac{Y(t; U(tz), U(z); y)}{Z(t; U(tz), U(z))} \cdot \frac{Z(t; U(tz), U(t^2 z))}{Y(t; U(tz), U(t^2 z); y)}.$$

After simplification $\mathscr{F}$ and $\mathscr{G}$ can be written as given in the proposition.

Iterating (4.55) with $z \mapsto t^2 z$ yields

$$Q(t; U(z); y) = \sum_{n=0}^{\infty} \mathscr{F}(t; t^{2n} z; y) \prod_{k=0}^{n-1} \mathscr{G}(t; t^{2k} z; y). \tag{4.56}$$

Since $\mathscr{F}(t; z; y)$ is a power series in t, with coefficients in $\mathbb{Z}[y, z]$, of the form $1 + t(y + 2z) + O(t^2)$, and likewise $\mathscr{G}(t; z; y)$ is a power series of the form $t^6 z^2 (y - 1) + O(t^7)$, the sum in (4.56) converges as a formal power series in t with coefficients in $\mathbb{Z}[y, z]$.

Finally, since $S(t; v)$ is the function satisfying $U(t; S(t; v)) = v$ and $S(t; v)$ is a power series in t with coefficients in $\mathbb{Z}[v]$ of the form $v + t(v + v^2) + O(t^2)$, we obtain (4.54). ∎

Once we have a solution to $Q(t; v; y)$, the final solution for $E(t; u, v; y)$ is easily obtained.

Theorem 4.32. *The generating function $E(t; u, v; y)$ for equilateral prudent walks above an impenetrable surface, which end on the right side of their bounding triangle and accrue a weight y with*

each step along the surface, is given by

$$E(t;u,v;y) = \frac{1}{K(t;u,v)}\left(X(t;u,v;y) - \frac{Y(t;u,v;y)}{Y(t;u,D(t;u);y)}(X(t;u,D(t;u);y)\right.$$

$$\left. + Z(t;u,D(t;u))Q(t;D(t;u);y)) + Z(t;u,v)Q(t;v;y)\right). \quad (4.57)$$

The overall generating function for equilateral prudent walks above an impenetrable surface is then

$$W_\triangle(t;u,v;y) = \sum_{n,i,j,\nu} w^\triangle_{n,i,j,\nu} t^n u^i v^j y^\nu$$

$$= 2E(t;u,v;y) - E(t;0,v;y)$$

where $w^\triangle_{n,i,j,\nu}$ is the number of n-step triangular prudent walks above an impenetrable surface which end a distance i from the top of their bounding triangle, a distance j from the surface and have ν steps along the surface.

Proof. To obtain $E(t;u,tu;y)$, substitute $v = D(t;u)$ into (4.53), cancelling the kernel, and rearrange to get

$$E(t;u,tu;y) = \frac{-1}{Y(t;u,D(t;u);y)}(X(t;u,D(t;u);y) + Z(t;u,D(t;u))Q(t;D(t;u);y)). \quad (4.58)$$

Substituting (4.58) into (4.53) then yields the given form of $E(t;u,v;y)$. The overall generating function is obtained by adding walks ending on the left and right of the bounding triangle, both of which are counted by E. However walks ending at the top corner of their triangle have been counted twice and thus we subtract $E(t;0,v;y)$. ∎

Now armed with the generating function $W_\triangle(t;u,v;y)$, we wish to determine its singularity structure. More specifically, we wish to know the location and nature of the dominant singularity $t_\triangle(y)$, and how it behaves as we vary the surface fugacity y from 0 to ∞.

Similarly to our study of 1-sided prudent walks, we are interested in two specialisations of the catalytic variables u and v: $(u,v) = (1,1)$, corresponding to all equilateral walks; and $(u,v) = (1,0)$, corresponding to equilateral loops. Thus, we are now only concerned with $W_\triangle(t;1,1;y)$ and $W_\triangle(t;1,0;y)$. Since $D(t;u)$ will only ever be evaluated at $u = 1$, we will henceforth just write D instead of $D(t;1)$.

Theorem 4.33. *The location of the dominant singularity $t_\triangle(y)$ of $E(t;1,1;y)$ is given by*

$$t_\triangle(y) = \begin{cases} \rho = \frac{\sqrt{17}-3}{4} \approx 0.281, & 0 \leq y \leq \chi = \frac{7+\sqrt{17}}{4} \approx 2.78 \\ \frac{y^2 - \sqrt{y(-4+8y-4y^2+y^3)}}{2y(y-1)}, & y > \chi \end{cases}$$

The dominant singularity is a simple pole for all $y \geq 0$ except at $y = \chi$, where it is a double pole. The free energy is thus $\varkappa_\triangle(y) = -\log t_\triangle(y)$. The derivative of $\varkappa_\triangle(y)$ is discontinuous at $y = \chi$, and thus the adsorption transition is first-order. The crossover exponent is $\phi = 1$.
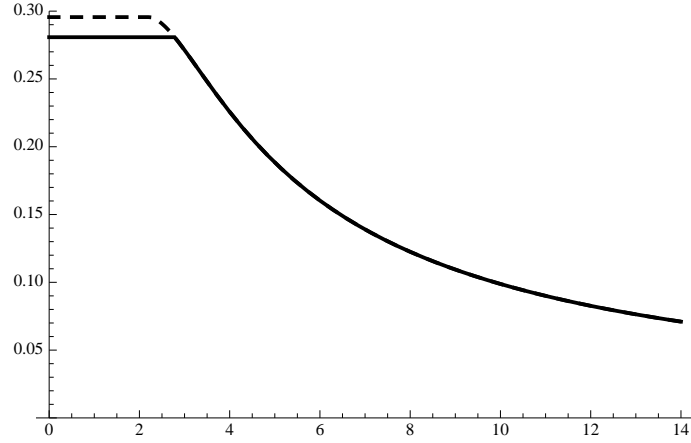
Figure 4.21: A plot of $t_\triangle(y)$ (solid) and $t_Y(y)$ (dashed) for $0 \leq y \leq 15$. Note that $t_Y(y)$ undergoes a smooth (second-order) phase transition while $t_\triangle(y)$ has a first-order transition.

In Figure 4.21 we plot the singularity functions $t_\triangle(y)$ and $t_Y(y)$.

We will prove Theorem 4.33 by showing that, for a given $y \geq 0$, $E(t; 1, 1; y)$ is singular at $t = t_\triangle(y)$ and non-singular for all $|t| < t_\triangle(y)$. As we have found no way to significantly simplify $E(t; u, v; y)$ beyond the formulation given in (4.57), we will consider its constituent parts individually.

We first note that

$$K(t; 1, 1) = 0 \text{ at } t = \frac{-3 \pm \sqrt{17}}{4}$$

and so the dominant singularity of $K(t; 1, 1)^{-1}$ is a simple pole at $t = \rho$.

Lemma 4.34. *There is a square-root singularity in D at $t = \beta = 0.295598...$, a root of $1 - 3\beta - \beta^2 - \beta^3 = 0$. Apart from $t = (y - 1)^{-1}$, the terms $X(t; 1, 1; y)$, $Y(t; 1, 1; y)$, $Z(t; 1, 1)$, $X(t; 1, D; y)$ and $Z(t; 1, D)$ have no singularities smaller in absolute value than β.*

Proof. This is clear by inspection for $X(t; 1, 1; y), Y(t; 1, 1; t)$ and $Z(t; 1, 1)$. There are other square-root singularities in D but all have absolute value at least 1, and also simple poles in D at $t = \pm\sqrt{2}$.

$X(t; 1, v; y)$ and $Z(t; 1, v; y)$ are also singular when $tv = 1$. But D is a power series in t with non-negative integer coefficients (this can easily be shown by considering the recurrence relations which determine the coefficients of D), so for $|t| < \beta$ (β is the radius of convergence of D), $|tD| < tD|_{t=\beta} = 0.21374...$ So $X(t; 1, D; y)$ and $Z(t; 1, D)$ produce no new singularities for $|t| < \beta$. ∎

We are left to consider the two infinite sums in $Q(t; 1; y)$ and $Q(t; D; y)$, as well as $Y(t; 1, D; y)^{-1}$.

Lemma 4.35. *For a given $y \geq 0$, the dominant singularity of $Y(t;1,D;y)^{-1}$ is given by*

$$t_Y(y) = \begin{cases} \beta \approx 0.296, & 0 \leq y \leq \gamma \approx 2.191 \\ \frac{y^2 - \sqrt{y(-4+8y-4y^2+y^3)}}{2y(y-1)}, & y > \gamma \end{cases}$$

where γ is a root of $-1 + 4\gamma - 6\gamma^2 + 2\gamma^3 = 0$. The singularity is a square-root singularity for $y < \gamma$, a pole of order $1/2$ at $y = \gamma$ and a simple pole for $y > \gamma$.

Proof. The y-dependent factor of $Y(t;1,v;y)^{-1}$ is

$$\frac{1 + t(1-y)}{1 + v(1 - t(1 + t(1-y)) - y)}.$$

(The other factors contribute only removable singularities at $t = 0, -1$.) The numerator is clearly non-singular for all y; the denominator is 0 when

$$\frac{v(1-t^2)}{1 + v - tv - t^2 v} = y^{-1}.$$

When $v = D$ this reduces to

$$U \equiv U(T) = \frac{1 - t^2 - \sqrt{(1-t)(1 - 3t - t^2 - t^3)}}{2} = y^{-1} \tag{4.59}$$

Now U is a power series in t with non-negative integer coefficients (as usual, this can be shown by considering the quadratic of which U is a root) and radius of convergence β. So for any value of $y^{-1} \in [0, \gamma^{-1})$ (where γ^{-1} is just $U|_{t=\beta}$), there will be a non-negative real solution t^* to (4.59) with $t^* \in [0, \beta)$. Moreover, any other solution t' must satisfy $|t'| > t^*$. The non-negative real solution t^* is given by $t_Y(y)$ for $y > \gamma$.

For $|t| < \beta$ we have $|U| < U|_{t=\beta} = \gamma^{-1}$. This implies that for $y^{-1} > \gamma^{-1}$, any solution to (4.59) must lie outside $|t| \leq \beta$, and hence cannot be a dominant singularity of $Y(t;1,D;y)^{-1}$.

The nature of the singularity $t_Y(y)$ is clear for $y < \gamma$ and $y > \gamma$. At $y = \gamma$, the square-root singularity and the simple pole coincide, and simple algebraic manipulation shows that $Y(t;1,D;\gamma)^{-1} \sim (t-\beta)^{-1/2}$ as $t \to \beta$. ∎

The following minor result compares the two y-dependent singularities examined thus far.

Corollary 4.36. *For $y \geq 0$, the function $t_Y(y)$ satisfies*

$$t_Y(y) < (y-1)^{-1}$$

Proof. This is clearly true for $y \leq \gamma$. For $y > \gamma$, both $t_Y(y)$ and $(y-1)^{-1}$ are monotone decreasing functions, and thus it is sufficient to compare their inverses, and show

$$t_Y^{-1}(t) = \frac{1}{U} < \frac{1+t}{t} \iff U > \frac{t}{1+t}.$$

But $U - t/(1+t)$ is a power series in t with non-negative coefficients, of the form $t^2 + 3t^4 + O(t^5)$, and is thus positive for $t \in (0, \beta)$. ∎

At this stage we have in fact located the dominant singularity of $E(t;1,1;y)$:

$$t_\triangle(y) = \min\{\rho, t_Y(y)\}.$$

The transition from ρ to $t_Y(y)$ occurs at χ because $t_Y(\chi) = \rho$, and there is a double pole at that point because both the y-dependent and -independent singularities contribute simple poles.

However, it remains to be shown that the infinite sums $Q(t;1;y)$ and $Q(t;D;y)$ are convergent and non-singular for $|t| \leq t_\triangle(y)$.

By considering the quadratic of which $\sigma(t;v)$ is a root, it can be shown that the series coefficients of $\sigma(t;v)$ are non-negative integers. So $\sigma(t;D)$ is a power series in t with non-negative integer coefficients, and since $|D| < 1$ for $|t| < \beta$, we have $|\sigma(t;D)| < |\sigma(t;1)|$ for $|t| < \beta$, with β being the radius of convergence of $\sigma(t;D)$.

Lemma 4.37. $\mathscr{F}(t;\sigma(t;1);y), \mathscr{G}(t;\sigma(t;1);y), \mathscr{F}(t;\sigma(t;D);y)$ and $\mathscr{G}(t;\sigma(t;D);y)$ are non-singular for $|t| \leq t_\triangle(y)$.

Proof. Clearly none of the terms in the numerators of $\mathscr{F}$ and $\mathscr{G}$ contribute non-removable singularities apart from the square root singularity at $t = \beta$. It suffices then to show that the three terms which appear in the denominators,

$$(1 + t^4 z), (1 - tz(1 - 2t^2)) \text{ and } (1 + t^2 z(1 - y(1 - t))),$$

are nonzero for $|t| \leq t_\triangle(y)$. In the following z can be specialised to either $t^{2n}\sigma(t;1)$ or $t^{2n}\sigma(t;D)$.

- $\mathbf{1 + t^4 z}$: For $|t| \leq \rho$,

$$|t^4 z| = |t^4||z| \leq \rho^4 |z| \leq \rho^4 \sigma(t;1)|_{t=\rho} \approx 0.0221.$$

  So

$$|1 + t^4 z| \geq 0.977$$

- $\mathbf{1 - t(1 - 2t^2)z}$: When $z = t^{2n}\sigma(t;1)$, we have that $t^{2n+1}(1 - 2t^2)\sigma(t;1)$ is a power series in t with non-negative integer coefficients. Then for $|t| \leq \rho$,

$$|t^{2n+1}(1 - 2t^2)\sigma(t;1)| \leq |t(1 - 2t^2)\sigma(t;1)| \leq t(1 - 2t^2)\sigma(t;1)|_{t=\rho} \approx 0.842.$$

  So

$$|1 - t^{2n+1}(1 - 2t^2)\sigma(t;1)| \geq 0.157$$

  Since $\sigma(t;D)$ is bounded above by $\sigma(t;1)$, the above precludes the existence of a root when $z = t^{2n}\sigma(t;D)$.

- $\mathbf{1 + t^2 z(1 - y(1 - t))}$: The root occurs when

$$\frac{t^2 z(1 - t)}{1 + t^2 z} = y^{-1}, \tag{4.60}$$

196

so define

$$\tau_n(t;v) := \frac{t^{2n+2}\sigma(t;v)(1-t)}{1+t^{2n+2}\sigma(t;v)}.$$

It is easy to show $\tau_n(t;v)$ is a power series in t with coefficients in $\mathbb{Z}^{\geq 0}[v]$, so for $v = 1$ or D and $y^{-1} \in [0, \tau_n(t;v)|_{t=\beta})$, the root of (4.60) of minimal absolute value will occur on the positive real axis. Also,

$$\frac{\tau_{n+1}(t;v)}{\tau_n(t;v)} = \frac{t^2 + t^{2n+4}\sigma(t;v)}{1+t^{2n+4}\sigma(t;v)} < 1 \quad \text{for } t \in [0,\beta),$$

so the τ_n are decreasing with n for relevant values of t.

Now

$$\tau_0(t;1) = \frac{1 - 2t + t^2 - \sqrt{1 - 4t + 2t^2 + t^4}}{2} = U - t + t^2$$

where U is defined in (4.59). Recall that the y-dependent component of $t_\triangle(y)$ is the solution to $U = y^{-1}$. So if $t = t_\triangle(y)$ and $t > 0$ (i.e. $y < \infty$),

$$\tau_0(t;1) = y^{-1} - t + t^2 < y^{-1},$$

and since $\tau_0(t;1)$ is a non-decreasing function of t on $[0,\beta)$, it follows that there are no roots of $\tau_0(t;1) = y^{-1}$ for $t \leq t_\triangle(y)$. Since $\tau_{n+1}(t;1) \leq \tau_n(t;1)$, this statement extends to all n, and similarly for $v = D$ since $\tau_n(t;D) \leq \tau_n(t;1)$. ∎

Lemma 4.38. *For a given $y \geq 0$, the sum*

$$Q(t;v;y) = Q(t;U(\sigma(t;v));y) = \sum_{n=0}^{\infty} \mathscr{F}(t;t^{2n}\sigma(t;v);y) \prod_{k=0}^{n-1} \mathscr{G}(t;t^{2k}\sigma(t;v);y)$$

converges when $|t| \leq t_\triangle(y)$ for $v = 1$ or D.

Proof. All of the factors of $\mathscr{F}(t;t^{2n}z;y)$ and $\mathscr{G}(t;t^{2n}z;y)$ with no y-dependence (where $z = \sigma(t;1)$ or $\sigma(t;D)$) can easily be bounded (in absolute value) uniformly with respect to n on the disc $|t| \leq \rho$, as can the term $(1 - y(1-t) + t^{2n+3}z)$ in the numerator of $\mathscr{G}$.

The term $1 + ty$ in the numerator of $\mathscr{F}$ has no dependence on n and can thus be trivially bounded for a given y.

The remaining factor of $\mathscr{F}$ and $\mathscr{G}$ is $f_n(t;y) = (1 + t^{2n+2}(1 - y(1-t))z)^{-1}$. For $|t| \leq t_\triangle(y)$, it is clear that

$$|1 + t^{2n+2}(1 - y(1-t))z| \to 1 \quad \text{as } n \to \infty,$$

and so for any $c \in (0,1)$ and $y < \infty$ there exists a positive integer N (depending on c and y) such that

$$|1 + t^{2n+2}(1 - y(1-t))z| > c \quad \text{for all } n \geq N \text{ and } |t| \leq t_\triangle(y).$$

So for a given c, we have $|f_n(t; y)| < c^{-1}$ for all $n \geq N$. For $n < N$, the f_n are all non-singular for $|t| \leq t_\triangle(y)$ by Lemma 4.37, and thus each is bounded in absolute value on this disc by some constant c_n. Then $|f_n(t; y)| < \max\{c^{-1}, c_0, c_1, \ldots, c_{N-1}\}$.

We now have the existence of positive values $c_{\mathscr{F}}(y)$ and $c_{\mathscr{G}}(y)$ such that $|\mathscr{F}(t; t^{2n} z; y)| < c_{\mathscr{F}}(y)$ and $|\mathscr{G}(t; t^{2n} z; y)/t^{4n}| < c_{\mathscr{G}}(y)$ for all n. Then

$$\left| \sum_{n=0}^\infty \mathscr{F}(t; t^{2n} z; y) \prod_{k=0}^{n-1} \mathscr{G}(t; t^{2k} z; y) \right| < \sum_{n=0}^\infty c_{\mathscr{F}}(y) \prod_{k=0}^{n-1} c_{\mathscr{G}}(y) |t|^{4n} = c_{\mathscr{F}}(y) \sum_{n=0}^\infty c_{\mathscr{G}}(y)^n |t|^{2n(n-1)}$$

which converges for $|t| \leq t_\triangle(y)$. ∎

We now have all the necessary ingredients to verify Theorem 4.33.

Proof of Theorem 4.33. By construction, for a given $y \geq 0$, $t_\triangle(y)$ is the location of the singularity of $E(t; 1, 1; y)$ closest to 0. The two components of $t_\triangle(y)$, for $y < \chi$ and $y > \chi$, are lines of simple poles, so their point of intersection at $(t, y) = (\rho, \chi)$ is thus a double pole in $E(t; 1, 1; y)$.

The derivative of $t_\triangle(y)$ is discontinuous at $y = \chi$, since

$$\left. \frac{d}{dy} t_Y(t) \right|_{t=\chi} = \frac{-393 + 95\sqrt{17}}{32} \approx -0.041.$$ ∎

Corollary 4.39. *The dominant singularity of $W_\triangle(t; 1, 1; y)$ for a given $y \geq 0$ is given by $t_\triangle(y)$.*

Proof. Since $E(t; 0, 1; y) = 1 + 2tQ(t; 1; y)$, and the singularities of $Q(t; 1; y)$ already appear in $E(t; 1, 1; y)$, there are no singularities in $W_\triangle(t; 1, 1; y)$ not already present in $E(t; 1, 1; y)$. ∎

This wraps up our analysis of equilateral prudent walks; we are left to consider the singular behaviour of equilateral prudent loops. Fortunately, most of the work is already done.

Lemma 4.40. *The location of the dominant singularity of $W_\triangle(t; 1, 0; y)$, the generating function of equilateral prudent loops, is given by $t_Y(y)$ for $y \geq 0$. This singularity is a square root singularity for $0 \leq y < \gamma$, a pole of order $1/2$ at $y = \gamma$ and a simple pole for $y > \gamma$.*

Since the derivative of $t_Y(y)$ exists and is continuous for $y \geq 0$, triangular prudent loops undergo a second-order phase transition at $y = \gamma$. The crossover exponent is $\phi = 1/2$.

Proof. We have that $W_\triangle(t; 1, 0; y) = 2E(t; 1, 0; y) - 1$, and with (4.52) and (4.58) it follows that

$$\begin{aligned}
E(t; 1, 0; y) &= \frac{1}{1 + t - ty}(1 + t(1 + t)E(t; 1, t; y)) \\
&= \frac{1}{1 + t - ty}\left(1 - \frac{t(1 + t)}{Y(t; 1, D; y)}(X(t; 1, D; y) + Z(t; 1, D)Q(t; D; y))\right).
\end{aligned}$$
(4.61)

The arguments in the proofs of Lemmas 4.37 and 4.38 can easily be modified to show that $\mathscr{F}(t; t^{2n}\sigma(t; D); y)$ and $\mathscr{G}(t; t^{2n}(\sigma(t; D); y)$ contain no poles for $|t| \leq t_Y(y)$, and that $Q(t; D; y)$ converges for $|t| \leq t_Y(y)$. So $Q(t; D; y)$ is non-singular for $|t| \leq t_Y(y)$, except for a square root singularity at $t = \beta$. It then follows by Lemmas 4.34 and 4.35 and Corollary 4.36 that the location of the dominant singularity for triangular prudent loops is given by $t_Y(y)$. ∎

Corollary 4.41. *For $y \geq 0$, the density of monomers in the surface for equilateral prudent loops is given by*

$$\delta_\ell(y) = \begin{cases} 0, & 0 \leq y \leq \gamma \approx 2.191 \\ \dfrac{4-12y+12y^2-5y^3+y^4+y(y-3)\sqrt{y\left(-4+8y-4y^2+y^3\right)}}{2(y-1)\left(-4+8y-4y^2+y^3\right)}, & y > \gamma \end{cases}$$

where γ is a root of $-1+4\gamma-6\gamma^2+2\gamma^3 = 0$, and for triangular prudent walks the density is given by

$$\delta_\triangle(y) = \begin{cases} 0, & 0 \leq y < \chi = \frac{7+\sqrt{17}}{4} \approx 2.78 \\ \dfrac{4-12y+12y^2-5y^3+y^4+y(y-3)\sqrt{y\left(-4+8y-4y^2+y^3\right)}}{2(y-1)\left(-4+8y-4y^2+y^3\right)}, & y > \chi \end{cases}$$

We plot these density functions in Figure 4.22.

These results are both pleasing and somewhat disappointing. On one hand we have explicitly solved a two-dimensional model of polymer adsorption which is not characterised by a directedness restriction – the SAWs which feature in the model are able to step in all six directions on the lattice. To our knowledge this is the first such model to be solved exactly. We were also able to find explicit expressions for the free energy and surface monomer density of this model.

On the other hand, despite the considerable difference in the complexity of their respective generating functions, the free energy of equilateral prudent walks is exactly the same as that of 1-sided prudent walks. This is not entirely surprising – as an equilateral walk takes more and more inflating steps, its triangle grows bigger and bigger. This in turn means that a walk ending on, say, the east side of its triangle must take successively more north-west steps to reach the top corner of its triangle before it can 'turn the corner' and step onto the west side. This effectively means that in the limit of infinitely long walks, the 'average' equilateral prudent walk behaves essentially like a 1-sided prudent walk.

199
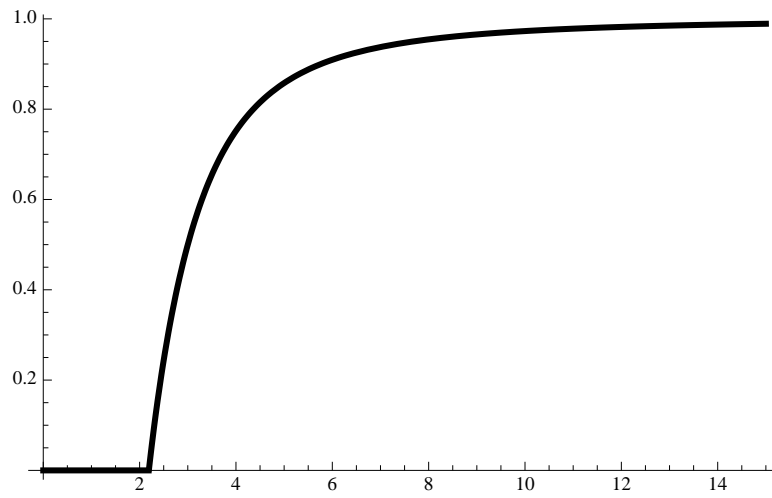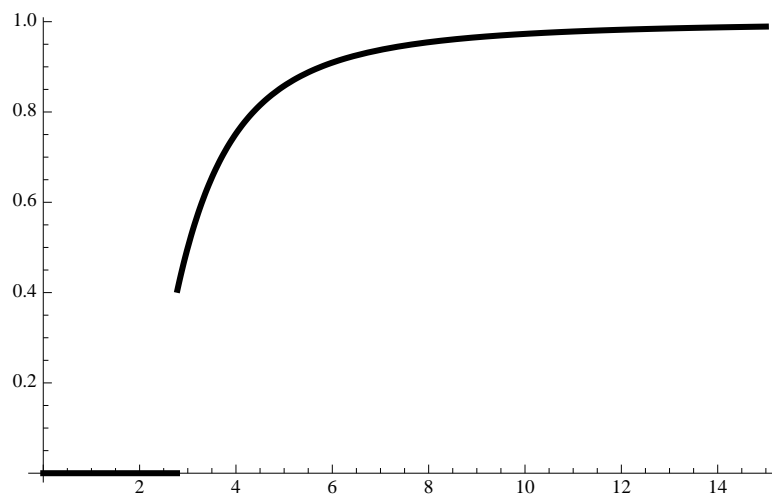
Figure 4.22: Plots of the densities of monomers (edges) in the surface versus the interaction strength y, for (a) equilateral prudent loops and (b) equilateral prudent walks. Note that the first-order adsorption transition for walks results in a jump discontinuity in the density at $y = \chi \approx 2.78$.

Chapter 5

Summary

We have considered a number of aspects of self-avoiding walks and polygons in this thesis. The results of Chapters 2 and 3 dealt almost entirely with enumerative properties of SAWs and SAPs, and there we considered the general models as well as subclasses. In Chapter 4 we turned our focus to the applications of SAWs in the modelling of polymer adsorption.

In Chapter 2 we examined the recent proof by Duminil-Copin and Smirnov [34] of the exact value of the growth constant of SAWs on the honeycomb lattice. While their methods do not seem to provide a way to obtain an analogous result for the square or triangular lattices, we did find that they can be used to compute estimates for those cases. We calculated series for SAWs in strips of small widths (up to 15 on the square lattice and 11 on the triangular), and found values of the step fugacity x which satisfied identities similar to those found by Duminil-Copin and Smirnov for the honeycomb lattice. As expected, when we extrapolated these values we found that the limits matched the current estimates for the critical point (i.e. the reciprocal of the growth constant) for each lattice. The precision of our estimates is only 1–2 digits less than the current benchmarks. (Though, as we noted in Chapter 2, our estimates are biased, as they rely on unproven assumptions regarding the limiting behaviour of certain generating functions.) This new method for computing numerical estimates seems very promising, and in the next chapter we discuss some possibilities for further applications.

In Chapter 3 we turned our attention from general SAWs to subclasses of SAWs and SAPs; specifically, subclasses for which we can find recursive constructions and, hopefully, explicit expressions for generating functions. A large part of the chapter was devoted to the enumeration of 3-sided prudent polygons by area. While the recursive construction and resulting generating function were relatively easy to obtain, we found that the singularity structure of that generating function is surprisingly complicated. A very detailed analysis was thus required in order to calculate the exact asymptotic form of the number of 3-sided prudent polygons. In addition, this asymptotic form has some very unusual properties, including a transcendental-valued critical exponent (where we would normally expect a rational value) and an oscillating function instead of a constant critical amplitude. There remain a number of open questions regarding this model

and its unusual asymptotic behaviour, and we mention some of these in the next chapter.

We also considered a number of other new subclasses of SAWs and SAPs, including perimeter and quasi-prudent walks and polygons. These are generalisations of the analogous prudent objects, and are thus rather more difficult to construct and solve. We found solutions for some classes of each, and as expected observed the growth constants to be larger than those of the corresponding prudent objects. However, the most general models remain unsolved at this time, and so we instead gave numerical estimates of the relevant critical values. Finally, we examined some new models on the honeycomb and triangular lattices. For both lattices we believe that these new models are more numerous (in terms of the size of growth constant) than any other currently-known solved models.

The third main part of the thesis, Chapter 4, dealt with two-dimensional models of polymer adsorption. We first returned to the methodology of Duminil-Copin and Smirnov, and showed that one of their key identities could be generalised to account for a fugacity y associated with vertices in an impenetrable surface. One of the terms in this identity vanishes at $y = 1 + \sqrt{2}$, the value conjectured by Batchelor and Yung [7] to be the critical surface fugacity for this model. We showed that our identity could be used to prove this fact. Our proof depends on the generating function of bridges of height T disappearing in the limit $T \to \infty$, and the proof of this result is given in Appendix A. We also considered an alternative geometry (essentially a 90° rotation of the original) for which there is also a conjecture [6] for the critical surface fugacity, and demonstrated that a similar identity and proof could be derived.

We then applied the same reasoning used in Chapter 2 to models of polymer adsorption on the square and triangular lattices. In particular, we showed that our identity for the honeycomb lattice could be used to calculate estimates of the critical surface fugacities for SAWs on other two-dimensional lattices. We once again computed series for walks strips of small width, but this time the series variable was the surface fugacity y. We then found values of y which satisfied identities like the one for the honeycomb lattice. The extrapolations of these values matched existing estimates for critical fugacities on the square lattice, and their precision exceeded that of the existing estimates by several orders of magnitude. On the triangular lattice we were unable to find any existing estimates, and thus believe our values to be the first of their kind. We do note, however, that our estimates are biased, and thus at this stage there is no guarantee that they are accurate.

Finally, we introduced some new solvable models of polymer adsorption, based on prudent walks. These models differ from existing solvable models in that they are not directed – the walks in question are able to take steps in all directions on their respective lattices. They thus 'interpolate' between the directed walk models and the more general SAW model. We obtained solutions for several cases, including an asymmetric model on the square lattice and a symmetric model on the triangular lattice. We proved the exact form of the free energy for the triangular model, and found the model to display some unexpected critical behaviour. In particular,

we observed a first-order adsorption phase transition in the case when only one end of a walk must be attached to the surface, and a second-order transition when both ends are attached. We attempted to provide an intuitive explanation for this behaviour, but matters are further complicated by the square lattice model, which does not exhibit the same phase transition behaviour.

Chapter 6

Future work

In this chapter we discuss possible directions for future investigation related to this thesis. As such we will divide it into three sections, broadly corresponding to Chapters 2, 3 and 4.

General SAWs and SAPs

- In light of the developments of Section 2.2, the most obvious next step is to investigate how our methodology can be improved so as to give more accurate estimates. Further advances in algorithms for enumerating SAWs in a strip (see Section B.1), or simply more powerful computers, should enable us to generate series for strips of width greater than those used in Chapter 2. Given how close our estimates are, in terms of precision, to the best current estimates for the critical points on the square and honeycomb lattices, it seems reasonable to suppose that only a small number of extra widths (perhaps two or three) would be needed before we could actually improve on the precision of the best known estimates.

- We have so far only applied this technique to the regular two-dimensional lattices. The generalisation to non-regular two-dimensional lattices seems straightforward – all that should be required is to generate series for the A_T and B_T generating functions in strips for small T and then apply the same methods used in Chapter 2.

  There is also the possibility that our techniques could be used to compute estimates for higher-dimensional lattices, though the way forward there is rather less obvious. To begin with, we have no reason to believe there is any three-dimensional lattice for which there exists a parafermionic observable (or some higher-dimensional equivalent) satisfying an identity like (2.3). If we could find one, this might suggest what form a "domain identity" (like (2.10)) should take. Without such a guide, it is unclear even what geometry we should restrict walks to – possibilities include *slabs* and *tubes*.

- We also mention here some recent work by Elvey Price et al. [36] which explores a generalisation of the identity (2.3) to values of x *away from the critical value x_c*. The authors use

that result to prove an inequality relating two critical exponents of walks in a half-plane, and to provide further evidence in support of a conjecture [35] regarding the probability distribution of the winding angles of SAWs in a half-plane. It seems possible that their methods (currently restricted to the honeycomb lattice) could be applied to other two-dimensional lattices in a similar manner to our adaptation of the results of Duminil-Copin and Smirnov.

Solvable subclasses

- There are a number of questions which arise from our results regarding 3-sided prudent polygons. The most obvious (and, perhaps, the most difficult to answer) is why oscillations are present in the asymptotic form of $pa_n^{(3)}$, when such behaviour has never been observed in any other polygon model. Can we find a *combinatorial* argument to explain the complexity of the singular structure of the generating function?

  Our results also raise the question as to what other polygon or walk models display oscillatory (or perhaps even more complex) asymptotic behaviour. Certainly there are no rigorous results which preclude such behaviour being present in the general models of SAWs or SAPs (in low dimensions, anyway), and the series we currently have at hand are likely far too short to enable detection of such subtle variations. One is even forced to wonder if there are any *solved* models in the literature where oscillatory behaviour has been overlooked, with the authors simply taking for granted that the usual asymptotic form $cn^g\mu^n$ is applicable.

  It may also be possible to solve the area-perimeter generating function of 3-sided prudent polygons. If so, it would be interesting to study the scaling behaviour of said function around its critical point, and to see if it is somehow set apart from other polygon models. (See, for example, [101, 102].)

- We observe that there are ways of "combining" some of the models discussed in Chapter 3, and we will mention two here. The first, a combination of prudent and perimeter walks, would involve k-sided walks (recall that the definition of k-sided is the same for prudent and perimeter walks) which must take prudent steps on one or more of their k sides, but not on the others. For example, a 3-sided walk (which must end on the north, east or west sides of its box) might be required to take prudent steps on the east and west sides of its box, but not on the north side. Such a model would likely have a growth rate strictly larger than that of 3-sided prudent walks, but may not be much more complicated to solve.

  Another possibility involves a generalisation of the "weakly directed walks" of Bacher and Bousquet-Mélou [2]. Recall that these are walks which, between any two visits to a horizontal line, must be partially directed. We could instead define "weakly prudent walks", where between any two visits to a horizontal line, a walk must be *prudent* (or perhaps

2-sided prudent, etc.). Given that weakly directed walks have a growth rate considerably larger than that of the next-best solvable walk model (2-sided perimeter walks), it seems likely that, if such a model is solvable, it would have an even greater growth rate.

- Our study of solvable models on the triangular and honeycomb lattices was restricted to prudent walks. There is an obvious way to define prudent *polygons* on those lattices, and these should be solvable whenever the corresponding walk model is. It would be particularly interesting to discover if the asymptotic enumeration of prudent polygons by area on these lattices displays any similarities to the square lattice case.

  We can likewise define perimeter and quasi-prudent walks and polygons on the triangular and honeycomb lattices, and these will almost certainly be more numerous (in terms of the growth constants) than the prudent objects we have discussed in this thesis.

- There were several models discussed in Chapter 3 for which we were able to derive functional equations that we were unable to solve. In most cases the problem was that the generating functions had too many catalytic variables. We have techniques, like the kernel method and its generalisations, for solving functional equations involving up to two catalytic variables, but we have been unable to apply these techniques to equations with three or more catalytic variables. We would very much like to have a method for solving these equations, or at least have a definitive answer one way or the other as to whether they *can* be solved.

- As with Chapter 2, we have restricted our investigations to models on the regular two-dimensional lattices. Possibilities abound for solvable models on non-regular lattices.

  Far more tantalising (though also more difficult) are possibilities for solvable models in three dimensions. There exist some numerical results for prudent and perimeter walks on the simple cubic lattice [9], though it is likely that any solvable model in three dimensions would have to be far more restricted than either of those cases.

- All of the walk models we considered in Chapter 3 have (or are believed to have) metric exponent $\nu = 1$; that is, the mean squared end-to-end distance of walks of length n scales as $O(n^2)$. As we mentioned in Subsection 3.3.3, there also exist solvable models with $\nu = 1/2$, but those are rather unsatisfying, owing to the fact that the number of such walks does not grow exponentially. It is widely believed that for general SAWs, the metric exponent is $\nu = 3/4$. It would thus be of great interest to find a solvable model with $\nu = 3/4$, or at least $1/2 < \nu < 1$.

Interacting polymer models

- In Section 4.1 we consider two geometries of honeycomb lattice half-planes, and in both cases we are able to derive identities involving a fugacity y associated with vertices on

206

the surface. There are a number of other possible "geometries" that could be considered, including

- a quarter-plane, rather than a half-plane, with (possibly different) fugacities associated with vertices on each of the two boundaries;
- more generally, a wedge of angle $\alpha \in (0, 2\pi)$;
- a "striped" surface (see for example [71]), with two different fugacities associated with alternating vertices in the surface.

- We pointed out earlier in this chapter that the estimates we obtained in Chapter 2 could be improved with longer series or greater strip widths. The same is true of the results of Section 4.2, though of course those results are (to our knowledge) already the most precise estimates of the critical fugacities of the models we consider.

- As with Chapter 2, we restricted our numerical studies to the three regular two-dimensional lattices. To our knowledge there are no existing results for adsorption models on non-regular lattices, but it seems like our methods could be adapted to those settings with minimal difficulty.

  We also hope that there might be a way of applying our methods to adsorption models in three dimensions, though there are a number of problems to address. Firstly, it is unclear what geometry we would need to work in; a slab, with the adsorbing surface located on one of the two bounding planes, seems the most likely candidate, but there are other possibilities. Computing series in three dimensions is also fraught with difficulty, and it seems likely that any series we could generate using current algorithms and hardware would be too short for reliable estimates. Finally, the only estimates we have for the critical points x_c for three-dimensional lattices are very imprecise, and this may limit our ability to compute reliable estimates for the critical surface fugacities.

- Proving Conjectures 4.22 and 4.23 would nicely round out our study of adsorbing prudent walks. However, it would still leave the question as to why prudent walks on the square lattice undergo a first-order transition regardless of whether the endpoint is on the surface or not, while equilateral triangular walks undergo a second-order transition when they end on the surface. We still believe that there should be an intuitive argument explaining this difference.

- The prudent walk adsorption models we considered used only *edge* weights. We could also examine the corresponding *vertex*-weighted models; these should be solvable whenever the edge-weighted models are solvable. It seems reasonable to expect that the adsorption transitions will be of the same order as the edge-weighted models. We could also consider the models with both edges *and* vertices weighted (with different fugacities); such an idea (using SAWs, not solvable walks) is discussed in [103].

- The prudent models can be further enhanced by including another fugacity associated with *stiffness* – the tendency of a walk to take collinear steps, rather than "turn corners" (see for example [94, 70]). The inclusion of such a fugacity should not affect the solvability of a model, though any singularity analyses will almost certainly be made rather more complicated than before. We can also consider the effects of a *pulling force* (see for example [87, 95]).

- Like directed walks, prudent walks could also be used to model inhomogeneous adsorption, where either the polymer or the surface (or both) is comprised of multiple types of units, typically in some kind of regular pattern (see [71]). However, it is unclear at this stage whether such models would be solvable.

- Polymer adsorption is also observed to occur at *penetrable* surfaces, such as the interface between two layers of different fluids. Such phenomena can easily be modelled with SAWs – a fugacity is still associated with vertices or edges in the surface, but now walks are no longer restricted to remain on one side of the surface. Our prudent walks models could easily be adapted to such a setting, though it is unclear precisely which ones would be solvable.

- The model of adsorbing 2-sided prudent walks could be made simpler by placing the impenetrable surface on the line $y = -x$. We would then regain reflective symmetry, through the line $y = x$, and would only need to use one generating function instead of two.[1]

- We finally note that all of the models considered in Chapter 4 are of adsorbing *polymers*, and are based on self-avoiding walks. We can also model the adsorption of *vesicles* by considering self-avoiding *polygons* interacting with a surface (see for example [110, 111]). The techniques we used in Chapter 3 to enumerate solvable classes of SAPs should be adaptable to the adsorption problem in a straightforward manner.

---

[1]Kind thanks to an examiner for pointing this out.

Bibliography

[1] S. E. Alm and S. Janson, *Random self-avoiding walks on one-dimensional lattices*, Comm. Statist. Stochastic Models **6** (1990), no. 2, 169–212.

[2] A. Bacher and M. Bousquet-Mélou, *Weakly directed self-avoiding walks*, J. Combin. Theory Ser. A **118** (2011), no. 8, 2365–2391.

[3] R. Balian and G. Toulouse, *Critical exponents for transitions with $n = -2$ components of the order parameter*, Phys. Rev. Lett. **30** (1973), no. 12, 544–546.

[4] J. R. Banavar, A. Maritan, and A. Stella, *Critical behavior of two-dimensional vesicles in the deflated regime*, Phys. Rev. A **43** (1991), no. 10, 5752–5754.

[5] C. Banderier, M. Bousquet-Mélou, A. Denise, P. Flajolet, D. Gardy, and D. Gouyou-Beauchamps, *Generating functions for generating trees*, Discrete Math. **246** (2002), no. 1-3, 29–55.

[6] M. T. Batchelor, D. Bennett-Wood, and A. L. Owczarek, *Two-dimensional polymer networks at a mixed boundary: Surface and wedge exponents*, Eur. Phys. J. B **5** (1998), no. 1, 139–142.

[7] M. T. Batchelor and C. M. Yung, *Exact results for the adsorption of a flexible self-avoiding polymer chain in two dimensions*, Phys. Rev. Lett. **74** (1995), no. 11, 2026–2029.

[8] N. R. Beaton, M. Bousquet-Mélou, J. de Gier, H. Duminil-Copin, and A. J. Guttmann, *The critical fugacity for surface adsorption of self-avoiding walks on the honeycomb lattice is $1 + \sqrt{2}$*, Preprint, arXiv:1109.0358, 2012.

[9] N. R. Beaton, P. Flajolet, T. M. Garoni, and A. J. Guttmann, *Some new self-avoiding walk and polygon models*, Fund. Inform. **117** (2012), no. 1, 19–33.

[10] N. R. Beaton, P. Flajolet, and A. J. Guttmann, *The unusual asymptotics of three-sided prudent polygons*, J. Phys. A: Math. Theor. **43** (2010), no. 34, 342001, 10pp.

[11] ———, *The enumeration of prudent polygons by area and its unusual asymptotics*, J. Combin. Theory Ser. A **118** (2011), no. 8, 2261–2290.

[12] N. R. Beaton, A. J. Guttmann, and I. Jensen, *A numerical adaptation of self-avoiding walk identities from the honeycomb to other 2D lattices*, J. Phys. A: Math. Theor. **45** (2012), no. 3, 035201, 18pp.

[13] ———, *Two-dimensional self-avoiding walks and polymer adsorption: critical fugacity estimates*, J. Phys. A: Math. Theor. **45** (2012), no. 5, 055208, 12pp.

[14] K. Binder, *Critical behaviour at surfaces*, Phase Transitions and Critical Phenomena, vol. 8, Academic Press, 1983.

[15] H. W. J. Blöte and H. J. Hilhorst, *Spiralling self-avoiding walks: an exact solution*, J. Phys. A: Math. Gen. **17** (1984), no. 3, L111–L115.

[16] M. Bousquet-Mélou, *A method for the enumeration of various classes of column-convex polygons*, Discrete Math. **154** (1996), no. 1-3, 1–25.

[17] ———, *Families of prudent self-avoiding walks*, J. Combin. Theory Ser. A **117** (2010), no. 3, 313–344.

[18] ———, *Counting permutations with no long monotone subsequence via generating trees and the kernel method*, J. Alg. Combin. **33** (2011), no. 4, 571–608.

[19] M. Bousquet-Mélou and R. Brak, *Exactly solved models*, Polygons, Polyominoes and Polycubes, Lecture Notes in Physics, vol. 775, Springer, Dordrecht, 2009, pp. 43–78.

[20] S. R. Broadbent and J. M. Hammersley, *Percolation processes. I. Crystals and mazes*, Proc. Cambridge Philos. Soc. **53** (1957), 629–641.

[21] R. Bulirsch and J. Stoer, *Fehlerabschätzungen und extrapolation mit rationalen funktionen bei verfahren vom Richardson-Typus*, Numer. Math. **6** (1964), 413–427.

[22] J. Cardy, 2011, Private communication.

[23] A. R. Conway, I. G. Enting, and A. J. Guttmann, *Algebraic techniques for enumerating self-avoiding walks on the square lattice*, J. Phys. A: Math. Gen. **26** (1993), no. 7, 1519–1534.

[24] N. G. de Bruijn, *Asymptotic Methods in Analysis*, third ed., Dover, 1970.

[25] P. G. de Gennes, *Exponents for the excluded volume problem as derived by the Wilson method*, Phys. Lett. A **38** (1972), no. 5, 339–340.

[26] M. Delest and X. G. Viennot, *Algebraic languages and polyominoes enumeration*, Theoret. Comput. Sci. **34** (1984), no. 1-2, 169–206.

[27] J. C. Dethridge and A. J. Guttmann, *Prudent self-avoiding walks*, Entropy **10** (2008), no. 3, 309–318.

[28] P. Di Francesco, P. Mathieu, and D. Sénéchal, *Conformal Field Theory*, Springer-Verlag, New York, 1997.

[29] E. Domany, D. Mukamel, B. Nienhuis, and A. Schwimmer, *Duality relations and equivalences for models with O(N) and cubic symmetry*, Nuclear Phys. B **190** (1981), no. 2, 279–287.

[30] M. Drmota and W. Szpankowski, *A master theorem for discrete divide and conquer recurrences*, Proceedings of the Twenty-Second Annual ACM-SIAM Symposium on Discrete Algorithms (Philadelphia, PA), SIAM, 2011, pp. 342–361.

[31] E. Duchi, *On some classes of prudent walks*, FPSAC'05 (Taormina, Italy), 2005.

[32] P. Dumas, *Mean asymptotic behaviour of radix-rational sequences and dilation equations (extended version)*, Technical report, arXiv:0807.1523, 2008.

[33] H. Duminil-Copin and A. Hammond, *Self-avoiding walk is sub-ballistic*, Preprint, arXiv:1205.0401, 2012.

[34] H. Duminil-Copin and S. Smirnov, *The connective constant of the honeycomb lattice equals* $\sqrt{2+\sqrt{2}}$, Ann. of Math. **175** (2012), no. 3, 1653–1665.

[35] B. Duplantier and H. Saleur, *Winding-angle distributions of two-dimensional self-avoiding walks from conformal invariance*, Phys. Rev. Lett. **60** (1988), no. 23, 2343–2346.

[36] A. Elvey Price, J. de Gier, A. J. Guttmann, and A. Lee, *Off-critical parafermions and the winding angle distribution of the O(n) model*, Preprint, arXiv:1203.2959, 2012.

[37] I. G. Enting, *Generating functions for enumerating self-avoiding rings on the square lattice*, J. Phys. A: Math. Gen. **13** (1980), no. 12, 3713–3722.

[38] K. J. Falconer, *The Geometry of Fractal Sets*, Cambridge Tracts in Mathematics, vol. 85, Cambridge University Press, Cambridge, 1985.

[39] M. E. Fisher, A. J. Guttmann, and S. G. Whittington, *Two-dimensional lattice vesicles and polygons*, J. Phys. A: Math. Gen. **24** (1991), no. 13, 3095–3106.

[40] P. Flajolet, P. Dumas, and X. Gourdon, *Mellin transforms and asymptotics: Harmonic sums*, Theoret. Comput. Sci. **144** (1995), no. 1-2, 3–58.

[41] P. Flajolet, S. Gerhold, and B. Salvy, *On the non-holonomic character of logarithms, powers, and the nth prime function*, Electron. J. Combin. **11** (2005), no. 2, 1–16.

[42] P. Flajolet, P. Grabner, P. Kirschenhofer, H. Prodinger, and R. F. Tichy, *Mellin transforms and asymptotics: Digital sums*, Theoret. Comput. Sci. **123** (1994), no. 2, 291–314.

[43] P. Flajolet and B. Richmond, *Generalized digital trees and their difference-differential equations*, Random Structures & Algorithms **3** (1992), no. 3, 305–320.

[44] P. Flajolet and R. Sedgewick, *Analytic Combinatorics*, Cambridge University Press, Cambridge, 2009.

[45] P. J. Flory, *The configuration of real polymer chains*, J. Chem. Phys. **17** (1949), no. 3, 303–310.

[46] G. Forgacs, V. Privman, and H. L. Frisch, *Adsorption-desorption transition of polymer chains interacting with surfaces*, J. Chem. Phys. **90** (1989), no. 6, 3339–3345.

[47] T. M. Garoni, A. J. Guttmann, I. Jensen, and J. C. Dethridge, *Prudent walks and polygons*, J. Phys. A: Math. Theor. **42** (2009), no. 9, 095205, 16pp.

[48] G. Gasper and M. Rahman, *Basic Hypergeometric Series*, Encyclopedia of Mathematics and its Applications, vol. 35, Cambridge University Press, Cambridge, 1990.

[49] P. Grassberger and R. Hegger, *Comment on "Surface critical exponents of self-avoiding walks on a square lattice with an adsorbing linear boundary: A computer simulation study"*, Phys. Rev. E **51** (1995), no. 3, 2674–2676.

[50] I. Guim and T. W. Burkhardt, *Transfer-matrix study of the adsorption of a flexible self-avoiding polymer chain in two dimensions*, J. Phys. A: Math. Gen. **22** (1989), no. 8, 1131–1140.

[51] A. J. Guttmann, *Two conjectures in the theory of self-avoiding random walks*, Proceedings of the First Australian Conference on Combinatorial Mathematics (Newcastle), Univ. of Newcastle Res. Associates, 1972, pp. 5–13.

[52] ———, *Analysis of coefficients*, Phase Transitions and Critical Phenomena, vol. 13, Academic Press, 1989.

[53] ———, *History and introduction to polygon models and polyominoes*, Polygons, Polyominoes and Polycubes, Lecture Notes in Physics, vol. 775, Springer, Dordrecht, 2009, pp. 1–21.

[54] A. J. Guttmann and I. Jensen, *Self-avoiding walks, neighbour-avoiding walks and trails on semiregular lattices*, J. Phys. A: Math. Gen. **31** (1998), no. 40, 8137–8145.

[55] ———, *Series analysis*, Polygons, Polyominoes and Polycubes, Lecture Notes in Physics, vol. 775, Springer, Dordrecht, 2009, pp. 181–202.

[56] A. J. Guttmann and G. S. Joyce, *On a new method of series analysis in lattice statistics*, J. Phys. A: Gen. Phys. **5** (1972), no. 9, L81–L84.

[57] A. J. Guttmann, T. Prellberg, and A. L. Owczarek, *On the symmetry classes of planar self-avoiding walks*, J. Phys. A: Math. Gen. **26** (1993), no. 23, 6615–6623.

[58] A. J. Guttmann and S. G. Whittington, *Two-dimensional lattice embeddings of connected graphs of cyclomatic index two*, J. Phys. A: Math. Gen. **11** (1978), no. 4, 721–729.

[59] A. J. Guttmann and N. C. Wormald, *On the number of spiral self-avoiding walks*, J. Phys. A: Math. Gen. **17** (1984), no. 5, L271–L274.

[60] J. M. Hammersley, *Percolation processes. II. The connective constant*, Proc. Cambridge Philos. Soc. **53** (1957), 642–645.

[61] ———, *The number of polygons on a lattice*, Proc. Cambridge Philos. Soc. **57** (1961), 516–523.

[62] J. M. Hammersley, G. M. Torrie, and S. G. Whittington, *Self-avoiding walks interacting with a surface*, J. Phys. A: Math. Gen. **15** (1982), no. 2, 539–571.

[63] J. M. Hammersley and D. J. A. Welsh, *Further results on the rate of convergence to the connective constant of the hypercubical lattice*, Q. J. Math. **13** (1962), no. 1, 108–110.

[64] T. Hara and G. Slade, *The lace expansion for self-avoiding walks in five or more dimensions*, Rev. Math. Phys. **4** (1992), no. 2, 235–327.

[65] ———, *Self-avoiding walk in five or more dimensions. I. The critical behaviour*, Comm. Math. Phys. **147** (1992), no. 1, 101–136.

[66] G. H. Hardy and S. Ramanujan, *Asymptotic formulæ for the distribution of integers of various types*, Proc. London Math. Soc. Ser. 2 **16** (1917), 112–132.

[67] P. Henrici, *Applied and Computational Complex Analysis*, vol. 1, John Wiley & Sons, New York, 1974.

[68] B. D. Hughes, *Random Walks and Random Environments, vol. 1: Random walks*, Clarendon Press, Oxford, 1995.

[69] Y. Ikhlef and J. Cardy, *Discretely holomorphic parafermions and integrable loop models*, J. Phys. A: Math. Theor. **42** (2009), no. 10, 102001, 11pp.

[70] G. K. Iliev, E. Orlandini, and S. G. Whittington, *Directed walk models of adsorbing semi-flexible polymers subject to an elongational force*, J. Phys. A: Math. Theor. **43** (2010), no. 31, 315202, 16pp.

[71] ———, *Polymers undergoing inhomogeneous adsorption: exact results and Monte Carlo simulations*, J. Phys. A: Math. Theor. **44** (2011), no. 40, 405004, 11pp.

[72] I. Jensen, *A parallel algorithm for the enumeration of self-avoiding polygons on the square lattice*, J. Phys. A: Math. Gen. **36** (2003), no. 21, 5731–5745.

[73] ———, *Enumeration of self-avoiding walks on the square lattice*, J. Phys. A: Math. Gen. **37** (2004), no. 21, 5503–5524.

[74] ———, *Improved lower bounds on the connective constants for two-dimensional self-avoiding walks*, J. Phys. A: Math. Gen. **37** (2004), no. 48, 11521–11529.

[75] ———, *Self-avoiding walks and polygons on the triangular lattice*, J. Stat. Mech. **2004** (2004), P10008, 25pp.

[76] ———, *Honeycomb lattice polygons and walks as a test of series analysis techniques*, J. Phys. Conf. Ser. **42** (2006), 163–178.

[77] I. Jensen and A. J. Guttmann, *Self-avoiding polygons on the square lattice*, J. Phys. A: Math. Gen. **32** (1999), no. 26, 4867–4876.

[78] H. Kesten, *On the number of self-avoiding walks*, J. Math. Phys. **4** (1963), 960–969.

[79] D. E. Knuth, *The average time for carry propagation*, Nederl. Akad. Wetensch. Indag. Math. **40** (1978), no. 2, 238–242.

[80] ———, *The Art of Computer Programming, vol. 1: Fundamental algorithms*, third ed., Addison-Wesley, Reading, MA, 1997.

[81] ———, *The Art of Computer Programming, vol. 2: Seminumerical algorithms*, third ed., Addison-Wesley, Reading, MA, 1997.

[82] ———, *The Art of Computer Programming, vol. 3: Sorting and searching*, second ed., Addison-Wesley, Reading, MA, 1998.

[83] G. F. Lawler, O. Schramm, and W. Werner, *On the scaling limit of planar self-avoiding walk*, Fractal geometry and applications: A jubilee of Benoît Mandelbrot, Part 2, Proc. Sympos. Pure Math., vol. 72, Amer. Math. Soc., Providence, RI, 2004, pp. 339–364.

[84] D. MacDonald, S. Joseph, D. L. Hunter, L. L. Moseley, N. Jan, and A. J. Guttmann, *Self-avoiding walks on the simple cubic lattice*, J. Phys. A: Math. Gen. **33** (2000), no. 34, 5973–5983.

[85] N. Madras and G. Slade, *The Self-Avoiding Walk*, Probability and its Applications, Birkhäuser, Boston, MA, 1993.

[86] I. Majid, N. Jan, A. Coniglio, and H. E. Stanley, *Kinetic growth walk: A new model for linear polymers*, Phys. Rev. Lett. **52** (1984), no. 15, 1257–1260.

[87] P. K. Mishra, S. Kumar, and Y. Singh, *Force-induced desorption of a linear polymer chain adsorbed on an attractive surface*, Europhys. Lett. **69** (2005), no. 1, 102–108.

[88] J. L. Monroe, *Extrapolation and the Bulirsch-Stoer algorithm*, Phys. Rev. E **65** (2002), no. 6, 066116, 8pp.

[89] B. Nienhuis, *Exact critical point and critical exponents of $O(n)$ models in two dimensions*, Phys. Rev. Lett. **49** (1982), no. 15, 1062–1065.

[90] G. L. O'Brien, *Monotonicity of the number of self-avoiding walks*, J. Stat. Phys. **59** (1990), no. 3–4, 969–979.

[91] A. M. Odlyzko, *Periodic oscillations of coefficients of power series that satisfy functional equations*, Adv. in Math. **44** (1982), no. 2, 180–205.

[92] ———, *Asymptotic enumeration methods*, Handbook of Combinatorics, vol. 2, Elsevier, Amsterdam, 1995, pp. 1063–1229.

[93] W. J. C. Orr, *Statistical treatment of polymer solutions at infinite dilution*, Trans. Faraday Soc. **43** (1947), 12–27.

[94] A. L. Owczarek, *Exact solution for semi-flexible partially directed walks at an adsorbing wall*, J. Stat. Mech. **2009** (2009), P11002, 15pp.

[95] ———, *Effect of stiffness on the pulling of an adsorbing polymer from a wall: an exact solution of a partially directed walk model*, J. Phys. A: Math. Theor. **43** (2010), no. 22, 225002, 16pp.

[96] P. Préa, *Exterior self-avoiding walks on the square lattice*, Unpublished manuscript, 1997.

[97] T. Prellberg and R. Brak, *Critical exponents from nonlinear functional equations for partially directed cluster models*, J. Stat. Phys. **78** (1995), no. 3–4, 701–730.

[98] V. Privman and N. M. Švrakić, *Directed Models of Polymers, Interfaces, and Clusters: Scaling and Finite-Size Properties*, Lecture Notes in Physics, vol. 338, Springer-Verlag, Berlin, 1989.

[99] H. Prodinger, *The kernel method: a collection of examples*, Sém. Lothar. Combin. **50** (2003/04), Art. B50f, 19pp.

[100] A. Rechnitzer, *Haruspicy 2: The anisotropic generating function of self-avoiding polygons is not D-finite*, J. Combin. Theory Ser. A **113** (2006), no. 3, 520–546.

[101] C. Richard, I. Jensen, and A. J. Guttmann, *Scaling function and universal amplitude combinations for self-avoiding polygons*, J. Phys. A: Math. Gen. **34** (2001), no. 36, L495–501.

[102] _____ , *Area distribution and scaling function for punctured polygons*, Electron. J. Combin. **15** (2008), #R53, 50pp.

[103] G. Rychlewski and S. G. Whittington, *Self-avoiding walks and polymer adsorption: edge and vertex weighting*, J. Phys. A: Math. Theor. **44** (2011), no. 9, 095006, 7pp.

[104] _____ , *Self-avoiding walks and polymer adsorption: Low temperature behaviour*, J. Stat. Phys. **145** (2011), no. 3, 661–668.

[105] U. Schwerdtfeger, *Exact solution of two classes of prudent polygons*, European J. Combin. **31** (2010), no. 3, 765–779.

[106] S. Smirnov, *Discrete complex analysis and probability*, Proceedings of the International Congress of Mathematicians, vol. I (Hyderabad, India), Hindustan Book Agency, 2010, pp. 595–621.

[107] H. E. Stanley, *Dependence of critical properties on dimensionality of spins*, Phys. Rev. Lett. **20** (1968), no. 12, 589–592.

[108] W. Szpankowski, *Average-Case Analysis of Algorithms on Sequences*, Discrete Mathematics and Optimization, John Wiley & Sons, New York, 2001.

[109] H. N. V. Temperley, *Graph Theory and Applications*, Ellis Horward Ltd., Chichester, 1981.

[110] E. J. J. van Rensburg, *Adsorbing staircase walks and staircase polygons*, Ann. Comb. **3** (1999), no. 2–4, 451–473.

[111] _____ , *Collapsing and adsorbing polygons*, J. Phys. A: Math. Gen. **31** (1999), no. 41, 8295–8306.

[112] _____ , *The statistical mechanics of interacting walks, polygons, animals, and vesicles*, Oxford University Press, 2000.

[113] _____ , *Statistical mechanics of directed models of polymers in the square lattice*, J. Phys. A: Math. Gen. **36** (2003), no. 15, R11–R61.

[114] _____ , *Adsorbing Motzkin paths*, J. Phys. A: Math. Theor. **43** (2010), no. 48, 485006, 19pp.

[115] E. J. J. van Rensburg, E. Orlandini, and S. G. Whittington, *Self-avoiding walks in a slab: rigorous results*, J. Phys. A: Math. Gen. **39** (2006), no. 45, 13869–13902.

[116] E. J. J. van Rensburg, T. Prellberg, and A. Rechnitzer, *Partially directed paths in a wedge*, J. Combin. Theory Ser. A **115** (2008), no. 4, 623–650.

[117] E. J. J. van Rensburg and A. Rechnitzer, *Multiple Markov chain Monte Carlo study of adsorbing self-avoiding walks in two and in three dimensions*, J. Phys. A: Math. Gen. **37** (2004), no. 27, 6875–6898.

[118] S. G. Whittington, *Self-avoiding walks terminally attached to an interface*, J. Chem. Phys. **63** (1975), no. 2, 779–785.

[119] _____, *A directed-walk model of copolymer adsorption*, J. Phys. A: Math. Gen. **31** (1999), no. 44, 8797–8803.

[120] _____, *Lattice polygons and related objects*, Polygons, Polyominoes and Polycubes, Lecture Notes in Physics, vol. 775, Springer, Dordrecht, 2009, pp. 23–41.

[121] D. V. Widder, *The Laplace Transform*, Princeton Mathematical Series, vol. 6, Princeton University Press, Princeton, NJ, 1941.

[122] D. Zeilberger, *Symbol-crunching with the transfer-matrix method in order to count skinny physical creatures*, Integers 0 (2000), A9, 34pp.

Appendix A

$B_T(x_c) \to 0$ as $T \to \infty$

Define $B_T(x)$ as in Section 2.1; that is, B_T is the generating function of self-avoiding bridges on the honeycomb lattice which span a strip of width T. In this appendix we present a proof of the following theorem.

Theorem A.1. *The generating function $B_T(x)$ of self-avoiding bridges on the honeycomb lattice spanning a strip of width T satisfies*

$$\lim_{T \to \infty} B_T(x_c) = 0, \tag{A.1}$$

where $x_c = 1/\sqrt{2 + \sqrt{2}}$ is the critical point of SAWs on the honeycomb lattice.

This result, together with Proposition 4.9, completes the proof of Theorem 4.2 – that the critical surface fugacity for adsorbing SAWs on the honeycomb lattice is $1 + \sqrt{2}$.

Before starting the proof, let us introduce some additional notation; some of this notation will differ from the rest of the thesis, but as this Appendix is almost entirely self-contained, we hope that no confusion will arise. The set of mid-edges of the honeycomb lattice is denoted by $\mathbb{H}$. The coordinates of a point v will be denoted $(x(v), y(v))$. (Note that we take the edges of the lattice to be of unit length.) The lattice has an origin $a \in \mathbb{H}$, at coordinates $(0,0)$. We consider SAWs that start and end at a mid-edge. A SAW γ is denoted by the sequence $(\gamma_0, \ldots, \gamma_n)$ of its mid-edges. The length of γ, that is, the number of vertices of the lattice it visits, is denoted $|\gamma| = n$. To lighten notation, we often omit floor symbols, especially in indices: for instances, γ_t should be understood as $\gamma_{\lfloor t \rfloor}$.

In Chapters 2 and 4 we discussed bridges in a vertical strip of width T. In this Appendix we rotate the strip by 90°, so that bridges now span a *horizontal* strip of height T. The convention is chosen in such a way that a bridge of height T of minimal length contains exactly T vertical edges, one of them being split into two half-edges (see Figure A.1).[1] In general, we call *bridge* any

---

[1] This is in contrast to Chapters 2 and 4, where we would define such a bridge to have $T + 1$ vertical edges. Redefining the height in this way makes things notationally simpler here.
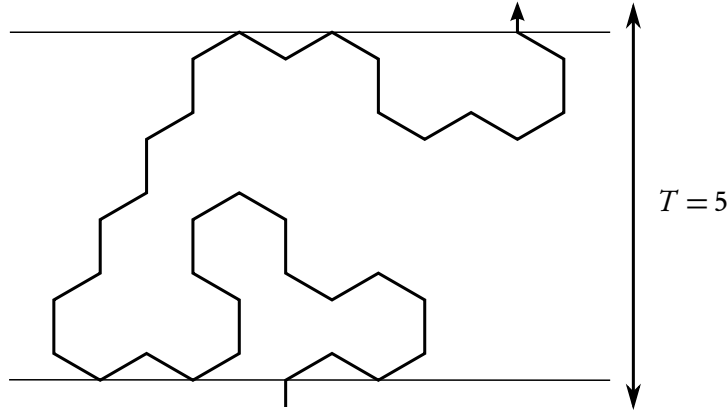
218

Figure A.1: A bridge of height $T = 5$ on the honeycomb lattice.

SAW $\gamma = (\gamma_0, \ldots, \gamma_n)$ that is a bridge of height T for some T. Equivalently, $y(\gamma_0) < y(\gamma_i) < y(\gamma_n)$ for $0 < i < n$. The set of bridges of length n is denoted by SAB_n.

The set R_γ of *renewal points* of $\gamma \in \mathrm{SAB}_n$ is the set of points of the form γ_i with $0 \leq i \leq n$, for which $\gamma_{[0,i]} := (\gamma_0, \ldots, \gamma_i)$ and $\gamma_{[i,n]} := (\gamma_i, \ldots, \gamma_n)$ are bridges. We denote by $\mathbf{r}_0(\gamma), \mathbf{r}_1(\gamma), \ldots$ the indices of the renewal points. That is, $\mathbf{r}_0(\gamma) = 0$ and $\mathbf{r}_{k+1}(\gamma) = \inf\{j > \mathbf{r}_k(\gamma) : \gamma_j \in \mathsf{R}_\gamma\}$ for each $k \leq |\mathsf{R}_\gamma| - 1$. When no confusion is possible, we often denote $\mathbf{r}_k(\gamma)$ by $\mathbf{r}_k$.

A bridge $\gamma \in \mathrm{SAB}_n$ is *irreducible* if its only renewal points are γ_0 and γ_n. Let iSAB be the set of irreducible bridges of arbitrary length starting from a. Every bridge γ is the concatenation of a finite number of irreducible bridges, the decomposition is unique and the set R_γ is the union of the initial and terminal points of the bridges that comprise this decomposition.

Kesten's relation for irreducible bridges (see [85, Section 4.2] or [78]) on the hypercubic lattice $\mathbb{Z}^d$ can be easily adapted to the honeycomb lattice. It gives

$$\sum_{\gamma \in \mathrm{iSAB}} x_{\mathrm{c}}^{|\gamma|} = 1.$$

This enables us to define a probability measure $\mathbb{P}_{\mathrm{iSAB}}$ on iSAB by setting $\mathbb{P}_{\mathrm{iSAB}}(\gamma) = x_{\mathrm{c}}^{|\gamma|}$. Let $\mathbb{P}_{\mathrm{iSAB}}^{\otimes \mathbb{N}}$ denote the law on semi-infinite walks $\gamma : \mathbb{N} \to \mathbb{H}$ formed by the concatenation of infinitely many independent samples $\gamma^{[1]}, \gamma^{[2]}, \ldots$ of $\mathbb{P}_{\mathrm{iSAB}}$. We refer to [85, Section 8.3] for details on related measures in the case of $\mathbb{Z}^d$. The definition of R_γ and the indexation of renewal points extend to this context (we obtain an infinite sequence $(\mathbf{r}_k)_{k \in \mathbb{N}}$).

Observe that a bridge γ of length n has height $\mathsf{H}(\gamma) = \frac{2}{3}y(\gamma_n)$ (since edges have unit length; in particular a bridge of length 2 has height 1). We define its *width* by

$$\mathsf{W}(\gamma) = \frac{1}{\sqrt{3}} \max\{x(\gamma_k) - x(\gamma_{k'}), 0 \leq k, k' \leq n\},$$

so that a bridge of length 2 has width $1/2$.

It was proved in Section 4.1 that $B_T(x_{\mathrm{c}}, 1)$ converges as $T \to \infty$. We provide here an alternative proof, and relate the limit value to the average height of irreducible bridges.

Lemma A.2. *As $T \to \infty$,*

$$B_T(x_c, 1) \to \frac{1}{\mathbb{E}_{\text{iSAB}}(\mathsf{H}(\gamma))}.$$

Proof. The result follows from standard renewal theory. We can for instance apply [85, Theorem 4.2.2(b)] to the sequence

$$f_T := \sum_{\gamma \in \text{iSAB}: \, \mathsf{H}(\gamma) = T} x_c^{|\gamma|}.$$

Indeed, with the notation of this theorem, $v_T = B_T(x_c, 1)$ and $\sum_k k f_k = \mathbb{E}_{\text{iSAB}}(\mathsf{H}(\gamma))$. ∎

Thus Theorem A.1 is equivalent to

$$\mathbb{E}_{\text{iSAB}}(\mathsf{H}(\gamma)) = \infty.$$

We will prove this by contradiction. Under the assumption that $\mathbb{E}_{\text{iSAB}}(\mathsf{H}(\gamma)) < \infty$, we first show that $\mathbb{E}_{\text{iSAB}}(\mathsf{W}(\gamma)) < \infty$. Then, we show that under these two conditions, an infinite bridge is very narrow. The last step consists of proving that this cannot be the case. The argument uses a "stickbreak" operation which perturbs a bridge by selecting a subpath and rotating it clockwise by $\frac{\pi}{3}$. The new path is a self-avoiding bridge for an adequately chosen subpath. But its width is relatively large, contradicting the fact that bridges are narrow. The strategy of proof is inspired by a recent paper by Duminil-Copin and Hammond, where SAWs are proved to be sub-ballistic [33]. (That is, they proved that if the metric exponent ν, governing (among other things) the mean squared end-to-end distance of SAWs, exists, then it is less than 1.)

Proposition A.3. *If $\mathbb{E}_{\text{iSAB}}(\mathsf{H}(\gamma)) < \infty$, then $\mathbb{E}_{\text{iSAB}}(\mathsf{W}(\gamma)) < \infty$.*

Proof. Consider the rectangular domain $R_{T,L}$ of $\mathbb{H}$ depicted in Figure A.2, with its boundary partitioned into four subsets $\mathscr{A}$, $\mathscr{B}$, $\mathscr{E}^-$ and $\mathscr{E}^+$ (the mid-edges of $\mathscr{E}^+$ point up, those of $\mathscr{E}^-$ point down). We do not consider any kind of interactions here. As in Section 2.1, we define four generating functions counting SAWs in the rectangle, going from a to a mid-edge of the boundary. First, we set

$$\tilde{A}_{T,L}(x) := \sum_{\gamma: a \rightsquigarrow \mathscr{A} \backslash \{a\}} x^{|\gamma|},$$

and then the generating functions $\tilde{B}_{T,L}(x)$, $\tilde{E}_{T,L}^-(x)$ and $\tilde{E}_{T,L}^+(x)$ are defined similarly. We now use the local identity of Lemma 2.1 with $n = 0$, $\theta = \pi/2$, $\sigma = 5\pi/24$, and $x_c^{-1} = 2\cos(\pi/8)$ to prove the following global identity, analogous to (2.10):

$$1 = \alpha \tilde{A}_{T,L}(x_c) + \tilde{B}_{T,L}(x_c) + \varepsilon^+ \tilde{E}_{T,L}^+(x_c) + \varepsilon^- \tilde{E}_{T,L}^-(x_c), \tag{A.2}$$

where $\alpha = \cos(\frac{3\pi}{8})$, $\varepsilon^- = \cos(\frac{\pi}{4})$ and $\varepsilon^+ = \cos(\frac{\pi}{8})$.

Since we always evaluate our generating functions at $x = x_c$, we will almost systematically omit the variable x_c, so that $\tilde{A}_{T,L}$ now means $\tilde{A}_{T,L}(x_c)$, and so on.
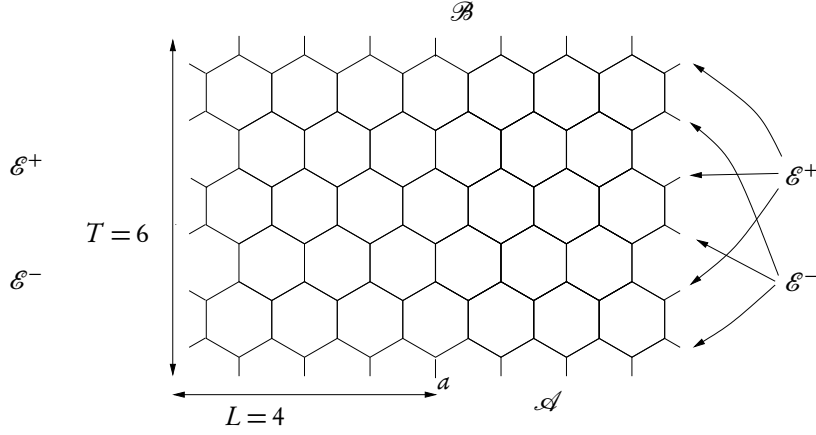
Figure A.2: The rectangular domain $R_{T,L}$ with $T = 6$ and $L = 4$.

As in Section 2.1, we would like $\tilde{E}^{\pm}_{T,L}$ to tend to 0 as the size of the rectangle increases. This holds for fixed T as L increases, with the same argument as before, but now we want both T and L to tend to infinity, so the matter is a bit more delicate. Recall that a *loop* is a SAW starting from a, confined to the upper half-plane, and ending on the line $y = 0$. For $L \in \mathbb{N}$, let $\mathfrak{a}_L(x)$ be the generating function of loops ending L cells to the right of a. We will bound $\tilde{E}^{\pm}_{T,L}$ in terms of $\mathfrak{a}_{2L}$.

For $m \in \mathbb{N}$, let $\mathfrak{e}^+_m(x)$ be the generating function of walks in $R_{T,L}$ ending on the the right side of the rectangle, on the mth row of $\mathscr{E}^+$, so that, by symmetry, $\tilde{E}^+_{T,L} = 2\sum_{m \leq \lfloor \frac{T}{2} \rfloor} \mathfrak{e}^+_m$. Using a reflection argument and the Cauchy-Schwarz inequality, we find

$$\left(\tilde{E}^+_{T,L}\right)^2 \leq 4\lfloor \tfrac{T}{2} \rfloor \sum_{m \leq \lfloor \frac{T}{2} \rfloor} (\mathfrak{e}^+_m)^2 \leq 4\lfloor \tfrac{T}{2} \rfloor \, x_{\mathrm{c}}^{-1} \mathfrak{a}_{2L}. \tag{A.3}$$

The second inequality comes from the fact that one can concatenate two walks contributing to $\mathfrak{e}^+_m$ (after reflecting the second one) by adding a step between them in order to create a loop contributing to $\mathfrak{a}_{2L}$. We obtain a similar upper bound for $\tilde{E}^-_{T,L}$ with $\lfloor \frac{T}{2} \rfloor$ replaced by $\lceil \frac{T}{2} \rceil$.

Assume that we couple $T \equiv T_k$ and $L \equiv L_k$ so that both tend to infinity as k grows, and $T\mathfrak{a}_{2L} \longrightarrow 0$. Then $\tilde{E}^+_{T,L}$ and $\tilde{E}^-_{T,L}$ tend to 0. Moreover, $\tilde{A}_{T,L}$ increases with L and T, and converges to $A \equiv A(x_{\mathrm{c}})$, where $A(x)$ is the generating function of loops (this is finite, as discussed just before Proposition 4.9). Returning to (A.2) shows that $\tilde{B}_{T,L}$ must also converge, and gives

$$\lim_k \tilde{B}_{T_k,L_k} = 1 - \alpha A(x_{\mathrm{c}})$$
$$= \lim_T B_T(x_{\mathrm{c}}, 1) \qquad \text{by (2.12)}$$
$$> 0 \qquad \text{by assumption.} \tag{A.4}$$

Let us now return to random infinite bridges and use them to give an upper bound on $\tilde{B}_{T,L}$.

Let $0 < \delta < 1/\mathbb{E}_{\text{iSAB}}(\mathsf{H}(\gamma))$. We have

$$
\begin{aligned}
\tilde{B}_{T,L} &= \sum_{\gamma:a \rightsquigarrow \mathscr{B}} x_{\mathrm{c}}^{|\gamma|} \\
&\leq \mathbb{P}_{\text{iSAB}}^{\otimes \mathbb{N}}\left(\exists n \in \mathbb{N} : \mathsf{H}(\gamma_{[0,\mathbf{r}_n]}) = T \text{ and } \mathsf{W}(\gamma_{[0,\mathbf{r}_n]}) \leq 2L\right) \\
&\leq \mathbb{P}_{\text{iSAB}}^{\otimes \mathbb{N}}\left(\mathsf{H}(\gamma_{[0,\mathbf{r}_{\delta T}]}) \geq T\right) + \mathbb{P}_{\text{iSAB}}^{\otimes \mathbb{N}}\left(\exists n \geq \delta T : \mathsf{H}(\gamma_{[0,\mathbf{r}_n]}) = T \text{ and } \mathsf{W}(\gamma_{[0,\mathbf{r}_n]}) \leq 2L\right).
\end{aligned}
$$

Let $\gamma^{[i]}$ be the ith irreducible bridge of γ. Since the $\gamma^{[i]}$'s are independent, we obtain

$$
\begin{aligned}
\tilde{B}_{T,L} &\leq \mathbb{P}_{\text{iSAB}}^{\otimes \mathbb{N}}\left(\mathsf{H}(\gamma_{[0,\mathbf{r}_{\delta T}]}) \geq T\right) + \mathbb{P}_{\text{iSAB}}^{\otimes \mathbb{N}}\left(\forall i \leq \delta T, \mathsf{W}(\gamma^{[i]}) \leq 2L\right) \\
&= \mathbb{P}_{\text{iSAB}}^{\otimes \mathbb{N}}\left(\mathsf{H}(\gamma_{[0,\mathbf{r}_{\delta T}]}) \geq T\right) + \mathbb{P}_{\text{iSAB}}(\mathsf{W}(\gamma) \leq 2L)^{\delta T} \\
&\leq \mathbb{P}_{\text{iSAB}}^{\otimes \mathbb{N}}\left(\mathsf{H}(\gamma_{[0,\mathbf{r}_{\delta T}]}) \geq T\right) + \exp\left(-\delta T \, \mathbb{P}_{\text{iSAB}}(\mathsf{W}(\gamma) > 2L)\right).
\end{aligned}
$$

Note that

$$
\mathsf{H}(\gamma_{[0,\mathbf{r}_{\delta T}]}) = \sum_{i=1}^{\delta T} \mathsf{H}(\gamma^{[i]}).
$$

Hence the law of large numbers, together with the fact that $\delta \cdot \mathbb{E}_{\text{iSAB}}(\mathsf{H}(\gamma)) < 1$, implies that $\mathbb{P}_{\text{iSAB}}^{\otimes \mathbb{N}}\left(\mathsf{H}(\gamma_{[0,\mathbf{r}_{\delta T}]}) \geq T\right)$ tends to 0 as $T \to \infty$. Hence, if we can couple $T \equiv T_k$ and $L \equiv L_k$ in such a way that $T \mathbb{P}_{\text{iSAB}}(\mathsf{W}(\gamma) > 2L)$ tends to infinity, then $\tilde{B}_{T,L}$ tends to zero.

We now argue ad absurdum. Assume that $\mathbb{E}_{\text{iSAB}}(\mathsf{W}(\gamma)) = \infty$. Then

$$
\limsup_{L \to \infty} \frac{\mathbb{P}_{\text{iSAB}}(\mathsf{W}(\gamma) > 2L)}{\mathfrak{a}_{2L}} = \infty,
$$

since $\mathfrak{a}_L$ is the term of a converging series (namely, the generating function $A(x_{\mathrm{c}})$ of loops) and $\mathbb{P}_{\text{iSAB}}(\mathsf{W}(\gamma) > L)$ is non-increasing in L and is the term of a diverging series (indeed, it sums to $\mathbb{E}_{\text{iSAB}}(\mathsf{W}(\gamma)) = \infty$). Let $(L_k)_k$ be a sequence such that

$$
\lim_{k \to \infty} \frac{\mathbb{P}_{\text{iSAB}}(\mathsf{W}(\gamma) > 2L_k)}{\mathfrak{a}_{2L_k}} = \infty,
$$

and take

$$
T_k = \left\lfloor \frac{1}{\sqrt{\mathfrak{a}_{2L_k} \mathbb{P}_{\text{iSAB}}(\mathsf{W}(\gamma) > 2L_k)}} \right\rfloor.
$$

Then

$$
T_k \, \mathbb{P}_{\text{iSAB}}(\mathsf{W}(\gamma) > 2L_k) \to \infty \quad \text{and} \quad T_k \, \mathfrak{a}_{2L_k} \to 0.
$$

According to our two estimates of $\tilde{B}_{T,L}$, this means that $\lim_k \tilde{B}_{T_k,L_k}$ is both zero and a positive number, an absurdity. Therefore, $\mathbb{E}_{\text{iSAB}}(\mathsf{W}(\gamma)) < \infty$. $\blacksquare$

Let Ω be the set of bi-infinite walks $\gamma : \mathbb{Z} \to \mathbb{H}$ such that $\gamma_0 = a$. Let $(\gamma^{[i]}, i \in \mathbb{Z})$ be a bi-infinite sequence of irreducible bridges sampled independently according to $\mathbb{P}_{\text{iSAB}}$. Let $\mathbb{P}_{\text{iSAB}}^{\otimes \mathbb{Z}}$ denote the law on Ω formed by concatenating the bridges $\gamma^{[i]}, i \in \mathbb{Z}$ in such a way that $\gamma^{[1]}$

starts at a. Let $\mathscr{F}$ be the σ-algebra generated by events depending on a finite number of vertices of the walk.

We extend the indexation of renewal points to these bi-infinite bridges (we obtain a bi-infinite sequence $(\mathbf{r}_n(\gamma))_{n \in \mathbb{Z}}$ such that $r_0(\gamma) = 0$). Let $\tau : \Omega \to \Omega$ be the *shift* defined by $\tau(\gamma)_i = \gamma_{i + \mathbf{r}_1(\gamma)} - \gamma_{\mathbf{r}_1(\gamma)}$ for every $i \in \mathbb{Z}$. (This is only defined if $\mathbf{r}_1$ exists, but this is the case with probability 1 under $\mathbb{P}_{\text{iSAB}}^{\otimes \mathbb{Z}}$.) The shift translates the walk so that $\mathbf{r}_1(\gamma)$ is now at the origin a of the lattice. Note that $\mathbf{r}_i(\tau(\gamma)) = \mathbf{r}_{i+1}(\gamma) - \mathbf{r}_1(\gamma)$. Let σ denote the reflection in the real axis.

Proposition A.4. *The measure $\mathbb{P}_{\text{iSAB}}^{\otimes \mathbb{Z}}$ satisfies the following properties.*

(P$_1$) *It is invariant under the shift τ.*

(P$_2$) *The shift τ is ergodic for $(\Omega, \mathscr{F}, \mathbb{P}_{\text{iSAB}}^{\otimes \mathbb{Z}})$.*

(P$_3$) *Under $\mathbb{P}_{\text{iSAB}}^{\otimes \mathbb{Z}}$, the random variables $(\sigma \gamma_n)_{n \leq 0}$ and $(\gamma_n)_{n \leq 0}$ are independent and identically distributed.*

Proof. Property (P$_1$) is fairly straightforward. Indeed, for every $n > 0$, the law of $\gamma_{[\mathbf{r}_{-n}(\gamma), \mathbf{r}_n(\gamma)]}$ determines, in the high-n limit, the law of γ (since we work with the σ-algebra $\mathscr{F}$). Now, the laws of $\tau(\gamma_{[\mathbf{r}_{-n+1}(\gamma), \mathbf{r}_{n+1}(\gamma)]})$ and $\gamma_{[\mathbf{r}_{-n}(\gamma), \mathbf{r}_n(\gamma)]}$ are the same by construction (both are the law of $2n$ concatenated independent irreducible bridges). Thus (P$_1$) follows by letting $n \to \infty$.

Let us turn to (P$_2$). Consider a shift-invariant event A. We want to show that $\mathbb{P}_{\text{iSAB}}^{\otimes \mathbb{Z}}(A) \in \{0,1\}$. Let $\varepsilon > 0$. There exists $n > 0$ and an event A_n depending only on the vertices $\gamma_{-n}, \ldots, \gamma_n$ such that $\mathbb{P}_{\text{iSAB}}^{\otimes \mathbb{Z}}(A_n \Delta A) \leq \varepsilon$, where Δ denotes the symmetric difference. In particular, $|\mathbb{P}_{\text{iSAB}}^{\otimes \mathbb{Z}}(A) - \mathbb{P}_{\text{iSAB}}^{\otimes \mathbb{Z}}(A_n)| \leq \varepsilon$. By extension, A_n depends only on vertices in $\gamma_{\mathbf{r}_{-n}}, \ldots, \gamma_{\mathbf{r}_n}$. Invariance of A under τ implies that $A = \tau^{-2n}(A)$, so that

$$\mathbb{P}_{\text{iSAB}}^{\otimes \mathbb{Z}}(A) = \mathbb{P}_{\text{iSAB}}^{\otimes \mathbb{Z}}\left(A \cap \tau^{-2n}(A)\right). \tag{A.5}$$

Moreover,

$$\left|\mathbb{P}_{\text{iSAB}}^{\otimes \mathbb{Z}}\left(A \cap \tau^{-2n}(A)\right) - \mathbb{P}_{\text{iSAB}}^{\otimes \mathbb{Z}}\left(A_n \cap \tau^{-2n}(A_n)\right)\right|$$
$$\leq \mathbb{P}_{\text{iSAB}}^{\otimes \mathbb{Z}}\left(A \Delta A_n\right) + \mathbb{P}_{\text{iSAB}}^{\otimes \mathbb{Z}}\left(\tau^{-2n}(A) \Delta \tau^{-2n}(A_n)\right) \leq 2\varepsilon.$$

Using (A.5) and the independence between the walk before and after $\mathbf{r}_n$, this reads

$$|\mathbb{P}_{\text{iSAB}}^{\otimes \mathbb{Z}}(A) - \mathbb{P}_{\text{iSAB}}^{\otimes \mathbb{Z}}(A_n)^2| \leq 2\varepsilon,$$

which, combined with $|\mathbb{P}_{\text{iSAB}}^{\otimes \mathbb{Z}}(A) - \mathbb{P}_{\text{iSAB}}^{\otimes \mathbb{Z}}(A_n)| \leq \varepsilon$, implies

$$|\mathbb{P}_{\text{iSAB}}^{\otimes \mathbb{Z}}(A) - \mathbb{P}_{\text{iSAB}}^{\otimes \mathbb{Z}}(A)^2| \leq 4\varepsilon.$$

By letting ε tend to 0, we obtain that $\mathbb{P}_{\text{iSAB}}^{\otimes \mathbb{Z}}(A) = \mathbb{P}_{\text{iSAB}}^{\otimes \mathbb{Z}}(A)^2$ and therefore $\mathbb{P}_{\text{iSAB}}^{\otimes \mathbb{Z}}(A) \in \{0,1\}$. Hence (P$_2$) is proved.
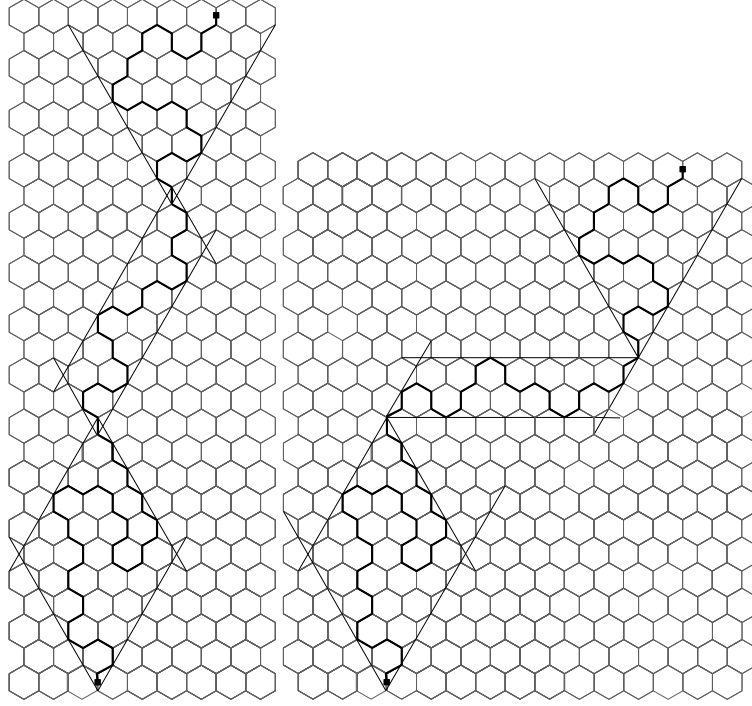
Figure A.3: *Left*: A bridge having 3 diamond points. *Right*: A stickbreak operation applied to this bridge.

Since the law of irreducible bridges is invariant (up to a translation) under reflection with respect to a horizontal line, (P_3) is straightforward. ∎

Renewal points separate a walk into two parts, located below and above the point. We now introduce a more restrictive notion, illustrated in Figure A.3 (left). A mid-edge γ_k of a walk γ is said to be a *diamond point* if

- it lies on a vertical edge of the lattice,

- the walk is contained in the cone

$$\left(\left(\gamma_k - \tfrac{i}{2}\right) + \mathbb{R}_+ e^{i\pi/3} + \mathbb{R}_+ e^{2i\pi/3}\right) \cup \left(\left(\gamma_k + \tfrac{i}{2}\right) - \mathbb{R}_+ e^{i\pi/3} - \mathbb{R}_+ e^{2i\pi/3}\right)$$

(recall that edges have length 1). The set of diamond points of γ is denoted by $\mathbf{D}_\gamma$. Of course, it is a subset of $\mathbf{R}_\gamma$. The following proposition tells us that, under our assumption, a positive fraction of renewal points are diamond points.

Proposition A.5. *If* $\mathbb{E}_{\mathrm{iSAB}}(\mathsf{H}(\gamma)) < \infty$, *then there exists* $\delta > 0$ *such that*

$$\mathbb{P}_{\mathrm{iSAB}}^{\otimes \mathbb{N}} \left(\liminf_{n \to \infty} \frac{|\mathbf{D}_\gamma \cap \{0, \ldots, \mathbf{r}_n\}|}{n} \geq \delta \right) = 1.$$

Let us first provide a heuristic argument. Since $\mathbb{E}_{\mathrm{iSAB}}(\mathsf{H}(\gamma))$ is finite, so is $\mathbb{E}_{\mathrm{iSAB}}(\mathsf{W}(\gamma))$ (Proposition A.3). Then $\mathbb{E}_{\mathrm{iSAB}}(x(\gamma_{|\gamma|})) = 0$, and the law of large numbers implies that the

prefixes of an infinite bridge are tall and skinny – that is, height grows linearly, width grows sub-linearly. So the probability of a bridge staying within a cone as thin as one likes is positive and a similar result going backwards. Thus, diamond points occur with positive density among renewal points.

Proof. We first prove that $\mathbb{P}_{\mathrm{iSAB}}^{\otimes \mathbb{Z}}(\gamma_0 \in \mathbf{D}_\gamma) > 0$. Proposition A.3 shows that $\mathbb{E}_{\mathrm{iSAB}}(W(\gamma)) < \infty$. Hence $\mathbb{E}_{\mathrm{iSAB}}(x(\gamma_{|\gamma|}))$ is well-defined, and is 0 since the law of an irreducible bridge is invariant under the reflection with respect to the imaginary axis. The law of large numbers thus implies that, $\mathbb{P}_{\mathrm{iSAB}}^{\otimes \mathbb{N}}$-almost surely, $x(\gamma_{\mathbf{r}_n})/n \longrightarrow 0$. Since the expected width of irreducible bridges is finite, a classical use of the Borel-Cantelli Lemma shows that $W(\gamma_{[\mathbf{r}_n, \mathbf{r}_{n+1}]})/n \longrightarrow 0$ almost surely. Thus

$$\frac{1}{n}\left(|x(\gamma_{\mathbf{r}_n})| + W(\gamma_{[\mathbf{r}_n, \mathbf{r}_{n+1}]})\right) \longrightarrow 0 \qquad \text{a.s.}$$

Since

$$W(\gamma_{[0,\mathbf{r}_n]}) \le 2\max\{|x(\gamma_{\mathbf{r}_k})| + W(\gamma_{[\mathbf{r}_k, \mathbf{r}_{k+1}]}), \ 0 \le k \le n-1\},$$

we find that, $\mathbb{P}_{\mathrm{iSAB}}^{\otimes \mathbb{N}}$-almost surely, $W(\gamma_{[0,\mathbf{r}_n]})/n \longrightarrow 0$.

Let us now apply the law of large numbers to $y(\gamma_{\mathbf{r}_n})$. We obtain that, $\mathbb{P}_{\mathrm{iSAB}}^{\otimes \mathbb{N}}$-almost surely, $y(\gamma_{\mathbf{r}_n})/n \longrightarrow \frac{3}{2}\mathbb{E}_{\mathrm{iSAB}}(H(\gamma)) > 0$.

We deduce that

$$I(\gamma) := \inf_{k \ge 0}\left(y(\gamma_k) - \sqrt{3}|x(\gamma_k)| + 1/2\right)$$

is finite $\mathbb{P}_{\mathrm{iSAB}}^{\otimes \mathbb{N}}$-almost surely. Note that for an infinite bridge $\gamma = (\gamma_0, \gamma_1, \ldots)$, the origin γ_0 is a diamond point if and only if $I(\gamma) \ge 0$. Let $K \in \mathbb{N}$ be such that $\rho_K := \mathbb{P}_{\mathrm{iSAB}}^{\otimes \mathbb{N}}(I(\gamma) \ge -K) > 0$. We are going to show that

$$\rho_0 \ge (2x_{\mathrm{c}}^4)^K \rho_K > 0. \tag{A.6}$$

To prove (A.6), consider an experiment under which the law $\mathbb{P}_{\mathrm{iSAB}}^{\otimes \mathbb{N}}$ is constructed by first concatenating K independent samples of $\mathbb{P}_{\mathrm{iSAB}}$ (starting from a) and then an independent sample γ' of $\mathbb{P}_{\mathrm{iSAB}}^{\otimes \mathbb{N}}$. If each of the K samples happens to be a walk of length 4 going from a to $a + 3i$ and $I(\gamma') \ge -K$, then the complete walk γ satisfies $I(\gamma) \ge 0$. The probability that the ith sample of $\mathbb{P}_{\mathrm{iSAB}}$ is a walk of length four going from a to $a + 3i$ is $2x_{\mathrm{c}}^4$. Thus, the experiment behaves as described with probability $(2x_{\mathrm{c}}^4)^K \rho_K$, and we obtain (A.6), that is, $\mathbb{P}_{\mathrm{iSAB}}^{\otimes \mathbb{N}}(\gamma_0 \in \mathbf{D}_\gamma) > 0$.

Using Property (P_3) of Proposition A.4, we deduce that

$$\delta := \mathbb{P}_{\mathrm{iSAB}}^{\otimes \mathbb{Z}}(\gamma_0 \in \mathbf{D}_\gamma) = \left(\mathbb{P}_{\mathrm{iSAB}}^{\otimes \mathbb{N}}(\gamma_0 \in \mathbf{D}_\gamma)\right)^2 > 0.$$

The shift τ being ergodic (cf. Property (P_2) of Proposition A.4), the ergodic theorem, applied to $\mathbb{1}_{\gamma_0 \in \mathbf{D}_\gamma}$, gives

$$\mathbb{P}_{\mathrm{iSAB}}^{\otimes \mathbb{Z}}\left(\lim_{n \to \infty} \frac{|\mathbf{D}_\gamma \cap \{0, \ldots, \mathbf{r}_n(\gamma)\}|}{n} = \delta\right) = 1.$$

Let γ be a bi-infinite bridge, and denote $\gamma^+ = \gamma_{[0,\infty)}$. Then for $n \geq 0$, $\mathbf{r}_n(\gamma) = \mathbf{r}_n(\gamma^+)$ and

$$\mathbf{D}_\gamma \cap \{0, \ldots, \mathbf{r}_n(\gamma)\} = \mathbf{D}_\gamma \cap \{0, \ldots, \mathbf{r}_n(\gamma^+)\} \subset \mathbf{D}_{\gamma^+} \cap \{0, \ldots, \mathbf{r}_n(\gamma^+)\}$$

since all diamond points of γ are diamond points of γ^+. This implies that

$$\mathbb{P}_{\mathrm{iSAB}}^{\otimes \mathbb{N}} \left(\liminf_{n \to \infty} \frac{|\mathbf{D}_\gamma \cap \{0, \ldots, \mathbf{r}_n(\gamma)\}|}{n} \geq \delta \right)$$
$$= \mathbb{P}_{\mathrm{iSAB}}^{\otimes \mathbb{Z}} \left(\liminf_{n \to \infty} \frac{|\mathbf{D}_{\gamma^+} \cap \{0, \ldots, \mathbf{r}_n(\gamma)\}|}{n} \geq \delta \right) \geq$$
$$\mathbb{P}_{\mathrm{iSAB}}^{\otimes \mathbb{Z}} \left(\liminf_{n \to \infty} \frac{|\mathbf{D}_\gamma \cap \{0, \ldots, \mathbf{r}_n(\gamma)\}|}{n} \geq \delta \right) = 1.$$

This concludes the proof of the proposition. ∎

We are finally ready to complete the proof.

Proof of Theorem A.1. By Lemma A.2, we want to prove that $\mathbb{E}_{\mathrm{iSAB}}(\mathsf{H}(\gamma)) = \infty$. We argue *ad absurdum*. Assume $\mathbb{E}_{\mathrm{iSAB}}(\mathsf{H}(\gamma)) < \infty$ and let $\nu > \mathbb{E}_{\mathrm{iSAB}}(\mathsf{H}(\gamma))$. Also, let $0 < \varepsilon < \delta/20$, where δ satisfies Proposition A.5.

We denote by Ω^+ the set of semi-infinite walks in the upper half-plane. That is, $\phi = (\phi_0, \phi_1, \ldots) \in \Omega^+$ if and only if $y(\phi_i) > 0$ for $i > 0$. For $\phi \in \Omega^+$ and γ a finite bridge, we denote $\gamma \triangleleft \phi$ if $\phi_{[0,|\gamma|]} = \gamma$ and $\phi_{|\gamma|}$ is a renewal point of ϕ. Note that

$$x_{\mathrm{c}}^{|\gamma|} = \mathbb{P}_{\mathrm{iSAB}}^{\otimes \mathbb{N}}(\phi \in \Omega^+ : \gamma \triangleleft \phi). \tag{A.7}$$

Let $\overline{\mathrm{SAB}}_n$ denote the set of finite bridges γ with exactly $n+1$ renewal points (meaning that $\mathbf{r}_n(\gamma) = |\gamma|$) such that

(C$_1$) $\mathsf{H}(\gamma) \leq \nu n$,

(C$_2$) $|\mathbf{D}_\gamma| \geq \delta n/2$.

Let us define $\overline{\mathrm{SAB}}_n^+ = \{\phi \in \Omega^+ : \exists \gamma \in \overline{\mathrm{SAB}}_n \text{ such that } \gamma \triangleleft \phi\}$. That is, the prefix of ϕ consisting of its n first irreducible bridges satisfies (C$_1$) and (C$_2$). It follows from (A.7) that

$$\mathbb{P}_{\mathrm{iSAB}}^{\otimes \mathbb{N}}\left(\overline{\mathrm{SAB}}_n^+\right) = \sum_{\gamma \in \overline{\mathrm{SAB}}_n} x_{\mathrm{c}}^{|\gamma|}. \tag{A.8}$$

We now prove that

$$\mathbb{P}_{\mathrm{iSAB}}^{\otimes \mathbb{N}}\left(\overline{\mathrm{SAB}}_n^+\right) \longrightarrow 1 \text{ as } n \to \infty. \tag{A.9}$$

We consider conditions (C$_1$) and (C$_2$) separately. Condition (C$_1$) for $\gamma \in \overline{\mathrm{SAB}}_n$ translates for $\phi \in \overline{\mathrm{SAB}}_n^+$ into $\mathsf{H}(\phi_{[0, \mathbf{r}_n(\phi)]}) \leq \nu n$. Since $\mathbb{E}_{\mathrm{iSAB}}(\mathsf{H}(\gamma)) < \nu$, the law of large numbers gives

$$\mathbb{P}_{\mathrm{iSAB}}^{\otimes \mathbb{N}}\left(\phi \in \Omega^+ : \mathsf{H}(\phi_{[0, \mathbf{r}_n(\phi)]}) \leq \nu n\right) \longrightarrow 1.$$

Let us now consider condition (C$_2$), which translates into $|\mathbf{D}_{\phi_{[0,\mathbf{r}_n(\phi)]}}| \geq \delta n/2$. But

$$\mathbf{D}_{\phi_{[0,\mathbf{r}_n(\phi)]}} \supset \mathbf{D}_\phi \cap \{0,\ldots,\mathbf{r}_n(\phi)\},$$

since the truncation operation $\phi \to \phi_{[0,\mathbf{r}_n(\phi)]}$ can only create (and not annihilate) diamond points. Thus Proposition A.5 yields

$$\mathbb{P}_{\text{iSAB}}^{\otimes\mathbb{N}}\big(|\mathbf{D}_{\phi_{[0,\mathbf{r}_n(\phi)]}}| \geq \tfrac{\delta}{2}n\big) \longrightarrow 1,$$

and we have proved (A.9).

We are now going to prove that

$$\mathbb{P}_{\text{iSAB}}^{\otimes\mathbb{N}}\Big(\mathsf{W}(\phi_{[0,\mathbf{r}_{vn+1}(\phi)]}) > \varepsilon n\Big) \geq \left(\frac{\delta n x_\text{c}}{10(vn+2)}\right)^2 \mathbb{P}_{\text{iSAB}}^{\otimes\mathbb{N}}\big(\overline{\text{SAB}}_n^+\big). \tag{A.10}$$

Since $\mathsf{W}(\phi_{[0,\mathbf{r}_{vn+1}(\phi)]})/n$ tends to zero $\mathbb{P}_{\text{iSAB}}^{\otimes\mathbb{N}}$-almost surely, as follows from the beginning of the proof of Proposition A.5, this contradicts (A.9) and proves that our assumption $\mathbb{E}_{\text{iSAB}}(\mathsf{H}(\gamma)) < \infty$ cannot hold.

Consider $\gamma \in \overline{\text{SAB}}_n$. Let $\mathbf{d}_i$ be the index of the ith diamond point of γ. For integers $i \in \left[\frac{\delta}{10}n, \frac{2\delta}{10}n\right]$ and $j \in \left[\frac{3\delta}{10}n, \frac{4\delta}{10}n\right]$, let $\mathsf{StickBreak}_{i,j}(\gamma)$ be the following walk (see Figure A.3, right):

$$\mathsf{StickBreak}_{i,j}(\gamma) = \gamma_{[0,\mathbf{d}_i]} \circ \tau \circ \rho\big(\gamma_{[\mathbf{d}_i,\mathbf{d}_j]}\big) \circ \tilde{\tau} \circ \gamma_{[\mathbf{d}_j,\mathbf{r}_n]}, \tag{A.11}$$

where $\circ$ stands for the concatenation of walks, ρ is the clockwise rotation of angle $\pi/3$, τ is a single right turn, and $\tilde{\tau}$ is a left turn. The definition of diamond points implies that $\mathsf{StickBreak}_{i,j}(\gamma)$ is not only self-avoiding, but also a bridge. Also, note that we used (C$_2$) in order to define $\mathsf{StickBreak}(\gamma)$ for all these values of i and j.

Let

$$\Phi = \left[\tfrac{\delta}{10}n, \tfrac{2\delta}{10}n\right] \times \left[\tfrac{3\delta}{10}n, \tfrac{4\delta}{10}n\right] \times \overline{\text{SAB}}_n,$$

and denote

$$S := \sum_{(i,j,\gamma)\in\Phi} x_\text{c}^{|\mathsf{StickBreak}_{i,j}(\gamma)|}.$$

One can express S in terms of $\mathbb{P}_{\text{iSAB}}^{\otimes\mathbb{N}}\big(\overline{\text{SAB}}_n^+\big)$. Indeed, $|\mathsf{StickBreak}_{i,j}(\gamma)| = |\gamma| + 2$, and therefore

$$S = \sum_{(i,j,\gamma)\in\Phi} x_\text{c}^{|\gamma|+2} = \left(\frac{\delta x_\text{c} n}{10}\right)^2 \sum_{\gamma\in\overline{\text{SAB}}_n} x_\text{c}^{|\gamma|} = \left(\frac{\delta x_\text{c} n}{10}\right)^2 \mathbb{P}_{\text{iSAB}}^{\otimes\mathbb{N}}\big(\overline{\text{SAB}}_n^+\big). \tag{A.12}$$

We used (A.8) for the last equality. We are now going to give an upper bound on S, which will imply (A.10).

Note that the walk $\gamma_{[\mathbf{d}_i, \mathbf{d}_j]}$ contains at least $\delta n / 10$ diamond points, and thus has height $h := \mathsf{H}(\gamma_{[\mathbf{d}_i, \mathbf{d}_j]}) \geq \delta n / 10$. Rotating this walk by $\pi/3$ results in a walk of height at most h and width at least $h/2$. Hence $\mathsf{StickBreak}_{i,j}(\gamma)$ has width at least $\delta n / 20 > \varepsilon n$. By $(\mathbf{C}_1)$, we also have $\mathsf{H}(\mathsf{StickBreak}_{i,j}(\gamma)) \leq \nu n + 1$ and therefore $\mathsf{StickBreak}_{i,j}(\gamma)$ has at most $\nu n + 2$ renewal points. Hence, for any $\phi \in \Omega^+$ such that $\mathsf{StickBreak}_{i,j}(\gamma) \lhd \phi$, we have $\mathbf{r}_{\nu n + 1}(\phi) \geq |\mathsf{StickBreak}_{i,j}(\gamma)|$ and therefore $\mathsf{W}(\phi_{[0, \mathbf{r}_{\nu n + 1}(\phi)]}) > \varepsilon n$. Thus, for any $(i, j, \gamma) \in \Phi$,

$$
\begin{aligned}
x_{\mathrm{c}}^{|\mathsf{StickBreak}_{i,j}(\gamma)|} &= \mathbb{P}_{\mathrm{iSAB}}^{\otimes \mathbb{N}}\big(\phi \in \Omega^+ : \mathsf{StickBreak}_{i,j}(\gamma) \lhd \phi\big) \\
&= \mathbb{P}_{\mathrm{iSAB}}^{\otimes \mathbb{N}}\big(\phi \in \Omega^+ : \mathsf{StickBreak}_{i,j}(\gamma) \lhd \phi \text{ and } \mathsf{W}(\phi_{[0, \mathbf{r}_{\nu n + 1}(\phi)]}) > \varepsilon n\big).
\end{aligned}
$$

Therefore,

$$
\begin{aligned}
S &= \sum_{(i,j,\gamma) \in \Phi} \mathbb{P}_{\mathrm{iSAB}}^{\otimes \mathbb{N}}\big(\phi \in \Omega^+ : \mathsf{StickBreak}_{i,j}(\gamma) \lhd \phi \text{ and } \mathsf{W}(\phi_{[0, \mathbf{r}_{\nu n + 1}(\phi)]}) > \varepsilon n\big) \\
&= \mathbb{E}_{\mathrm{iSAB}}^{\otimes \mathbb{N}}\Big(\big|\{(i,j,\gamma) \in \Phi : \mathsf{StickBreak}_{i,j}(\gamma) \lhd \phi\}\big| \cdot \mathbb{1}_{\{\mathsf{W}(\phi_{[0, \mathbf{r}_{\nu n + 1}(\phi)]}) > \varepsilon n\}}\Big) \\
&\leq (\nu n + 2)^2 \, \mathbb{P}_{\mathrm{iSAB}}^{\otimes \mathbb{N}}\big(\mathsf{W}(\phi_{[0, \mathbf{r}_{\nu n + 1}(\phi)]}) > \varepsilon n\big). \qquad\qquad\text{(A.13)}
\end{aligned}
$$

The last inequality follows from the fact that, for any given $\phi \in \Omega^+$, the number of elements (i, j, γ) of Φ such that $\mathsf{StickBreak}_{i,j}(\gamma) \lhd \phi$ is at most $(\nu n + 2)^2$. Indeed, the triple (i, j, γ) is completely determined if we specify in ϕ the renewal point that precedes the step denoted τ in (A.11) and the one that follows the step $\tilde{\tau}$. As both points occur before $\mathbf{r}_{\nu n + 1}$, as explained above, the bound (A.13) follows.

By combining (A.12) and (A.13) we obtain (A.10), which concludes the proof. ∎

Appendix B

Techniques for computer enumeration and series analysis

In this appendix we briefly discuss some of the methods used to obtain the numerical results presented in this thesis. These methods generally fall into one of two categories: algorithms for *series generation* and techniques for *series analysis*.

B.1 Enumeration of self-avoiding walks

There are three main types of enumerative algorithms used in this thesis. The most basic is *backtracking*, where walks or polygons are simply 'grown' one step at a time. Far more advanced (and faster) is the *finite lattice method*, where walks are restricted to a finite geometry and constructed by adding columns (or rows) to the lattice in a systematic way. Finally, models for which we have functional equations can be enumerated even more quickly (i.e. in polynomial time) by encoding these functional equations as recursions.

B.1.1 Backtracking

In a backtracking algorithm, walks of length n are enumerated simply by taking every walk of length $n - 1$ and attempting to add every possible step to the end. The structure of such an algorithm is roughly as follows:

- WALK is an object which tracks the vertices of the lattice which have been visited by a walk. Typically, we know beforehand that our walks will be contained in, say, an $n \times n$ grid, and then WALK would just be an $n \times n$ binary array.

- COUNTER is an object which tracks how many walks we have counted. If we will be counting walks of lengths up to m, then COUNTER might be an $m \times 1$ integer array, where COUNTER(n) would be the number of walks of length n counted thus far.

- N tracks the length of the walk we are currently considering.

- CONDITIONS is a function which will take WALK as an argument, and output YES if the walk satisfies the conditions required by the model being considered (e.g. self-avoidance, quasi-prudence, etc.) or NO if it does not.

(0) Start with the empty walk. Set $N = 0$. Go to (1).

(1) Add an east step to WALK. Set $N = N + 1$. Go to (X), then (2).

(2) Add a north step to WALK. Set $N = N + 1$. Go to (X), then (3).

(3) Add a west step to WALK. Set $N = N + 1$. Go to (X), then (4).

(4) Add a south step to WALK. Set $N = N + 1$. Go to (X).

(X) Evaluate CONDITIONS(WALK).

  - If YES, set COUNTER(N) = COUNTER(N)+1. Go to (1).
  - If NO, remove last step from WALK and set $N = N - 1$.

The running time for such an algorithm to count walks of lengths up to m will be roughly $O(\mu^m)$, where μ is the exponential growth constant for the model under consideration.[1] This means that, in practice, such algorithms are very slow, and for models in two or more dimensions they are only useful for counting walks of lengths up to perhaps 20 or 30. The upshot is that they are very easy to program, and they are thus useful for quickly checking series obtained via other methods. They are also sometimes the only option at our disposal – for example, when enumerating quasi-prudent walks, we have no functional equations, nor any idea how to adapt the finite lattice method.

B.1.2   Finite lattice method

This algorithm was used to produce some of the results of Sections 2.2 and 4.2. The following description is thus specific to the geometries featuring in those sections; namely, finite-width strips of the square, triangular and honeycomb lattices.

The algorithm used to enumerate SAWs on the square lattice builds on the pioneering work of Enting [37] who enumerated square lattice self-avoiding polygons using the finite lattice method. More specifically the algorithm is based in large part on the one devised by Conway, Enting and Guttmann [23] for the enumeration of SAWs. The details can be found in [73]. Below we shall only briefly outline the basics of the algorithm and describe the changes made for the particular problem studied in this work.
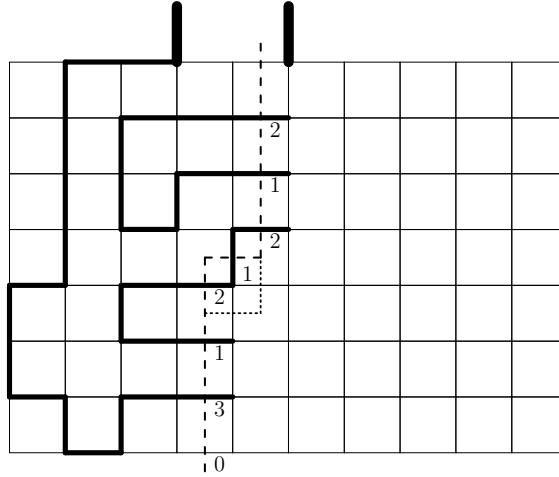
---

[1]Assuming such a constant exists, of course.

Figure B.1: A snapshot of the boundary line (dashed line) during the transfer matrix calculation of type A configurations on a strip of size 7×10. SAWs are enumerated by successive moves of the kink in the boundary line, as exemplified by the position given by the dotted line, so that one vertex and two edges at a time are added to the strip. To the left of the boundary line we have drawn an example of a partially completed SAW. The heavy lines at the top are the incoming and outgoing edges of the SAW.

The generating function for a rectangle was calculated using transfer matrix (TM) techniques. The most efficient implementation of the TM algorithm generally involves bisecting the finite lattice with a boundary (this is just a line in the case of rectangles) and moving the boundary in such a way as to build up the lattice vertex by vertex as illustrated in Figure B.1. If we draw a SAW and then cut it by a line we observe that the partial SAW to the left of this line consists of a number of loops connecting two edges (we shall refer to these as loop ends) in the intersection, and pieces which are connected to only one edge (we call these free ends). The other end of the free piece is an end point of the SAW so there are at most two free ends.

Each end of a loop is assigned one of two labels depending on whether it is the lower end or the upper end of a loop. Each configuration along the boundary line can thus be represented by a set of edge states $\{\sigma_i\}$, where

$$
\sigma_i = \begin{cases}
0 & \text{empty edge,} \\
1 & \text{lower loop-end,} \\
2 & \text{upper loop-end,} \\
3 & \text{free end.}
\end{cases}
\tag{B.1}
$$

If we read from the bottom to the top, the configuration or signature S along the intersection of the partial SAW in Figure B.1 is $S = \{031212120\}$. Since crossings are not permitted this encoding uniquely describes which loop ends are connected.

A few changes to the algorithm described in [73] are required in order to enumerate the

231

restricted SAWs we study here. Most importantly the SAWs must have a free end at the middle vertex of the top side of the strip. This is easily ensured by restricting the updating rules at this vertex (also signatures prior to passing this vertex can have at most one free end). Specifically the middle vertex is reached when the TM boundary has been moved halfway through the strip. At this point the incoming edge to the left of the middle vertex is either empty, an upper loop-end or free. In the *empty* case we have to insert a new free end (along either the horizontal or vertical outgoing edge). In the *upper* case the loop-end is terminated and the matching lower loop-end becomes a free end.

While in the *free* case the end is again terminated and all the edges connected to this free end form a SAW. However this is only a valid configuration if all other edges are empty since otherwise we would form configurations with more than one component. Secondly, in enumerating SAWs of type A (i.e. the walks we have referred to as *loops*) the second free end must lie in the top side of the rectangle; we chose to force the free end to lie to the left of the middle vertex and use symmetry to count all possible configurations. In counting bridges or SAWs of type B the second free end must lie at the bottom of the strip. Thirdly, in [73] the SAWs were forced to span the rectangle, that is touch all sides, but this restriction is lifted in this study.

The sum over all contributing graphs is calculated as the boundary is moved through the lattice. For each configuration of occupied or empty edges along the intersection we maintain a generating function G_S for partial walks with signature S. In exact enumeration studies G_S would be a truncated two-variable polynomial $G_S(x)$ where z is conjugate to the number of steps.

In a TM update each source signature S (before the boundary is moved) gives rise to a few new target signatures S' (after the move of the boundary line) and $n = 0, 1$ or 2 new edges are inserted leading to the update $G_{S'}(x) = G_{S'}(x) + x^n G_S(x)$. Once a signature S has been processed it can be discarded. In most studies the calculations were done using integer arithmetic modulo several primes with the full integer coefficients reconstructed at the end using the Chinese remainder theorem. Here we are not really interested in the exact coefficients. This makes life a little easier for us since we can use real coefficients with the generating functions truncated at some maximal degree M. The calculations were carried out using quadruple (or 128-bit) floating-point precision (achieved in FORTRAN with the REAL(KIND=16) type declaration).

In our calculations we truncated $A_T(x)$ and $B_T(x)$ at degree $M = 1000$ and used strips of half-length $L = M$. These choices of M and L more than suffice to ensure that numerical errors are negligible as evidenced by the fact that when we solve (2.18) to find x_c for the honeycomb lattice the estimate for x_c agrees with the exact value to at least 30 digits, that is, to within the numerical accuracy of the floating-point computation itself.

The computational complexity of the calculation required to obtain the number of walks in a strip of width T and length L can be easily estimated. Time (and memory) requirements are basically proportional to a polynomial in M and L times the maximal number of signatures, N_{Conf},

generated during the calculation. It is well established [72] that $N_{\mathrm{Conf}} \propto 3^T$ so the algorithm has exponential computational complexity.

The transfer-matrix algorithm is eminently suited to parallel computations and here we used the approach first described in [72] and refer the interested reader to this publication for further detail. The bulk of the calculations for this paper were performed on the cluster of the NCI National Facility, which provides a peak computing facility to researchers in Australia. The NCI peak facility is a Sun Constellation Cluster with 1492 nodes in Sun X6275 blades, each containing two quad-core 2.93GHz Intel Nehalem CPUs with most nodes having 3GB of memory per core (24GB per node). It took a total of about 1800 CPU hours to calculate $A_T(x)$ for T up to 15. So, the bulk of the time (almost 1250 hours) was spent calculating $A_{15}(x)$. In this case we used 48 processors and the split between actual calculations and communications was roughly 2 to 1 (with quite a bit of variation from processor to processor). Smaller widths can be done more efficiently in that communication needs are fewer and hence not as much time is used for this task.

On a technical issue we note that quad precision is not a supported data type in the MPI standard. So in order to pass messages containing the generating functions we used the MPI data type MPI-BYTE with each coefficient then having a length of 16 bytes.

The algorithm used for the triangular lattice is quite similar. The triangular lattice is represented as a square lattice with additional edges along one of the main diagonals. This poses an immediate problem since a boundary line drawn as in Figure B.1 would intersect $2T$ edges thus greatly increasing the number of possible signatures. In this case it is more efficient to draw the boundary line through the vertices of the lattice. We then again have T intersections, however a vertex may be in an additional state since a partial SAW can touch the boundary line without crossing it (see [75] for further details). The upshot is that the computational complexity grows exponentially as 4^T.

Computational complexity

The time $T(W)$ required to obtain the number of walks of in a strip of width W can be easily estimated. Time (and memory) requirements are basically proportional to a polynomial (in M) times the maximal number of configurations, N_{Conf}, generated during the calculation. When the boundary line is straight and intersects $W + 1$ edges we can find the exact number of configurations. First look at the situation with no free ends. The configurations correspond to Motzkin paths [26] (just map 0 to a horizontal step, 1 to a north-east step, and 2 to a south-east step) and the number of such paths M_n with n steps is easily derived from the generating function

$$M(x) = \sum_n M_n x^x = \frac{1 - x - \sqrt{(1+x)(1-3x)}}{2x^2}. \tag{B.2}$$

The number of transfer matrix configurations $N_{\mathrm{Conf}}(W)$ is simply obtained by inserting 0, 1 or 2 free ends into a Motzkin path and eliminating the path corresponding to a configuration of all

0's, hence

$$N_{\text{Conf}}(W) = M_{W+1} + (W+1)M_W + \frac{(W+1)WM_{W-1}}{2} - 1. \qquad \text{(B.3)}$$

With a kink in the boundary line the number of configurations lies between $N_{\text{Conf}}(W)$ and $N_{\text{Conf}}(W+1)$. From (B.2) it is clear that $N_{\text{Conf}}(W) \propto 3^W$.

Parallelisation

The transfer-matrix algorithm is eminently suited for parallel computations. The most basic concern in any efficient parallel algorithm is to minimise the communication between processors and ensure that each processor does the same amount of work and uses the same amount of memory. In practice one naturally has to strike some compromise and accept a certain degree of variation across the processors.

One of the main ways of achieving a good parallel algorithm using data decomposition is to try to find an invariant under the operation of the updating rules. That is we seek to find some property of the signature which does not alter in a single iteration. The algorithm for the enumeration of SAWs is quite complicated since an update at a vertex might change the state of an edge far removed, for example when two lower loop ends are joined we have to relabel one of the associated upper loop ends as a lower loop end in the new signature. However, there is still an invariant since any edge not directly involved in the update cannot change from being empty to being occupied and vice versa. That is only the kink edges can change their occupation status. This invariant allows us to parallelise the algorithm in such a way that we can do the calculation completely independently on each processor with just two redistributions of the data set each time an extra column is added to the strip. This scheme was first used for enumerating SAPs [72], and we refer the interested reader to this publications for further detail. For the larger widths the configuration numbers are too small to ensure a decent balance with just two redistributions (in fact some CPUs would be almost empty while others would exceed the memory limit per CPU). Redistributing the configurations more often increases the length of the invariant part of the signature and hence the number of configurations available for the redistribution. Naturally there is a time penalty to be paid. For larger widths we used 3 or 4 redistributions per column.

Including surface interactions

The algorithms we use to enumerate SAWs interacting with a surface on the honeycomb, square and triangular lattices builds on the algorithm outlined earlier in this subsection, and detailed descriptions can be found in these papers [73, 75, 76]. Suffice to say that the generating functions for a given strip were calculated using transfer matrix (TM) techniques.

The sum over all contributing graphs is calculated as the boundary is moved through the lattice. For each configuration of occupied or empty edges along the intersection we maintain

a generating function G_S for partial walks with configuration S. In exact enumeration studies G_S would be a truncated two-variable polynomial $G_S(x, y)$ where x is conjugate to the number of steps and y to the number of surface-contacts (sites or edges). In a TM update each source configuration S (before the boundary is moved) gives rise to a few new target configurations S' (after the move of the boundary line) and $n = 0, 1$ or 2 new edges and $m = 0$ or 1 new contacts are inserted leading to the update $G_{S'}(x, y) = G_{S'}(x, y) + x^n y^m G_S(x, y)$. Here we are primarily interested in the case where $A(x, y)$ or $B(x, y)$ are evaluated at the critical point $x = x_c$. This actually makes life easier for us since we can change to a single variable generating function $G_S(y)$ and update signatures as $G_{S'}(y) = G_{S'}(y) + x_c^n y^m G_S(y)$. Here $G_S(y)$ is a polynomial in the contact fugacity y with real coefficients truncated at some maximal degree M. The calculations were carried out using quadruple (or 128-bit) floating-point precision (achieved in FORTRAN with the REAL(KIND=16) type declaration).

In our calculations we truncated $A(x_c, y)$ at degree $M = 1000$ and used strips of half-length $L = M$. In Table B.1 we have listed estimates for $y_c^*(9)$ obtained from strips of width 9 and 10 (the crossing between $A_9(x_c, y)$ and $A_{10}(x_c, y)$) for various values of M and L. Clearly the choice $M = L = 1000$ suffices to estimate $y^*(9)$ to more than 10 digits accuracy.

Table B.1: The estimated value of $y^*(9)$ for the square lattice surface vertex model, truncated at degree M and using strips of half-length from M up to $10M$.

M	$L = M$	$L = 2M$	$L = 5M$	$L = 10M$
100	1.832547814756	1.778376701255	1.778024722094	1.778024722094
250	1.776250937231	1.775990603337	1.775990594686	1.775990594686
500	1.775990340341	1.775990291271	1.775990291271	
1000	1.775990291271			

The transfer-matrix algorithm is eminently suited for parallel computations and here we used the approach first described in [72] and refer the interested reader to this publication for further detail. The bulk of the calculations for this paper were performed on the cluster of the NCI National Facility, which provides a peak computing facility to researchers in Australia. The NCI peak facility is a Sun Constellation Cluster with 1492 nodes in Sun X6275 blades, each containing two quad-core 2.93GHz Intel Nehalem CPUs with most nodes having 3GB of memory per core (24GB per node). It took a total of about 3300 CPU hours to calculate $A_T(x_c, y)$ for T up to 15 on the square lattice. It is known [73] that the time and memory required to obtain the number of walks in a strip of width T grows exponentially as 3^T for the honeycomb and square lattices and as 4^T for the triangular lattice. So, the bulk of the time was spent calculating A_{15} and B_{15}, which amounted to almost 2300 hours in the square lattice case. In this case we used 48 processors and the split between actual calculations and communications

was roughly 2 to 1 (with quite a bit a variation from processor to processor). Smaller widths can be done more efficiently in that communication needs are lesser and hence not as much time is used for this task.

B.1.3   Recursions from functional equations

In the cases where we are able to derive a functional equation satisfied by the generating function of the objects in question, it is usually possible to generate long series much more quickly than would be possible with backtracking or the finite lattice method. We will illustrate this with the relatively simple example of 2-sided perimeter walks. Recall the functional equation (3.100) satisfied by the generating function of walks which end on the east side of their box:

$$
\left(1 - tv - \frac{t^2 v}{u - t} - \frac{t^2 uv}{1 - tu}\right) R(t; u, v) = \frac{1}{1 - tu} - \frac{tuv}{u - t} R(t; t, v)
$$
$$
+ \frac{t(2 - tu)}{1 - tu} R(t; t, 1) - \frac{t^2 u}{1 - tu} R(t; t, tu) + \frac{t^2 u(1 - v)}{1 - tu} R(t; 0, 1),
$$

where in $R(t; u, v)$ the variable t is conjugate to the length of a walk, u is conjugate to the distance from the endpoint of a walk to the north-east corner of its box and v is conjugate to the distance from the north-east corner of the box to the nearest occupied vertex on the north side of the box.

This equation was obtained by grouping walks according to the location and direction of their *last inflating step* – the last step which moved the north or east sides of the bounding box. Here, we are essentially implementing the same method, but writing it purely in terms of the coefficients $r_{n,i,j}$. The notation $X += Y$ means that we set $X = X + Y$.

- R will be a three-dimensional array which tracks the coefficients $r_{n,i,j}$ as we compute them, indexed by $[n, i, j]$. If the greatest length we wish to compute is n_{max}, then restricting R to $0 \le n, i, j \le n_{max}$ is sufficient. Initially we fill R with zeros.

- First add the walks with no inflating steps: i.e. set $R[n, n, 0] = 1$ for $0 \le n \le n_{max}$.

- Then loop over n, i, j: Taking $0 \le n \le n_{max} - 1$, $0 \le i \le n$ and $0 \le j \le n - i$ is sufficient.

  - If $R[n, i, j] = 0$, skip to the next iteration of the loop.
  - If $R[n, i, j] > 0$, we can add an inflating step east. Then,
    * North steps: For $1 \le l \le \min\{i - 1, n_{max} - n - 1\}$,

    $$R[n + l + 1, i - l, j + 1] += R[n, i, j].$$

    If $1 \le i \le n_{max} - n - 1$, then also

    $$R[n + i + 1, 0, 0] += R[n, i, j].$$

* South steps: If $i = 0$, then for $0 \leq l \leq n_{max} - n - 1$,

$$R[n+l+1, l, 0] \mathrel{+}= R[n, i, j].$$

If $i > 0$, then for $0 \leq l \leq n_{max} - n - 1$,

$$R[n+l+1, l+i, j+1] \mathrel{+}= R[n, i, j].$$

- If $R[n, i, j] > 0$ and $i \leq n_{max} - n - 1$, we can reflect the walk through $y = x$ and add an inflating step north and i steps east, then, for $0 \leq l \leq \min\{j, n_{max} - n - i - 1\}$,

$$R[n+i+l+1, l, 0] \mathrel{+}= R[n, i, j].$$

- At the end, for each n we sum $R[n, i, j]$ over all i and j, doubling the terms with $i = 0$. The result will be the number of 2-sided perimeter walks of length n.

It can be easily seen that the running time of algorithms such as this will be polynomial in n_{max}; in this case, the running time is $O(n_{max}^4)$.

B.2   Series analysis

In this section we will briefly discuss two types of methods used in this thesis for analysing series of coefficients. The first and most rudimentary is generally known as the *ratio method*, and was first used by Sykes in 1951 (see [55]). The second and far more advanced is the method of *differential approximants*, first developed by Guttmann and Joyce [56] in 1972.

B.2.1   Ratio method

Recall from Chapter 1 that the number c_n of SAWs of length n is known to satisfy

$$c_n \sim \exp(\varkappa n + o(n)),$$

with $\varkappa$ a lattice-dependent constant. Though it has not been proven, it is generally expected that c_n satisfies

$$c_n \sim A n^{\gamma - 1} \mu^n, \tag{B.4}$$

with A, γ and $\mu = e^\varkappa$ constant. Nor has it been proven that

$$\lim_{n \to \infty} \frac{c_{n+1}}{c_n} = \mu,$$

though clearly if this limit does exist it must be equal to μ.

This gives us a particularly simple way of using the first terms in the sequence $\{c_n\}$ to search for an estimate for μ: compute the sequence $\{c_{n+1}/c_n\}$, and attempt to determine the limit as $n \to \infty$ (if it appears that such a limit exists). This will apply equally well to any sequence a_n
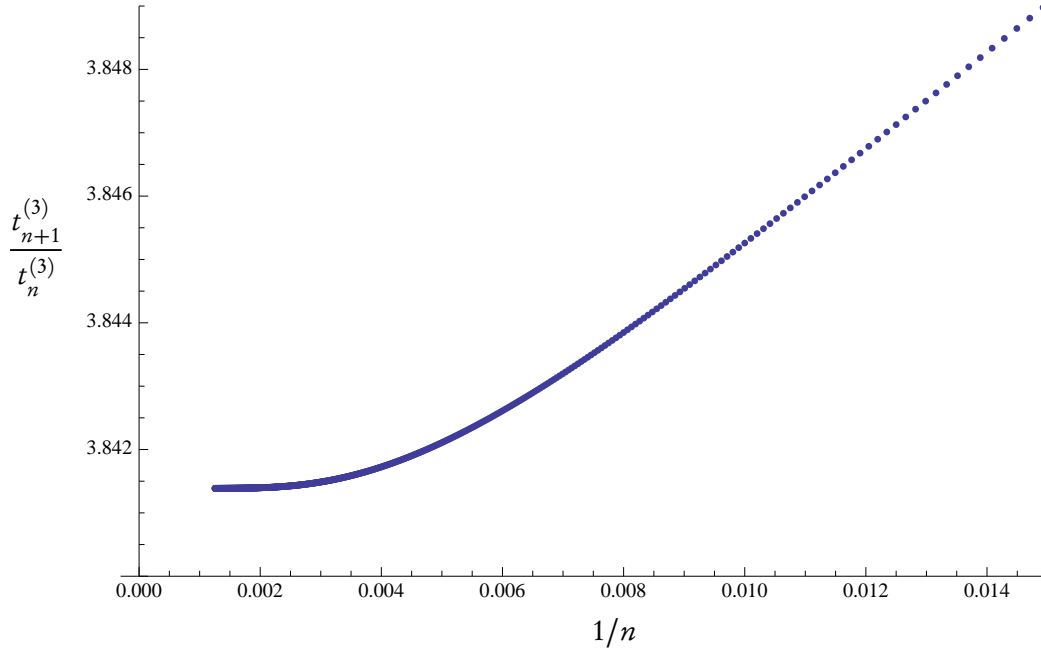
Figure B.2: Plot of the sequence $\{t^{(3)}_{n+1}/t^{(3)}_n\}$ against $1/n$ up to $n = 800$.

whose asymptotic behaviour is expected to resemble (B.4). For example, we can use the method explained in Subsection B.1.3 to generate 800 terms in the sequence $\{t^{(3)}_n\}$ of 3-sided prudent walks on the triangular lattice. In Figure B.2 we plot the sequence $\{t^{(3)}_{n+1}/t^{(3)}_n\}$ against $1/n$, and we can estimate that the sequence is approaching the limit 3.8413871.

The ratio method is not just limited to estimates for the growth rate μ, however. If we expect (B.4) to be the asymptotic form of our coefficients, then it follows that

$$\frac{c_{n+1}}{c_n} \sim \mu \left(1 + \frac{\gamma - 1}{n} + o\left(\frac{1}{n}\right) \right).$$

Thus, if we plot c_{n+1}/c_n against $1/n$, and we are able to ignore the $o(1/n)$ correction term (which should be appropriate for large n), then we would expect the plot to become linear in $1/n$, with gradient $\mu(\gamma - 1)$. In Figure B.2 we can observe that as n increases, the gradient of the curve is approaching 0, suggesting that $\gamma = 1$ for 3-sided prudent walks on the triangular lattice (and thus that the dominant singularity of $T^{(3)}(t)$ is a simple pole).

B.2.2 Differential approximants

This method is much more sophisticated than the ratio method, but as such its implementation is significantly more involved. Since it has been in use for decades, we will only give a very brief overview here, and direct interested readers to other references [56, 52] for more details.

The basic idea of the method of differential approximants is to find approximations of an unknown generating function $F(x)$ by searching for solutions to differential equations with

polynomial coefficients. Since the singular behaviour of ODEs is a well-studied problem, it is straightforward to then extract estimates for the singularities and corresponding exponents of $F(x)$.

An Mth-order differential approximant to a generating function $F(x)$ is calculated by finding polynomials $Q_k(x)$ and $P(x)$, of degrees N_k and L respectively, which satisfy

$$\sum_{k=0}^{M} Q_k(x) \left(x \frac{d}{dx} \right)^k \tilde{F}(x) = P(x),$$

where $\tilde{F}(x)$ is a function whose first $N = L + \sum_k (N_k + 1)$ series coefficients match those of $F(x)$. In order to find a solution it is necessary to specify one of the coefficients of the Q_k or P, since as it is written now there are $N + 1$ unknown coefficients. It is customary to set $Q_M(0) = 1$.

The theory of ODEs then states that the singularities of $\tilde{F}(x)$ are found among the N_M zeros of $Q_M(x)$, and these are then estimates of the singularities of $F(x)$. There are standard methods for computing the exponents of these singularities; for example, if x_i is a single root of Q_M then the exponent is

$$\lambda_i = M - 1 - \frac{Q_{M-1}(x_i)}{x_i Q_M'(x_i)},$$

so that

$$\tilde{F}(x) \underset{x \to x_i}{\sim} A(1 - x/x_i)^{\lambda_i}$$

for a constant A.