Markov models for Nucleosome dynamics during Transcription: Breathing and Sliding

Opheusden, S.C.F. van

S.C.F van Opheusden

Markov models for Nucleosome dynamics during Transcription: Breathing and Sliding

Bachelor thesis, September 10, 2010

Advisors: prof. F. Redig, prof. H. Schiessel

Mathematisch Instituut, Lorentz Institute for Theoretical Physics

Universiteit Leiden

Contents

1 Abstract

We study the dynamics of nucleosomes, DNA-wrapped proteins, along a DNA chain. First we show that a single nucleosome makes a simple symmetric random walk with respect to the DNA sequence. To obtain an estimate for the diffusion coefficient, we study a specific random walk in the quarter plane, absorbed by its boundary. There are exact results in two limiting cases, and in general we derive a continuum approximation. Then we show that a DNA chain filled with multiple nucleosomes cannot be transcribed by RNA polymerase if there is only hard-core interaction between the polymerase and the nucleosome. In the end, we suggest an alternative interaction between RNA polymerase and nucleosomes which allows the DNA to be transcribed without the help of other proteins.

2 Preface

This thesis is the result of my bachelor project in physics and mathematics. It is written for an audience of mathematicians and physicists of at least bachelor level. I assume the reader to be familiar with basic analytical tools, and to have some background knowledge of biophysics and probability theory. In fact, I use quite a lot of statistical physics and theory of stochastic processes without much explanation. For readers who are not familiar with these subjects, I recommend the very well-written books of Liggett [9], Spitzer [12] and van Kampen [7].

3 Introduction

3.1 DNA and nucleosomes

DNA is the key to life. All of the genetic information about an organism is stored on one or a couple of DNA molecules. These DNA molecules consist of two polymers of sugar and phosphate groups, wrapped together in a double helix structure. Each sugar group is attached to a base, and the hydrogen bonds between these bases is what keeps the two polymers together. There are four possible bases that can be bound to this sugar group: guanine (G), adenine (A), thymine (T), and cytosine (C). A thymine will only pair with a cytosine, and thymine only with adenine. The bases on both strands are exactly matched, so that each pair of adjacent bases is compatible with each other. These matched pairs are called **base pairs**.

A standard human DNA molecule is very long, consisting of 10^7 or 10^8 base pairs. A normal polymer of this length will form a blob with a diameter of about $100\mu m$, whereas the diameter of a cell nucleus will not exceed $10\mu m$. The DNA has to fit inside the nucleus, so there has to be some compaction mechanism which reduces the size of the DNA coil. Actually, there are many forms of compaction on different length scales, but we will focus on the very first, the **nucleosomes**.

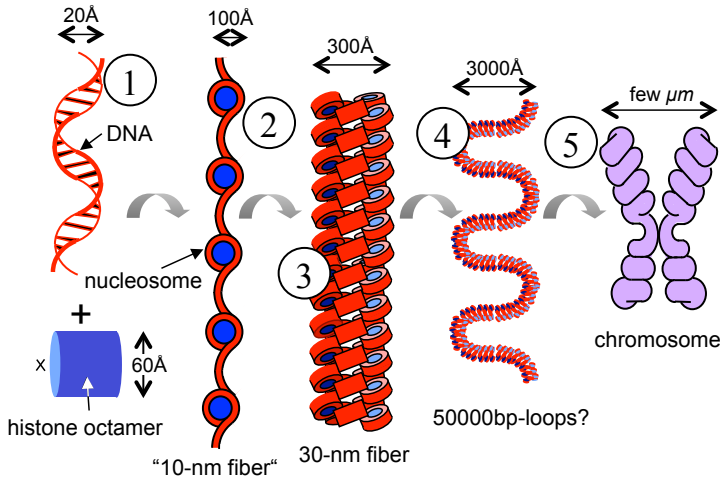

Figure 1: Compaction of DNA into chromatin [10]. We focus on the first level of compaction, the nucleosomes.

A nucleosome consists of a histone octamer, which is strongly bound to a piece of DNA (Details about the molecular structure and dynamics of nucleosomes can be found in [14]). This binding causes the DNA to wrap almost twice around the nucleosome core particle. This comes at a cost, however, as the DNA has to bend very sharply in order to wrap. An estimate for the bending energy can be obtained from the worm like chain model (WLC). As it turns out, the bending energy is just a little bit lower than the binding energy.

3.2 Breathing

When we zoom in on a nucleosome, we see that there are only fourteen points where the DNA actually makes contact with the octamer. At each of these fourteen **binding sites**, the DNA and the octamer are held together by hydrogen bonds as well as electrostatic attraction. Because the energy gained by establishing this bond is higher than the energy required to bend the DNA, the DNA will be wrapped. That is, if the system is in its lowest energy state. But there are always thermal fluctuations which put the system out of that ground state.

Let us now take the thermal fluctuations of the system into account. Suppose the DNA is fully wrapped around the nucleosome. All of the binding sites are very stable, except the two on the outer ends. If any of those bonds would break, the DNA would immediately straighten, and a lot of energy would be gained. Once the first binding site has opened up, the second binding site has a chance to open, and so on. This wrapping and unwrapping process is called **breathing**. In principle, the breathing of the nucleosome could cause the DNA

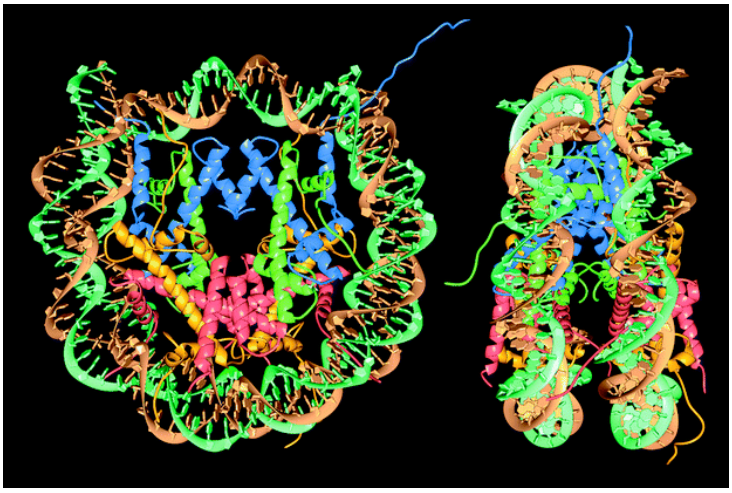

Figure 2: The crystal structure of a nucleosome [14].

to completely detach. However, breaking any bond will always cost more energy than it will gain, so wrapping is always favorable to unwrapping. Furthermore, the DNA is negatively charged, so the two turns of DNA repel each other. Once one of the turns has unwrapped, this repulsion won't be present, and the unwrapping process will be much slower. In effect, we can say that the probability for a nucleosome to fall off the DNA chain is negligible. This agrees with experiments conducted by Polach and Widom [22], which measure the probability for a given binding site to be open at a specific time. A dynamical study of the breathing rates is performed by Koopmans and van Noort [19].

3.3 Sliding

Apart from breathing, there are other thermal fluctuations that affect the nucleosome. If a nucleosome is fully wrapped, there are 147 base pairs associated with the nucleosome. But not all of these base pairs can directly attach to the nucleosome, as there are only fourteen binding sites. In order to minimize the bending energy, the DNA binds to the nucleosome every $10bp$, and the last $10bp$ on either end are essentially straight. But again, this is only the ground state. It could be that the DNA segments between two consecutive binding sites are 11 or 9 base pairs. These disturbances are called **defects** and **antidefects**, respectively. The defects and antidefects can only form at the ends of the nucleosome, and from the WLC it follows that the energy needed to form a defect or antidefect is equal (See also [23]).

When a stretch of DNA between two binding sites has a defect (or antidefect), the tension can resolve by moving either one of its binding sites by 1 bp. This causes a defect to appear in the neighboring stretch. But if the binding

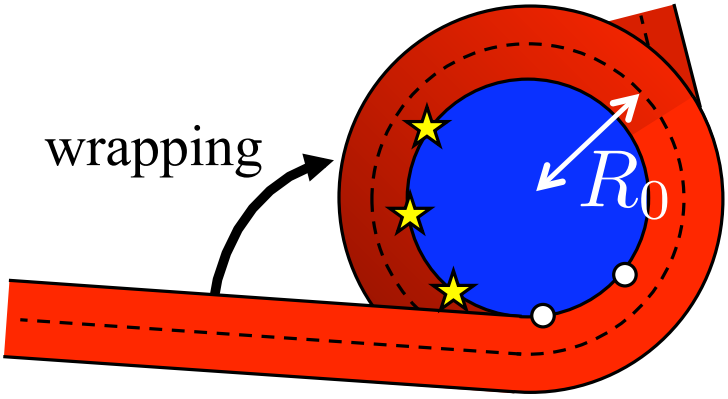

Figure 3: A schematic version of the breathing process [10]. The outer binding sites open and the length of DNA attached to the nucleosome decreases.

sites happens to be the last one on the nucleosome, the defect simply disappears. Suppose now a defect is generated at one end of the nucleosome, then moves through the structure, and eventually falls off at the other end. Then the nucleosome has effectively shifted $1bp$ with respect to DNA molecule. This process is called **sliding**. It has been verified experimentally that nucleosomes move in this way along the DNA [16]. Because the defects and antidefects are generated at a very low rate, there will almost never be more than one defect (or antidefect) in the structure.

3.4 RNA transcription

Until now, we have discussed the dynamics of the DNA and the nucleosome, but we haven't said anything about the purpose of the DNA. Actually, the DNA is just a very sophisticated storage device. The important part is the information stored on the DNA, the specific sequence of base pairs. This DNA sequence is read off by a polymerase molecule, which translates it into RNA.

At this point, there is a problem. How is it possible that an RNA polymerase can read the information on the DNA, if that DNA is attached to a nucleosome? Both the polymerase and the nucleosome are large proteins, and they cannot move through each other. As it happens, the breathing and sliding of the nucleosomes are crucial to answering this question.

In the first part of this thesis, we will focus on the dynamics of a single nucleosome attached to a DNA chain, and ignore its interaction with any other proteins. We will discover that the nucleosome makes a simple symmetric random walk with respect to the underlying DNA sequence, and we try to calculate the diffusion coefficient of this random walk. Then, in the second part, we will zoom out a bit, and look at the large scale behaviour of a DNA chain filled with many nucleosomes, using a simplified model of the single nucleosome dynamics.

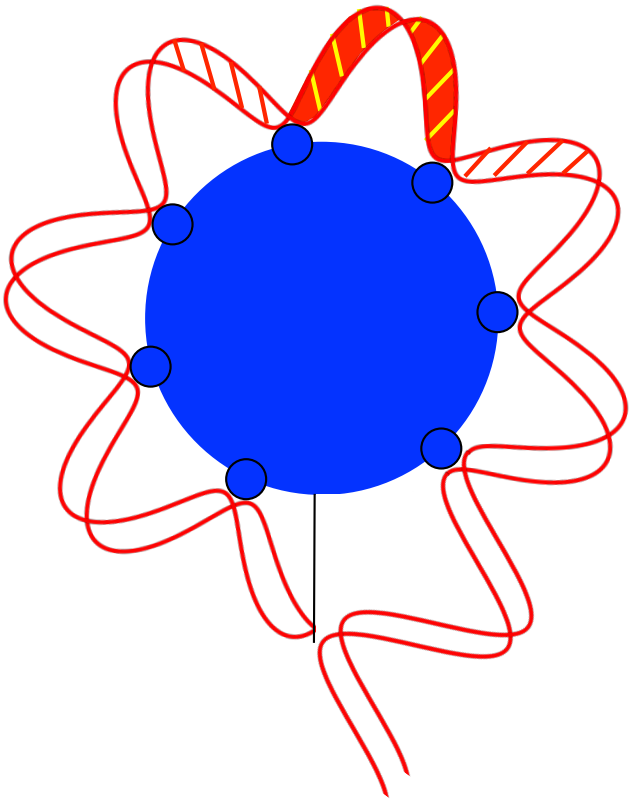

Figure 4: The sliding process [10]. A defect is created at one of the ends, moves to the other end, and then disappears. As a result, the nucleosome has shifted with respect to the DNA sequence.

4 Single nucleosome dynamics

We consider a single nucleosome with a DNA chain wrapped around it. The relative position of this nucleosome with respect to the DNA sequence can only change through the motion of defects and antidefects. Remember that there will never exist more than one defect or antidefect simultaneously. This (anti)defect makes a complicated random walk, but its effect on the position of the nucleosome is determined by three factors only: whether it is a defect or an antidefect, where it enters, and where it exits. If the entrance site and exit site are the same, there is no effect. If a defect goes from right to left, the nucleosome moves forward. If it goes from left to right on the other hand, the nucleosome moves backwards. For an antidefect the effect is exactly the opposite.

Defects and antidefects are generated at the same rate at both ends, so for a single one, there is a probability of $\frac{1}{2}$ that it is a defect, and $\frac{1}{2}$ that it is

an antidefect. Also, the probability that it enters at the right is equal to the probability that it enters at the left. All of this is independent of the exact details of the internal random walk. However, the probability that the exit site and entrance site are different does depend on the details of the random walk.

4.1 Simplifications

In order to do any mathematical analysis, we have to consider a simplified model. We assume that all binding sites are equally strong, and that the binding and bending energies are independent of the underlying DNA sequence. Of course, this is not true. The effect of the DNA sequence on the bending energy can actually be quite strong (there exists a specific DNA sequence, the Widom-601 sequence with an affinity multiple orders of magnitude higher than random DNA [21]). But, if we neglect any DNA sequence effects, the situation becomes highly symmetrical, as there is no real difference anymore between defects and antidefects. Even more, it does not matter where the defect or antidefect enters. Therefore, we may assume without loss of generality that we are dealing with a defect, which enters at the left. This defect makes a simple symmetric random walk through the nucleosome, and we want to know the probability that it exits at the right.

Meanwhile, the nucleosome is also breathing. At both ends, the nucleosome will unwrap some of its binding sites. Let us assume -incorrectly, of course- that the rate at which the outer ends unwrap and rewrap are the same, independent of how much have already unwrapped. This means that the nucleosome can completely disassemble, but that does not pose a real problem. The sliding process is observed to be much faster than the breathing, and we only look at a single defect moving through. This defect will fall off long before the nucleosome detaches from the DNA.

4.2 Markov model

Let us start by labeling the segments between consecutive binding sites of the nucleosome with the numbers 1 to 13. Then we can describe the entire state of a nucleosome with a defect in one of its loops by three parameters: the most unwrapped loop from the left (a_t), from the right (b_t), and the position of the defect (D_t). If there are no loops unwrapped from the left, we put $a_t = 0$, and we set $b_t = 14$ if this happens at the right. Then each of the numbers a_t, b_t and D_t makes a simple symmetric random walk on the set $\{0, 1, \ldots, 14\}$, independent of each other. In the beginning, $a_0 = 0, D_0 = 1, b_0 = 14$, and the process ends whenever $a_t = D_t$ or $D_t = b_t$. Let us say that a_t and b_t move with rate λ, and D_t with rate μ. It will be convenient to normalise λ and μ so that

$$4\lambda + 2\mu = 1. \tag{1}$$

The dynamics of the defect are modeled with the following Markov process.

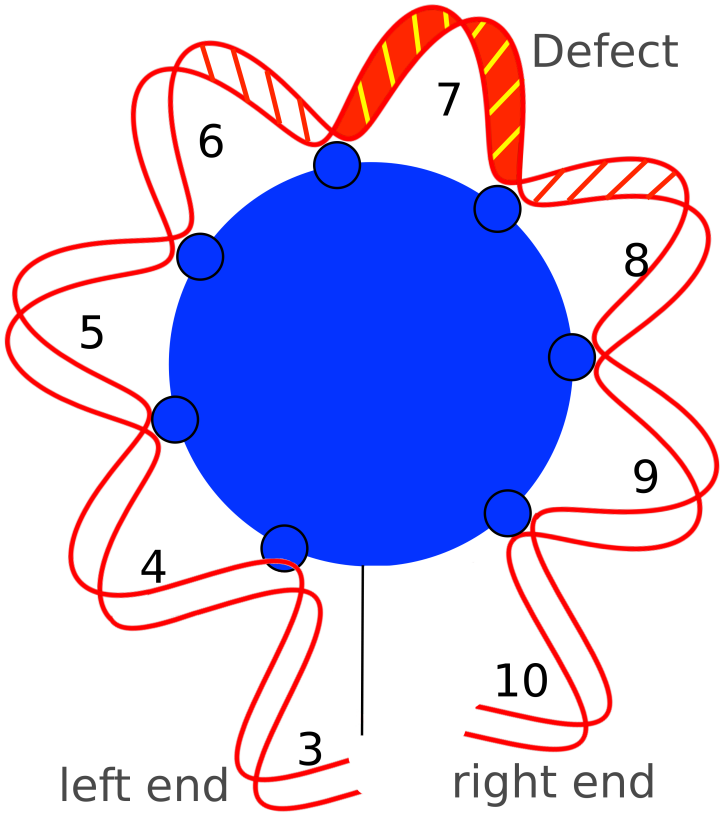

Figure 5: The definition of a_t, b_t and D_t. In this case, $a_t = 3$, $b_t = 10$ and $D_t = 7$.

Definition 1. *Let $N = 14$, and consider the continuous time Markov chain with state space $\{(a, D, b) \in \mathbb{Z}^3 : 0 \le a \le D \le b \le N\}$ and transition rates*

$$\begin{array}{ll} a \to a+1 & \lambda \\ a \to a-1 & \lambda \\ b \to b+1 & \lambda \\ b \to b-1 & \lambda \\ D \to D+1 & \mu \\ D \to D-1 & \mu \end{array}$$

for all $0 < a < D < b < N$. If $a = 0$ or $b = N$, the jumps to $a = -1$ and $b = N + 1$ are prohibited. Now set $\tau := \inf\{t \ge 0 : a_t = D_t \vee D_t = b_t\}$, and define

$$f(a, D, b) := P(b_\tau = D_\tau | a_0 = a, D_0 = D, b_0 = b). \tag{2}$$

Remark that f depends also on μ, λ and N, but we suppress this dependence as these are model parameters.

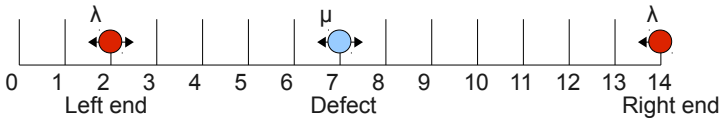

Figure 6: The Markov process: Both ends make a symmetric random walk with rate λ, the defect moves with rate μ. The process ends whenever the defect hits either end.

4.3 Infinite lattice approximation

This process turns out to be too difficult to calculate in full detail (except when $\lambda = 0$, when it is trivial), so we start with a rather crude approximation. We ignore the fact that the nucleosome has only 14 binding sites. This means that the ends a_t and b_t can diffuse away to $\pm\infty$. But the ends move very slowly, so they will not drift too far out. The upshot is that now the relative positions $x_t = D_t - a_t$ and $y_t = b_t - D_t$ perform a translation invariant random walk. This

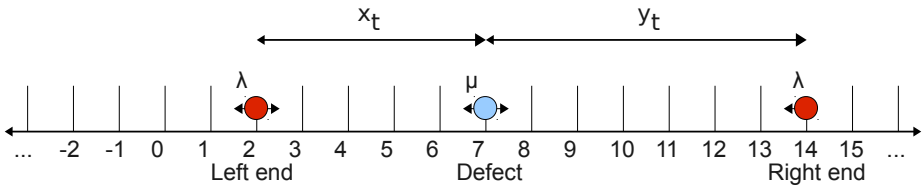

Figure 7: The process is extended to an infinite lattice. The state of the process is then determined by x_t and y_t.

random walk has state space $\mathbb{Z}^2_+ := \{(x, y) \in \mathbb{Z}^2 : x \geq 0, y \geq 0\}$ and transition rates:

$$\begin{aligned}
(x, y) &\to (x, y + 1) && \lambda \\
(x, y) &\to (x, y - 1) && \lambda \\
(x, y) &\to (x + 1, y) && \lambda \\
(x, y) &\to (x - 1, y) && \lambda \\
(x, y) &\to (x - 1, y + 1) && \mu \\
(x, y) &\to (x + 1, y - 1) && \mu
\end{aligned}$$

It starts at a position $(x, y) = (1, N-1)$, and it ends whenever $x_t = 0$ or $y_t = 0$. Then we have $\tau = \inf\{t \geq 0 : x_t = 0 \vee y_t = 0\}$, and

$$f(x, y) = P(y_\tau = 0 | y_0 = y, x_0 = x). \tag{3}$$

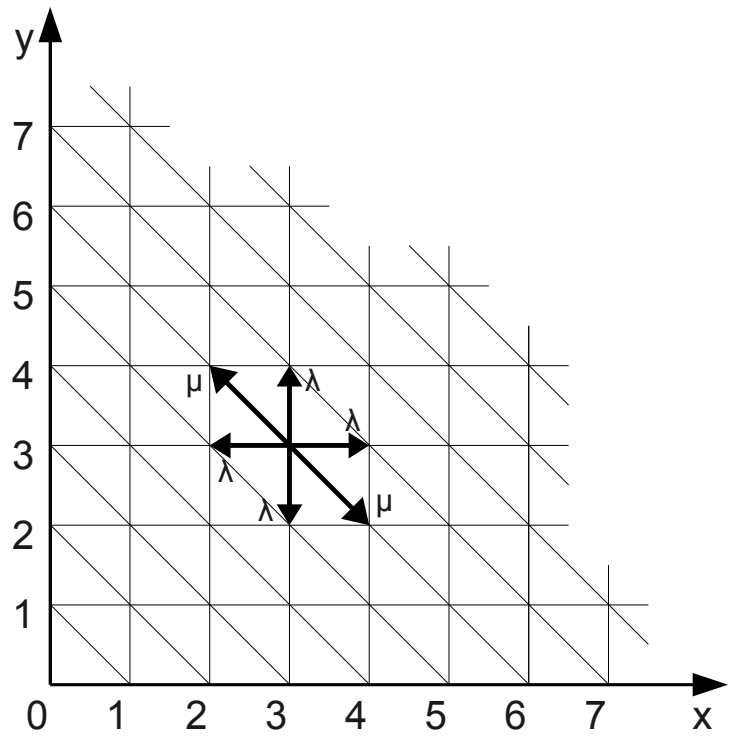

Figure 8: The transition probabilities of (x_t, y_t).

Now that the model is accurately described, we begin with the mathematical analysis. The main points of the following sections are summarized in theorem 1, theorem 4, eqn. 44, and theorem 8.

4.3.1 Asymptotic relations

In the infinite lattice approximation, the starting point is not really special. The process will always start with $x = 1$, but the value of y can just as easily be generalized to be any positive integer. We will now show that $\lim_{n\to\infty} f(1, n) = 0$, and we find an asymptotic relation for $f(1, n)$. But before we get to that, there are some basic results we will need later on.

Lemma 1. *The probability function $f(x, y)$ satisfies the following relations for all $x, y > 0$:*

(a) $f(x+1,y) \geq f(x,y) \geq f(x,y+1)$,

(b) $f(x,y) + f(y,x) = 1$,

(c) $f(0,y) = 0$,

(d) $f(x,0) = 1$,

(e) $f(x,y) = \lambda[f(x,y+1) + f(x,y-1) + f(x-1,y) + f(x+1,y)]$
$+ \mu[f(x+1,y-1) + f(x-1,y+1)]$,

Proof.

(a) Because the random walk has only nearest neighbour jumps, any path leading from the starting point $(1, N-1)$ to the x-axis will cross the line $y = 1$. So if we define $\tau' := \inf\{t \geq 0 : x_t = 0 \vee y_t = 1\}$, then $\tau' \leq \tau$ almost surely. Furthermore, if $x_{\tau'} = 0$, then $x_\tau = 0$. Therefore:

$$P(y_\tau = 0) = P(y_\tau = 0 | y_{\tau'} = 1) P(y_{\tau'} = 1) \leq P(y_{\tau'} = 1). \tag{4}$$

For a random walk started at $(x, y+1)$, this means:

$$\begin{aligned} f(x,y+1) &= P(y_\tau = 0 | y_0 = y+1, x_0 = x) \\ &\leq P(y_{\tau'} = 1 | y_0 = y+1, x_0 = x) \\ &= P(y_\tau = 0 | y_0 = y, x_0 = x) = f(x,y). \end{aligned} \tag{5}$$

The other inequality can be proven in the same way.

(b) Because the transition probabilities are symmetric, we have

$$\begin{aligned} f(j,i) &= P\{y_\tau = 0 | x_0 = j, y_0 = i\} = P\{x_\tau = 0 | y_0 = j, x_0 = i\} \\ &= 1 - P\{y_\tau = 0 | x_0 = i, y_0 = j\} = 1 - f(i,j). \end{aligned} \tag{6}$$

(c,d,e) This is a direct application of theorem 14 of the appendix.

□

Theorem 1. *There exists a constant $c \geq 0$ such that*

$$f(1,n) \approx \frac{c}{n} \tag{7}$$

in the Cesaro sense, i.e.

$$\lim_{n\to\infty} \frac{\sum_{j=0}^{n} f(1,j)}{\sum_{j=0}^{n} \left(\frac{c}{n}\right)} = 1. \tag{8}$$

For the proof of this asymptotic relation, we use a method reminiscent of the method used in [17]. First we introduce generating functions, which converge on a particular domain. Inside that domain, these functions satisfy an important functional equation. We will then examine the functional equation to find the dominant singularity of the generating functions outside the domain of convergence. We then investigate the way in which that singularity is approached and finish the proof by means of a Tauberian Theorem.

Proof.

Definition 2. *Define the generating functions* $V : \mathbb{R}^2 \to \mathbb{R}$, *and* $V_{1,2} : \mathbb{R} \to \mathbb{R}$ *by*

$$V(x,y) = \sum_{i=1}^{\infty}\sum_{j=1}^{\infty} f(i,j)x^i y^j, \tag{9}$$

$$V_1(x) = \sum_{i=1}^{\infty} f(i,1)x^i, \tag{10}$$

$$V_2(y) = \sum_{j=1}^{\infty} f(1,j)y^j. \tag{11}$$

Lemma 2. *The power series* $V(x,y)$, $V_1(x)$ *and* $V_2(y)$ *converge uniformly for* $|x| < 1, |y| < 1$, *and in that region the following* ***functional equation*** *is satisfied:*

$$D(x,y)V(x,y) + xy\{(\mu y + \lambda)V_2(y) - (\mu x + \lambda)V_2(x)\} = \frac{x^2 y}{1-x}(\mu y + \lambda y - \mu x - \lambda), \tag{12}$$

where $D(x,y) = xy - \lambda(x^2 y - y^2 x - x - y) - \mu(x^2 - y^2)$.

Proof. Because the coefficients of $V(x,y)$ are probabilities, they are bounded, and as a result $V(x,y)$ converges uniformly on $\mathcal{D} := \{(x,y) \in \mathbb{R}^2 : |x| < 1, |y| < 1\}$. The same reasoning applies to $V_1(x)$ and $V_2(y)$. By applying the recurrence relation (eqn. 1) and a straightforward manipulation with power series, we can derive that

$$D(x,y)V(x,y) + xy(\mu x + \lambda)V_1(x) + xy(\mu y + \lambda)V_2(y) = \frac{(\mu + \lambda)x^2 y^2}{1-x}. \tag{13}$$

Finally, from lemma 1 we obtain $V_1(x) + V_2(x) = \frac{x}{1-x}$. □

We want to derive the leading asymptotic behaviour of $f(1,n)$, so we consider the generating function with those coefficients, which is $V_2(x)$. This function is continuous in the region $-1 < x < 1$, and we look how it diverges as $x \to 1$ or $x \to -1$. We will repeatedly use eqn. 12 in the limit $(x,y) \to (1,1)$, but along different curves. Each curve will provide some new information, which we eventually put together in order to finish the proof.

Lemma 3. $\lim_{x\to 1}(x-1)V_2(x) = 0$.

Proof. First, let y be constant, and let x go to 1. Then the functional equation simplifies to

$$(y-1)^2 \lim_{x\to 1}(x-1)V(x,y) + y \lim_{x\to 1}(x-1)V_2(x) = y(y-1), \tag{14}$$

provided those limits exists. But $V(x,y)$ is a power series with positive coefficients, so it is increasing for $0 < x < 1$ and $0 < y < 1$. Furthermore, we can estimate $V(x,y)$:

$$\begin{aligned} |\sum_{i=1}^{\infty}\sum_{j=1}^{\infty} f(i,j)x^i y^j| &\leq \sum_{i=1}^{\infty}\sum_{j=1}^{\infty} |f(i,j)x^i y^j| \\ \leq \sum_{i=1}^{\infty}\sum_{j=1}^{\infty} |x|^i |y|^j &= \frac{|xy|}{(1-|x|)(1-|y|)}. \end{aligned} \tag{15}$$

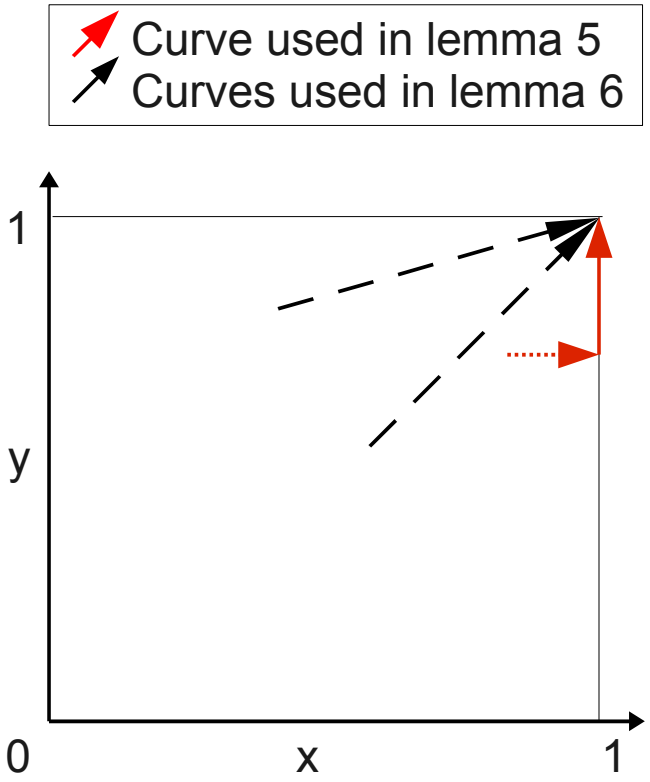

Figure 9: The various curves used in the proof. For lemma 3, we let x go to 1, and then y too. For lemma 4, we let x and y go to 1 simultaneously.

This means that $(x-1)(y-1)V(x,y)$ is increasing and bounded, so the limit exists. The same reasoning applies to $V_2(x)$.

Let us now take the limit $y \to 1$ at both sides of eqn. 14. By the estimate above, the first term goes to zero. The right hand side also goes to zero as $y \to 1$, so the second term has to go to zero. □

Lemma 4. *The limit*

$$L_a := \lim_{t \downarrow 0} [V_2(1-t) - V_2(1-at)] \tag{16}$$

exists for all $a > 0$.

Proof. Consider a curve given by $x_t = 1 - at, y_t = 1 - t$, and let t approach zero from above. Both x_t and y_t are increasing, and we can use the same argument as above to show that $(x_t - 1)(y_t - 1)V(x,y)$ converges. The factor

$\frac{D(x_t,y_t)}{(x_t-1)(y_t-1)}$ also approaches some finite value, so $D(x_t,y_t)V(x_t,y_t)$ converges. By direct computation, we see that the last term does not diverge. Thus, the limit of the second term must also exist. This second term is equal to $(\mu+\lambda)[V_2(1-t)-V_2(1-at)]+\mu t(V_2(1-at)-aV_2(1-t))$. The last contribution will vanish as $t \downarrow 0$, because $\lim_{t\downarrow 0} tV_2(1-t) = 0$. □

Remarkably, this last lemma is all we need to prove the theorem. But we have not used that much details about the transition probabilities of the random walk, which suggests that theorem 1 holds for other random walks too. We will not investigate the conditions necessary to invoke on the transition probabilities to ensure that theorem 1 holds, because that would take us too far from the original problem.

Lemma 5. *$L_a = c\log(a)$ for some $c > 0$.*

Proof. We start with showing that $L_a + L_b = L_{ab}$:

$$\begin{aligned} L_a + L_b &= \lim_{t\downarrow 0}[V_2(1-t) - V_2(1-at)] + \lim_{t\downarrow 0}[V_2(1-t) - V_2(1-bt)] \\ &= \lim_{t\downarrow 0}[V_2(1-t) - V_2(1-at)] + \lim_{\tilde{t}\downarrow 0}[V_2(1-a\tilde{t}) - V_2(1-ab\tilde{t})] \\ &= \lim_{t\downarrow 0}[V_2(1-t) - V_2(1-abt)] = L_{ab} \end{aligned} \tag{17}$$

Furthermore, we have that

$$a \leq b \iff 1-at \geq 1-bt \iff V_2(1-at) \geq V_2(1-bt) \iff L_a \leq L_b. \tag{18}$$

Therefore, $a \to L_a$ is a nondecreasing homomorphism from $(0,\infty)$ to $\mathbb{R}$. That means it has to be continuous, and, even stronger, $L_a = c\log(a)$ for some $c \geq 0$. □

Lemma 6. *$V_2(1-t) = -c\log(t) + o(\log(t))$ as $t \downarrow 0$.*

Proof. Define $W(t) := V_2(1-t) + c\log(t)$, and $\phi(x) =:= e^{W(\frac{1}{x})}$. Then for all $a > 0$ we have that $\lim_{t\downarrow 0}[W(t) - W(at)] = 0$, and therefore, $\lim_{x\to\infty} \frac{\phi(ax)}{\phi(x)} = 1$. In other words, ϕ is a slowly varying function. We can use the following fact about these functions, which can be found in [2]:

Theorem 2. *Let $f(x)$ be a slowly varying function. Then there exists $B \geq 0$ and $\eta, \varepsilon : \mathbb{R} \to \mathbb{R}$ such that $\lim_{x\to\infty} \eta(x)$ exists, $\lim_{x\to\infty} \varepsilon(x) = 0$, and for all $x \geq B$:*

$$f(x) = e^{\eta(x) + \int_B^x \frac{\varepsilon(t)}{t} dt}. \tag{19}$$

So we can write $W(\frac{1}{x}) = \eta(x) + \int_B^x \frac{\varepsilon(t)}{t} dt$, and therefore, $\lim_{x\to\infty} \frac{W(\frac{1}{x})}{\log x} = 0$. In other words, $\lim_{t\downarrow 0} \frac{W(t)}{\log t} = 0$, which proves that $V_2(x) = -c\log(1-x) + o(\log(1-x))$. □

Lemma 7. *$\lim_{x\to -1} V_2(x)$ exists.*

Proof. Because $f(1,j)$ is decreasing and always positive, the limit $\lim_{j\to\infty} f(1,j)$ exists. Suppose that this limit is equal to $\varepsilon > 0$. Then

$$V_2(x) = \sum_{j=1}^{\infty} f(1,j)x^j \geq \varepsilon \sum_{j=1}^{\infty} x^j = \frac{\varepsilon x}{1-x}. \tag{20}$$

But this means that $V_2(1-t)$ diverges like $\frac{1}{t}$ or faster as $t \downarrow 0$. Since that is in contradiction with the previous lemma, we must have that $\lim_{j\to\infty} f(1,j) = 0$. Therefore, by the Leibniz criterium, $V_2(-1) = \sum_{j=1}^{\infty}(-1)^j f(1,j)$ exists. And because a power series is always continuous inside its domain of convergence, $\lim_{x\to -1} V_2(x)$ exists. □

Now we can complete the proof of the main theorem by means of the following tauberian theorem [2].

Theorem 3. Karamata's Tauberian Theorem
Let a_n be a sequence of non-negative real numbers such that the power series $A(x) := \sum_{n=1}^{\infty} a_n x^n$ converges for $x \in [0,1)$. Then, for $c, \rho \geq 0$ and g a slowly varying function, the following are equivalent:

$$\sum_{k=0}^{n} a_k \sim \frac{c}{\Gamma(1+\rho)} n^{\rho} g(n) \text{ as } n \to \infty, \tag{21}$$

and

$$A(x) \sim \frac{c}{(1-x)^{\rho}} g\left(\frac{1}{1-x}\right) \text{ as } x \to 1, \tag{22}$$

When we apply theorem 3, with $a_n = f(1,n)$, $\rho = 0$, $g(n) = \log(n)$, we see that

$$V_2(x) \sim -c\log(1-x) \iff \sum_{j=1}^{n} f(1,j) \sim c\log(n), \tag{23}$$

which completes the proof. □

4.3.2 Limiting cases

By now we have some idea about the dependence of f on the starting point $(1,n)$. As a next step, we will focus on the effect on μ and λ. Remark that the asymptotic relation (theorem 1) holds for all λ and μ. With that in mind, we do not expect huge differences in the behaviour of f as we change the ratio of μ versus λ. It turns out that we can give exact results for the cases $\mu = 0$ and $\lambda = 0$.

Theorem 4. *If $\lambda = 0$, the probability $f(x,y)$ is given by*

$$f(x,y) = \frac{x}{x+y}. \tag{24}$$

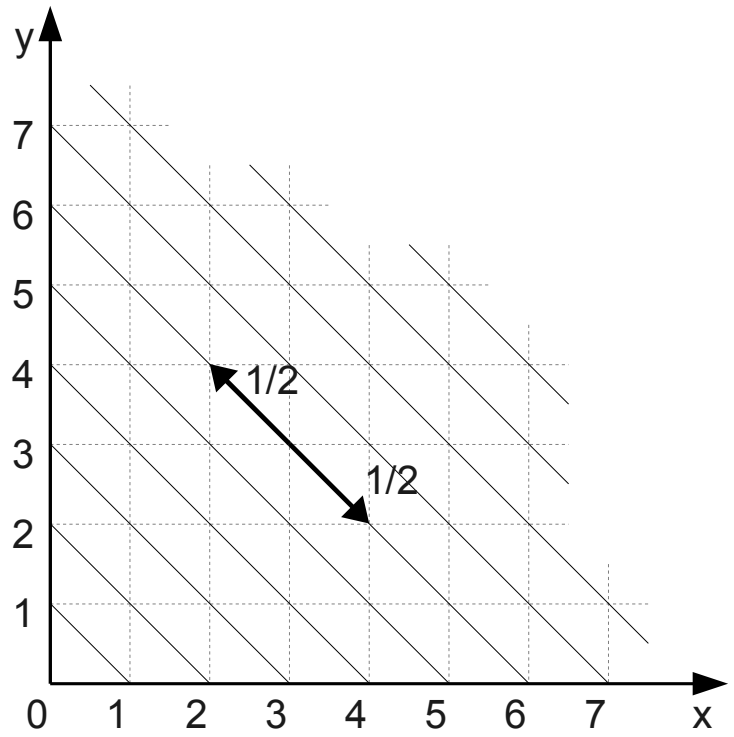

Figure 10: If $\lambda = 0$, the random walk has to stay on a single diagonal.

Proof. In this case, there are only diagonal jumps, so the random walk will stay on the set $\{(x', y') \in \mathbb{Z}_+^2 : x' + y' = x + y\}$. The resulting process is equivalent to a simple symmetric random walk on the set $\{1, 2, \ldots, x+y\}$, started at x and stopped upon reaching 0 or $x + y$. So we have to solve the Dirichlet problem $\phi(0) = 0$, $\phi(N) = 1$, $\Delta\phi = 0$, where $N := x + y$, and

$$\Delta\phi(x) = \phi(x-1) + \phi(x+1) - 2\phi(x). \tag{25}$$

The solution to this Dirichlet problem is given by the linear equation $\phi(x) = \frac{x}{N}$, which means that $f(x, y) = \frac{x}{x+y}$. □

Note that this formula is in agreement with the above asymptotic. The other limit, where μ goes to zero, is trickier to calculate. In the previous calculation the problem essentially reduced one dimension, and that did the trick. In the present case, however, the problem is still manifestly 2-dimensional. Luckily, the resulting random walk happens to be one of the most studied stochastic processes in history. It is again a simple symmetric random walk, only this time in two dimensions. The generator of the process is called the **discrete**

Laplacian, and it is given by

$$\Delta f(x,y) := \frac{1}{4}[f(x,y+1)+f(x,y-1)+f(x+1,y)+f(x-1,y)] - f(x,y). \quad (26)$$

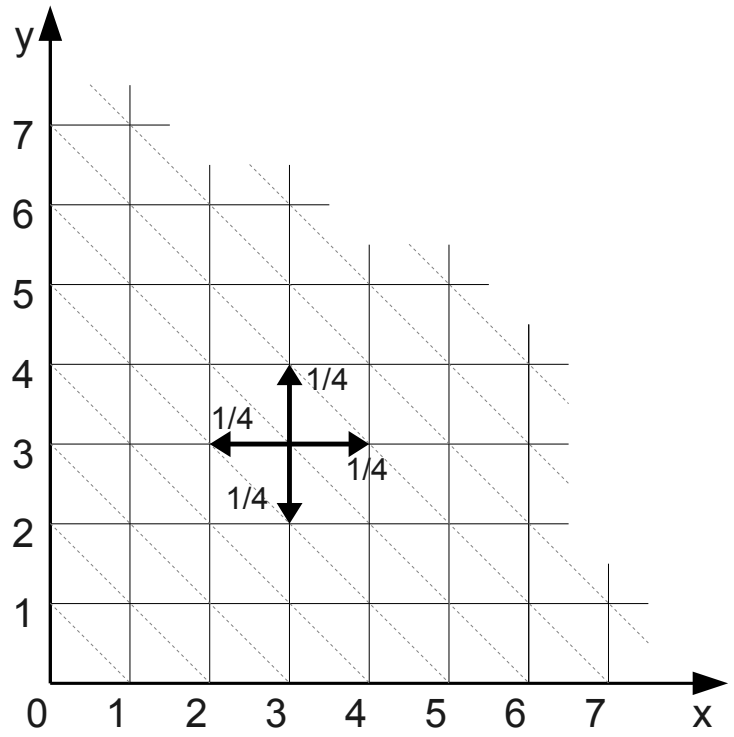

Figure 11: In the limit $\mu \to 0$, the diagonal jumps disappear. However, the process is still 2-dimensional.

The main tool we will use to solve the Dirichlet problem is the following theorem, which is explained and proven in the appendix. It is a slight modification of a result of Chung and Yau ([6, 15]). The theorem applies to a more general case than necessary for this paper, because in our case all edge weights will be equal to 1. However, we prefer to state the theorem in its strongest form.

Theorem 5. *Let x_t be a Markov chain with finite state space X, and edge weights w_{xy}. Consider $S \subset X$, and let $\mathbb{L}_S$ be the normalized Laplacian of S. Let $\{(\phi_i, \lambda_i), i \in \mathcal{I}\}$ be an orthonormal eigensystem of $\mathbb{L}_S$. The solution to the Dirichlet problem is then given by:*

$$f(x) = \sum_{i \in \mathcal{I}} \frac{1}{\lambda_i} \sum_{\substack{z \in S \\ z \sim y \in \delta S}} w_{yz} \phi_i(z) \sigma(y) d_z^{-1/2} d_x^{-1/2} \phi_i(x). \quad (27)$$

This theorem enables us to solve the Dirichlet problem for any given boundary condition, once we know the orthonormal eigenfunctions of the (normalized) Laplacian. However, the theorem only works for Markov chains with finite state space, while we are dealing with an infinite lattice. Therefore, we will adopt the following limiting procedure:

Definition 3. *Let τ_N be the first time the walker exits an $N \times N$-box, i.e.*

$$\tau_N = \inf\{t \geq 0 : x_t = 0 \vee y_t = 0 \vee x_t = N \vee y_t = N\}. \tag{28}$$

Let $f_N(x, y)$ be the probability that the $N \times N$-box is left because the walker hits the x-axis:

$$f_N(x, y) = P(y_{\tau_N} = 0 | x_0 = x, y_0 = y). \tag{29}$$

Because the random walk is recurrent, f_N converges to f as N goes to infinity. The functions f_N also satisfy the laplace equation, but with boundary conditions $f_N(0, y) = f_N(x, N) = f_N(y, N) = 0$ and $f_N(x, 0) = 1$ for all $0 < x < N$ and $0 < y < N$. So we can apply theorem 5 to find f_N.

Lemma 8. *The orthonormal eigenfunctions of the Laplacian of an $N \times N$-box are given by*

$$\phi_{mn}(x, y) = \frac{2}{N} \sin(\frac{\pi m x}{N}) \sin(\frac{\pi n y}{N}), \tag{30}$$

with corresponding eigenvalues

$$\lambda_{mn} = 1 - \frac{1}{2}\cos(\frac{n\pi}{N}) - \frac{1}{2}\cos(\frac{m\pi}{N}), \tag{31}$$

for $m, n \in \{1, 2, \ldots, N-1\}$.

Proof. The graph of an $N \times N$-box is regular in the sense that all vertices have the same degree, and the random walk is regular in the sense that all edge weights are equal to 1. Therefore, the normalized Laplacian is equal to the discrete Laplacian. It is easily verified that $\Delta\phi_{mn} = \lambda_{mn}\phi_{mn}$ for all $m, n < N$. □

Now we can apply theorem 5, which gives

$$\begin{aligned} f_N(k, l) &= \frac{1}{4} \sum_{m=1}^{N-1} \sum_{n=1}^{N-1} \frac{1}{\lambda_{mn}} \sum_{x=1}^{N-1} \phi_{mn}(x, 1)\phi_{mn}(k, l) && (32) \\ &= \frac{1}{N^2} \sum_{m=1}^{N-1} \sum_{n=1}^{N-1} \sum_{x=1}^{N-1} \frac{\sin(\frac{\pi m x}{N}) \sin(\frac{\pi n}{N}) \sin(\frac{\pi m k}{N}) \sin(\frac{\pi n l}{N})}{1 - \frac{1}{2}\cos(\frac{n\pi}{N}) - \frac{1}{2}\cos(\frac{m\pi}{N})} && (33) \end{aligned}$$

The summation over x can be done explicitely:

$$\sum_{x=1}^{N-1} \sin(\frac{\pi m x}{N}) = \begin{cases} 0 & m \text{ even} \\ \frac{\sin(\frac{m\pi}{N})}{1 - \cos(\frac{m\pi}{N})} & m \text{ odd} \end{cases} \tag{34}$$

which yields

$$f_N(k,l) = \frac{1}{N^2} \sum_{\substack{m=1 \\ m \text{ odd}}}^{N-1} \sum_{n=1}^{N-1} \frac{\sin(\frac{\pi m}{N}) \sin(\frac{\pi n}{N}) \sin(\frac{\pi mk}{N}) \sin(\frac{\pi nl}{N})}{(1-\cos(\frac{\pi m}{N}))(1-\frac{1}{2}\cos(\frac{n\pi}{N})-\frac{1}{2}\cos(\frac{m\pi}{N}))}. \quad (35)$$

Now, in the limit of $N \to \infty$, the summations become integrals. The first summation is only over the odd integers, so that results in an overall factor of $\frac{1}{2}$. Because the summand involves functions of $\frac{\pi n}{N}$ rather than $\frac{n}{N}$, there is also a factor of $\frac{1}{\pi^2}$.

$$f(k,l) = \frac{1}{2\pi^2} \int_0^\pi \int_0^\pi \frac{\sin(u)\sin(v)\sin(ku)\sin(lv)}{(1-\cos(u))(1-\frac{\cos(u)}{2}-\frac{\cos(v)}{2})} dudv \quad (36)$$

The integrand can be simplified by introducing the Chebyshev polynomials of the second kind, defined by $U_k(\cos(u)) = \frac{\sin((k+1)u)}{\sin(u)}$.

$$f(k,l) = \frac{1}{2\pi^2} \int_0^\pi \int_0^\pi \frac{\sin(u)^2 \sin(v)^2 U_{k-1}(\cos(u)) U_{l-1}(\cos(v))}{(1-\cos(u))(1-\frac{\cos(u)}{2}-\frac{\cos(v)}{2})} dudv \quad (37)$$

Using the trigonometric identities $\frac{\sin(u)^2}{1-\cos(u)} = 1+\cos(u)$, and $1-\frac{\cos(u)}{2}-\frac{\cos(v)}{2} = \sin^2(u/2)+\sin^2(v/2)$, we can rewrite the integral to

$$f(k,l) = \frac{1}{2\pi^2} \int_0^\pi \int_0^\pi (1+\cos(u)) \frac{\sin(v)^2 U_{k-1}(\cos(u)) U_{l-1}(\cos(v))}{\sin^2(u/2)+\sin^2(v/2)} dudv \quad (38)$$

From this point on, we describe a method to solve this integral, but won't keep track of the exact numbers. Note that the integral is a linear combination of integrals of the type

$$I_{mn} = \int_0^\pi \int_0^\pi \frac{\sin(v)^2 \cos^m(u) \cos^n(v)}{\sin^2(u/2)+\sin^2(v/2)} dudv \quad (39)$$

Those integrals can be transformed by changing variables to $x = \frac{u}{2}$ and $y = \frac{v}{2}$, and using the double angle formulas $\sin(2x) = 2\sin(x)\cos(x), \cos(2x) = 1 - 2\sin^2(x)$.

$$I_{mn} = 4 \int_0^{\frac{\pi}{2}} \int_0^{\frac{\pi}{2}} \frac{(\sin^2(x)-\sin^4(x))(1-2\sin^2(x))^m (1-2\sin^2(y))^n}{\sin^2(x)+\sin^2(y)} dxdy \quad (40)$$

Expanding the nominator produces a lot of integrals of the form

$$J_{mn} = \int_0^{\frac{\pi}{2}} \int_0^{\frac{\pi}{2}} \frac{\sin^{2m}(x)\sin^{2n}(x))}{\sin^2(x)+\sin^2(y)} dxdy. \quad (41)$$

Finally, we use long division to convert this integral into the sum of easier integrals, by noting that $\frac{x^m y^n}{x+y} = x^m y^{n-1} - x^{m+1} y^{n-2} + ... + (-1)^n \frac{x^{m+n}}{x+y}$. The resulting integrals can then be directly solved for all $m, n > 0$:

$$\int_0^{\frac{\pi}{2}} \int_0^{\frac{\pi}{2}} \sin^{2m}(x)\sin^{2n}(x) dxdy = \frac{\pi^2}{2^{2m}2^{2n}} \binom{2m-1}{m-1} \binom{2n-1}{n-1} \quad (42)$$

$$\int_0^{\frac{\pi}{2}} \int_0^{\frac{\pi}{2}} \frac{\sin^{2m}(x)}{\sin^2(x)+\sin^2(y)} dxdy = \frac{\pi 2^n \Gamma(\frac{m}{2})^2}{16\Gamma(m)} \tag{43}$$

The integral in eqn. 36 is a sum of these terms, with the appropriate prefactor. However, the number of terms necessary to compute the value of the integral gets out of hand quite quickly. We need to find the value of $f(1,13)$, which is just doable with a normal computer algorithm. The result is quite unusual:

$$f(1,13) = 42344121 - \frac{1198449065536}{9009\pi} \approx 0.048969\ldots \tag{44}$$

4.3.3 Remarks

We have obtained an expression for $f(1,13)$, so from a physics point of view, we are done. From a mathematical point of view, however, there are some interesting remarks to be made. The integral in eqn. 42 is always of the form π^2 times a rational number, whereas the second integral (eqn. 43) will be π^2 times a rational number if m is odd and π times a rational if m is even. Because f is a $\mathbb{Q}$-linear combination of these integrals, we know that

$$\forall m,n>0 : \exists q_{mn}, r_{mn} \in \mathbb{Q} : f(m,n) = q_{mn} + \frac{r_{mn}}{\pi}. \tag{45}$$

Because π is irrational, the only way that f can satisfy the recurrence relation (eqn. 1) is if that recurrence relation holds for q_{mn} and r_{mn} too.

From the table it appears that q_{mn} is always an integer (except on the diagonal). So we hypothesise:

$$\forall m \neq n : q_{mn} \in \mathbb{Z}. \tag{46}$$

If we look closer at the specific numbers, particularly just below and above the diagonal, there is another interesting observation:

$$q_{m,m+1} = \begin{cases} -m & m \equiv 0 \mod 2 \\ m+1 & m \equiv 1 \mod 2 \end{cases} \tag{47}$$

If eqn. 47 holds, then eqn. 46 follows by mathematical induction.

4.3.4 Continuum limit

This method does not generalise from the limiting cases $\mu = 0$ or $\lambda = 0$ to the general case $\mu, \lambda > 0$. In order to use theorem 5, it is necessary to know the eigenfunctions of the normalized Laplacian. For a general graph, determining the eigenfunctions is just as hard as solving the Dirichlet problem directly.

At this point we have to introduce a new approximation before we can proceed. This new approximation is the continuum limit. We will first try to explain the philosophy behind this approximation, then give a precise mathematical definition. The result in eqn. 44 is not quite what one expects beforehand, and we suspect this happens because we are dealing with a discrete lattice. Therefore,

0	550	-512	218	-48	6	$\frac{1}{2}$
0	121	-94	31	-4	$\frac{1}{2}$	-5
0	28	-16	4	$\frac{1}{2}$	5	49
0	7	-2	$\frac{1}{2}$	-3	-30	-217
0	2	$\frac{1}{2}$	3	17	95	513
0	$\frac{1}{2}$	-1	-6	-27	-120	-549
	1	1	1	1	1	1

(a) q_{mn}

0	$-\frac{60464}{35}$	$\frac{11264}{7}$	$-\frac{43088}{63}$	$\frac{47872}{315}$	$-\frac{60496}{3465}$	0
0	$-\frac{5696}{15}$	$\frac{31088}{105}$	$-\frac{10112}{105}$	$\frac{4384}{315}$	0	$\frac{60496}{3465}$
0	$-\frac{1312}{15}$	$\frac{256}{5}$	$-\frac{1184}{105}$	0	$-\frac{4384}{315}$	$-\frac{47872}{315}$
0	$-\frac{64}{3}$	$\frac{112}{15}$	0	$\frac{1184}{105}$	$\frac{10112}{105}$	$\frac{43088}{63}$
0	$-\frac{16}{3}$	0	$-\frac{112}{15}$	$-\frac{256}{5}$	$-\frac{31088}{105}$	$-\frac{11264}{7}$
0	0	$\frac{16}{3}$	$\frac{64}{3}$	$\frac{1312}{15}$	$\frac{5696}{15}$	$\frac{60464}{35}$
	0	0	0	0	0	0

(b) r_{mn}

Table 1: Specific values of q_{mn} and r_{mn}, for $m, n \leq 6$. It is most remarkable that all off-diagonal values of q_{mn} are integers.

we refine the lattice by adding extra vertices and edges in a reasonable way, in the hope that this effect smoothens out.

So we introduce vertices at the middle of every edge, and at the center of each square. Then we connect the new vertices such that the new graph is similar to the old one, but with a finer grid. The process on this grid will evolve much slower than the original one, so we also speed up the time by an appropriate factor. This procedure is iterated, and finally, we end up with a process on a continuous lattice. For more information on continuum limits, see [9, 1].

Definition 4. *Define $X_t = (x_t, y_t)$, viewed as a process on $\mathbb{R}^2$ instead of $\mathbb{Z}^2$ but started at $(x, y) \in \mathbb{Z}^2$. Then set $X_t^{(N)} = \frac{1}{N} X_{N^2 t}$. Furthermore, let X_t^c denote the process with state space $\mathbb{R}^2$ and generator*

$$L_c = \lambda \frac{\partial^2}{\partial x^2} + \lambda \frac{\partial^2}{\partial y^2} + \mu \left(\frac{\partial}{\partial x} - \frac{\partial}{\partial y} \right)^2. \tag{48}$$

Lemma 9. *$X_t^{(N)}$ converges to X_t^c in the sense that $S_N(t)f \to S_c(t)f$ uniformly for all $f \in C(X)$ and t in compact sets.*

Proof. In order to show convergence of the process, it is sufficient to show a certain type of convergence of the generator of the process. That type of convergence is made clear in the following definition and theorem by Trotter and Kurtz [9, 5].

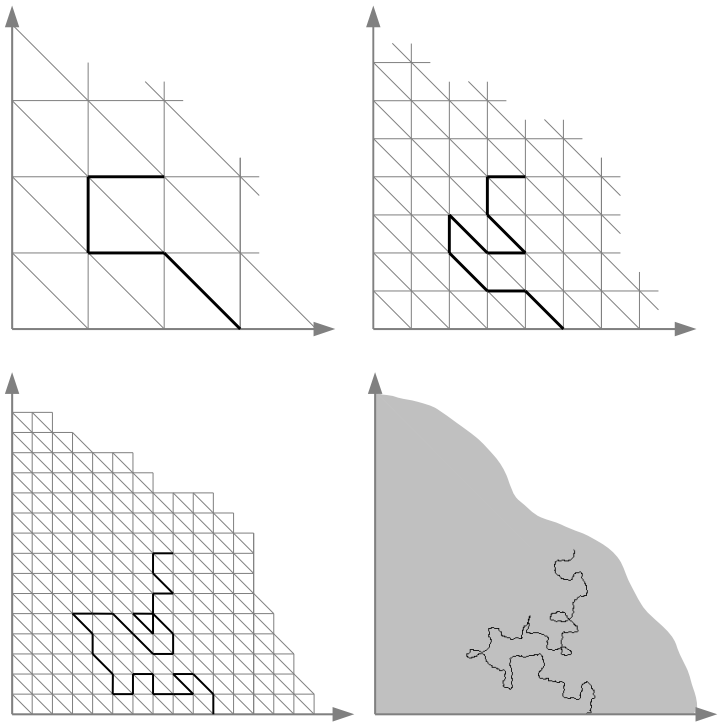

Figure 12: Visualization of the continuum limit. The process $X_t^{(N)}$ is defined on the state space $\frac{1}{N}\mathbb{Z}_+^2 = \{\left(\frac{x}{N}, \frac{y}{N}\right) : (x, y) \in \mathbb{Z}_+^2\}$. As $N \to \infty$, $X_t^{(N)}$ converges to a continuous process.

Definition 5. *A **core** for a Markov generator L is a linear subspace $\mathcal{D} \subset D(L)$ such that L is the closure of its restriction to $\mathcal{D}$.*

Theorem 6. Trotter-Kurtz Theorem
Let L_N and L be the generators of the semigroups $S_n(t)$ and $S(t)$, respectively. Suppose that there exists a core $\mathcal{D}$ for L such that $\mathcal{D} \subset D(L_n)$ for all n, and $L_n f \to Lf$ uniformly for all $f \in \mathcal{D}$. Then

$$S_n(t)f \to S(t)f \tag{49}$$

uniformly for all $f \in C(X)$ and t in compact sets.

Therefore, we compute the generator L_N of $X_t^{(N)}$:

$$\begin{aligned} L_N f(x,y) &= \lambda N^2[f(x,y+\frac{1}{N}) + f(x,y-\frac{1}{N}) - 2f(x,y) \\ &+ f(x-\frac{1}{N},y) + f(x+\frac{1}{N},y) - 2f(x,y)] \\ &+ \mu N^2[f(x+\frac{1}{N},y-\frac{1}{N}) + f(x-\frac{1}{N},y+\frac{1}{N}) - 2f(x,y)] \end{aligned} \tag{50}$$

It is possible to show that $L_N f \to L_c f$ uniformly for all $f \in C_0^\infty(X)$, the set of all infinitely differentiable functions for which all derivates go to zero uniformly at large distances. Because the transition kernel of Brownian motion decays exponentially, this set is mapped into itself by $S_c(t)$. Also, the set of those functions lies dense in $C_c(X)$, the space of all continuous functions with compact support, which contains the domain of L_c. the following theorem [5] then guarantees that $C_0^\infty(X)$ is a core for L_c.

Theorem 7. *Let L be the generator of a Markov process, and $S(t)$ the semigroup of that process. If D is a dense subset of $\mathcal{D}(L)$, and*

$$\forall f \in D, t \geq 0 : S(t)f \in D, \tag{51}$$

then D is a core for L.

□

Let us see what effect the continuum limit has on our object of study, $f(x,y)$. We define

$$\tau_N := \inf\{t \geq 0 : x_t^{(N)} = 0 \vee y_t^{(N)} = 0\}, \tag{52}$$

and

$$f_N(x,y) := P(x_{\tau_N}^{(N)} = 0 | x_0^{(N)} = x, y_0^{(N)} = y). \tag{53}$$

Analogously, for the continuous process, we have

$$\tau_c := \inf\{t \geq 0 : x_t^c = 0 \vee y_t^c = 0\}, \tag{54}$$

and

$$f_c(x,y) := P(x_{\tau_c}^c = 0 | x_0^c = x, y_0^c = y). \tag{55}$$

Because the continuum limit only involves scaling in space and time, f_N simplifies a lot. First of all, the event $\{x_t = 0 \vee y_t = 0\}$ is invariant under scaling of space coordinates, so $\tau_N = \tau$. Furthermore, the event that $x_\tau^{(N)} = 0$ is invariant under time rescaling. Combining these remarks, we can say that

$$\begin{aligned} f_N(x,y) &= P(x_{\tau_N}^{(N)} = 0 | x_0^{(N)} = x, y_0^{(N)} = y) \\ &= P(x_\tau^{(N)} = 0 | x_0^{(N)} = x, y_0^{(N)} = y) \\ &= P(x_\tau = 0 | x_0^{(N)} = x, y_0^{(N)} = y) \\ &= P(x_\tau = 0 | x_0 = Nx, y_0 = Ny) \\ &= f(Nx, Ny) \end{aligned} \tag{56}$$

Meanwhile, because $X_t^{(N)}$ converges to X_t^c, f_N converges to f_c at least pointwise, and therefore

$$f_c(x,y) = \lim_{N\to\infty} f_N(x,y) = \lim_{N\to\infty} f(\lfloor Nx\rfloor, \lfloor Ny\rfloor), \tag{57}$$

where we have to floor Nx and Ny because f is defined only on $\mathbb{Z}^2$, and f_c on $\mathbb{R}^2$. From this expression it immediately follows that

$$f_c(kx,ky) = f_c(x,y) \tag{58}$$

for all $k > 0$. Therefore, we can approximate

$$f(x,y) = f_N\left(\frac{x}{N}, \frac{x}{N}\right) \approx f_c\left(\frac{x}{N}, \frac{x}{N}\right) = f_c(x,y). \tag{59}$$

So it remains to calculate $f_c(x,y)$. By theorem 14, the function f_c is the solution to the Dirichlet problem

$$\begin{aligned} f_c(x,0) &= 1 \\ f_c(0,y) &= 0 \\ L_c f_c &= 0. \end{aligned} \tag{60}$$

Theorem 8. *The solution of the Dirichlet problem is given by*

$$f_c(x,y) = \frac{1}{2} + \frac{\arctan(\alpha\frac{x-y}{x+y})}{2\arctan(\alpha)}, \tag{61}$$

where $\alpha := \frac{1}{\sqrt{1+2\frac{\mu}{\lambda}}}$.

Proof. We start by rewriting the generator to

$$L_c = \frac{\lambda}{2}\left(\frac{\partial}{\partial x} + \frac{\partial}{\partial y}\right)^2 + \frac{\lambda+2\mu}{2}\left(\frac{\partial}{\partial x} - \frac{\partial}{\partial y}\right)^2, \tag{62}$$

then introduce the new coordinates $u := \frac{x+y}{\sqrt{\lambda}}$, $v := \frac{x-y}{\sqrt{\lambda+2\mu}}$. In these new coordinates, the generator takes on the simple form

$$L_c = \frac{1}{2}\left(\frac{\partial^2}{\partial u^2} + \frac{\partial^2}{\partial v^2}\right). \tag{63}$$

In other words, if $L_c f_c = 0$, then f_c is an harmonic function of the coordinates u and v. We will try to construct this harmonic function as the imaginary part of a holomorphic function ϕ:

$$f_c(u,v) = \Im(\phi(u+iv)). \tag{64}$$

The domain of this function is a wedge in the complex plane: $D_\phi := \{z \in \mathbb{C}\backslash\{0\} : -\arctan(\alpha) < \arg(z) < \arctan(\alpha)\}$. This domain can be mapped holomorphically onto the upperhalf complex plane $\mathcal{H} := \{z \in \mathbb{C} : \Im(z) > 0\}$ by

$$\begin{aligned} g : D_\phi &\to \mathcal{H} \\ z &\to iz^{\left(\frac{\pi}{2\arctan\alpha}\right)} \end{aligned} \tag{65}$$

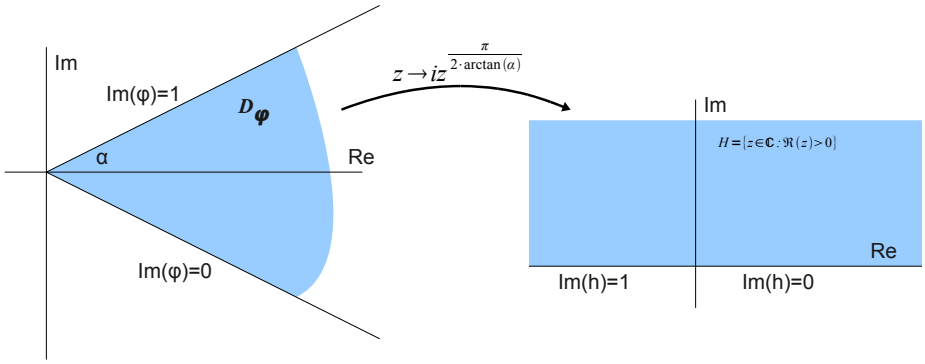

Figure 13: The function g is a biholomorphic mapping from the wedge to the upperhalf plane.

This function stretches the wedge so that its opening angle becomes π, then rotates it by $\frac{\pi}{2}$. The upper part of the boundary is mapped to the negative part of the real axis, and the lower part to the positive real axis. Now consider $h(z) = \frac{1}{\pi}\log(z)$, where log denotes the principal continuation of the real logarithm. This function maps the positive real line to the real numbers, and the negative real line to the complex numbers with imaginary part 1. So the boundary conditions are satisfied if we choose

$$\phi(z) = h(g(z)). \tag{66}$$

If we now work out the expression for $f_c(x, y)$, we obtain

$$f_c(x, y) = \frac{1}{2} + \frac{\arctan(\alpha \cdot \frac{x-y}{x+y})}{2\arctan(\alpha)}. \tag{67}$$

□

4.3.5 Quality of the continuum approximation

The continuum limit proved to be very effective in solving the Dirichlet problem, and we even obtained an exact solution. However, one may wonder how useful this formula is, because we approximated the discrete process $X_t^{(N)}$ with the continuous process X_t^c. In this section, we show that the continuous and the discrete case are indeed different, but that the difference is very small.

If f_c would be equal to f, then all the previous results about f should hold for f_c as well. We can verify that the asymptotic relation holds by making a

Taylor expansion of $f_c(1,n)$ in $\varepsilon = \frac{1}{n+1}$.

$$\begin{aligned} f_c(1,n) &= \frac{1}{2} + \frac{\arctan(\alpha(\frac{1-n}{1+n}))}{2\arctan(\alpha)} = \frac{\arctan\alpha - \arctan(\alpha - \frac{2\alpha}{n+1}))}{2\arctan\alpha} \\ &\approx \frac{2\alpha}{n+1}\frac{\frac{d}{dt}(\arctan t)_{t=\alpha}}{2\arctan\alpha} = \frac{\alpha}{(1+\alpha^2)\arctan\alpha}\frac{1}{n+1}. \end{aligned} \tag{68}$$

This expansion also suggests that the prefactor c is equal to $\frac{\alpha}{(1+\alpha^2)\arctan\alpha}$.

Let us now look at the behaviour of f_c around $\lambda = 0$. Small values of λ correspond to small values of α, so we make a Taylor expansion of f_c in terms of α:

$$f_c(x,y) = \frac{1}{2} + \frac{\arctan(\alpha\frac{x-y}{x+y})}{2\arctan(\alpha)} \approx \frac{1}{2} + \frac{\alpha(\frac{x-y}{x+y})}{2\alpha} = \frac{x}{x+y}. \tag{69}$$

So the formula is also correct in the limit $\lambda \to 0$. However, in the limit $\mu \to 0$, the result is different. If $\mu = 0$, then $\alpha = 1$, and $\arctan\alpha = \frac{\pi}{4}$, so

$$f_c(x,y) = \frac{1}{2} + \frac{2}{\pi}\arctan\left(\frac{x-y}{x+y}\right), \tag{70}$$

$$f_c(1,13) = \frac{1}{2} - \frac{2}{\pi}\arctan\left(\frac{6}{7}\right) \approx 0.048875\ldots \tag{71}$$

This number is actually rather close to the result of eqn. 44, but it is not the same. So the continuum limit is a very good approximation, but still an approximation.

Of course, after all these calculations there is still the question how the function $f(x,y,\mu,\lambda)$ actually looks like. To get an idea, we simulated the infinite lattice process, and fitted the data with the curve obtained by the continuum limit. To see how good the infinite lattice approximation is, we also simulated the original process. The curve from the infinite lattice process is almost exactly fitted by the continuum curve. However, the infinite lattice curve and the finite lattice curve are quite far apart, although similar in form. In the next section, we will refine the approximation in order to get more agreement with the original process.

4.4 Triangle approximation

Remember that we modeled the finite state space random walk with another random walk which has the same transition probabilities, but an infinite state space. This has the advantage that the calculations are easier, but it comes at a very high cost. The new process behaves in a similar manner as the original one, but the actual numbers are different. Now we would like to get more agreement between our approximation and the actual process. This will almost necessarily mean that the calculations become more difficult, but we hope that they're still doable.

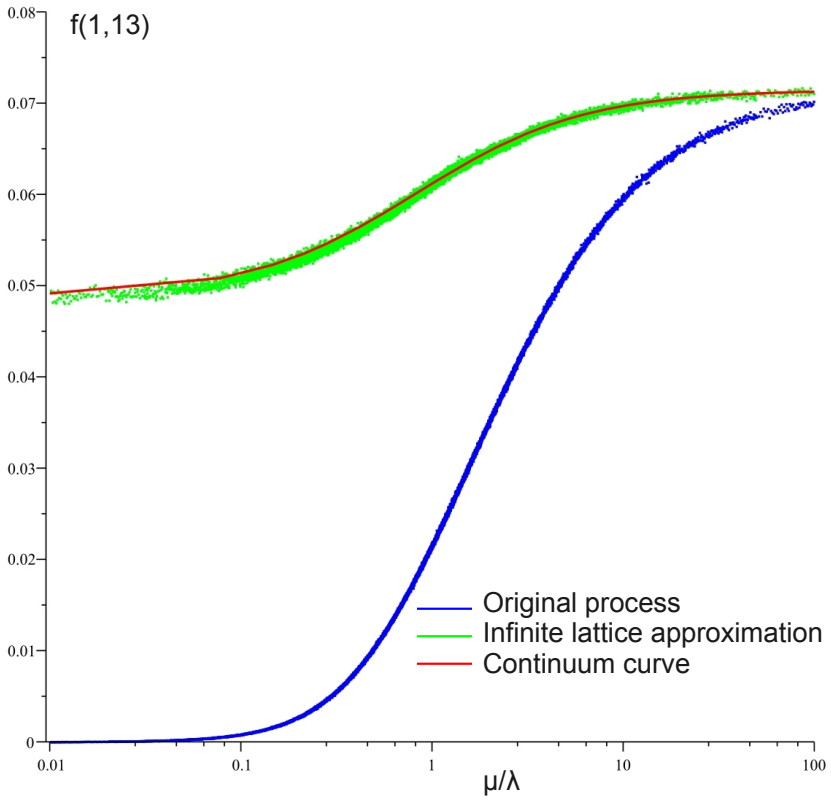

Figure 14: Graph of $f(1,13)$ as a function of $\frac{\mu}{\lambda}$. Shown are the original Markov chain, the infinite lattice approximation and the continuum limit. Remark that the horizontal axis has a logarithmic scale.

The crucial difference between the approximation and the original process is that on the infinite lattice, the endpoints a_t and b_t can move arbitrarily far apart. In the original process, however, the maximal distance $\max\{b_t - a_t, t \geq 0\}$ is 14. We can ensure that this is always the case, by conditioning on the event that $\forall t : x_t + y_t \leq 14$. This means that whenever $x_t + y_t = 14$, the jumps $x_t \to x_t + 1$ and $y_t \to y_t + 1$ are prohibited.

Motivated by earlier success, we will again switch to a continuum limit. In that limit, the process becomes the same anisotropic Brownian motion as before, but this time on the triangle $\{(x, y) \in \mathbb{R}_+^2 | x + y \leq 14\}$. Both coordinate axes are absorbing, and the hypothenuse is a reflecting boundary. We can remove this reflecting boundary by adding another triangle and form a square (see figure **??**). When the Brownian particle would be reflected by the boundary, imagine that it continues its path inside the other triangle. If we then identify the upper side of

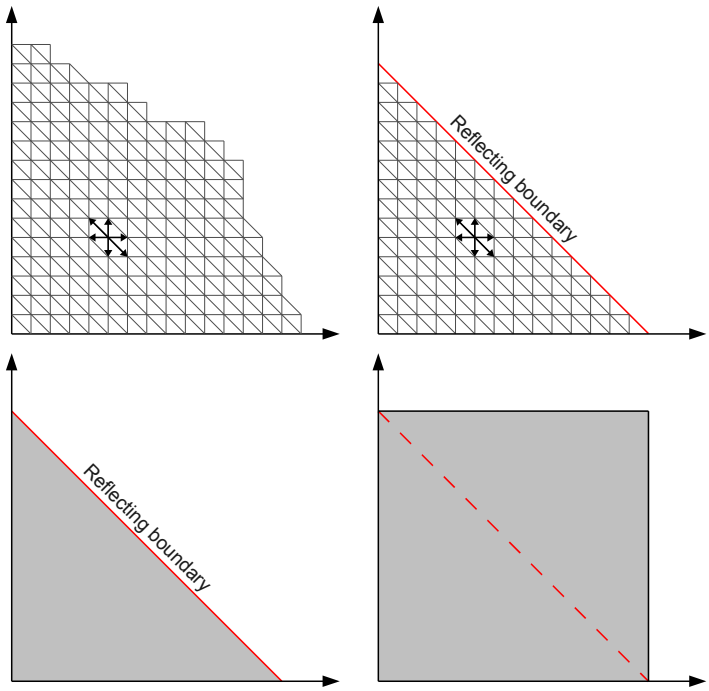

Figure 15: To get the triangle approximation, we use three steps. First, we confine the process to the triangle $\{(x, y)|x + y \leq 14\}$. Then we switch to continuum. Finally, we add another triangle to get a square.

the square with the left side, as well as the lower and right sides, all boundaries are absorbing. Because of these identifications, it doesn't matter whether the particle is absorbed at the bottom or at the right, or whether it hits the upper or left side of the square. It only matters which of these combinations (bottom + right vs. upper + left) it hits.

4.4.1 Calculations

Let us also give a mathematically precise statement of our model.

Definition 6. *The probability $\tilde{f}_c(x, y)$ is the solution to the Dirichlet problem*

$$\tilde{f}_c(0, y) = \tilde{f}_c(x, 14) = 0, \tag{72}$$

$$\tilde{f}_c(x, 0) = \tilde{f}_c(14, y) = 1, \tag{73}$$

$$\lambda \left(\frac{\partial}{\partial x} + \frac{\partial}{\partial y} \right)^2 \tilde{f}_c + (\lambda + 2\mu) \left(\frac{\partial}{\partial x} - \frac{\partial}{\partial y} \right)^2 \tilde{f}_c = 0. \tag{74}$$

We will try to solve this Dirichlet problem using the same techniques as before. So we again change coordinates to $u := \frac{x+y}{\sqrt{\lambda}}$, $v := \frac{x-y}{\sqrt{\lambda+2\mu}}$. Again,

the function $\tilde{f}_c$ has to be harmonic in u and v, so we try to find a holomorphic function ϕ such that $\Im(\phi)$ satisfies the right boundary conditions. At this point, the difficulties begin. In the previous case, the domain of ϕ was a wedge, and there was an obvious holomorphic mapping from D_ϕ to $\mathcal{H}$. In this case, the domain of ϕ is a rhombus with exterior angles θ, $\pi - \theta$,θ and $\pi - \theta$, where $\theta := 2\arctan(\alpha)$. We can find no simple holomorphic mapping to $\mathcal{H}$, but the following theorem from complex analysis [3] guarantees its existence.

Theorem 9. Riemann Mapping Theorem
Let $U \subset \mathbb{C}$ be a simply connected open subset of the complex plane, and $U \neq \mathbb{C}$. Then there exists a biholomorphic mapping $\psi : U \to D$, where $D = \{z \in \mathbb{C} : |z| < 1\}$. If π is another such map, then there a map $g(z) = \frac{az+b}{cz+d}$ with $ad - bc = 1$, such that $\pi = g \circ \psi$.

This theorem states that there exists a biholomorphic mapping between D_ϕ and D, but it does not tell how to find it. In general, this is a very difficult problem. Fortunately, we are not in a general case. The domain D_ϕ is a polygon, and for polygons the mapping can be made explicit [4].

Theorem 10. Christoffel-Schwarz Mapping
Let $P \subset \mathbb{C}$ be a polygon with exterior angles $\theta_1, \ldots, \theta_n$. Then, for any set of different constants $a_1, \ldots, a_n \in \mathbb{R}$ there exists a constant $K \in \mathbb{C}$ such that the map

$$z \to \int_0^z \frac{K}{(w - a_1)^{\theta_1/\pi} \ldots (w - a_n)^{\theta_n/\pi}} dw \tag{75}$$

is a biholomorphic mapping from $\mathcal{H}$ onto P.

It is possible to choose one of the a_i in this last theorem equal to $\pm\infty$. Then the corresponding term is effectively absorbed in the constant K. Let us now try to use this theorem to find the holomorphic mapping, and see how far it gets us. In order to use the Christoffel-Schwarz mapping, we have to choose the $a_i \in \mathbb{R}$. We can forget about one of the vertices, so we choose $-1, 0, 1, \infty$. Then we need to solve the integral

$$\int_0^z \frac{K}{(w+1)^{\theta/\pi} w^{1-\theta/\pi} (w-1)^{\theta/\pi}} dw. \tag{76}$$

This is an elliptic integral, and the answer can not be given in terms of elementary functions. That is something we can live with, but there is a bigger problem. The Christoffel-Schwarz mapping gives a map from $\mathcal{H}$ to D_ϕ. We need to find a map the other way around, so we need to invert this holomorphic function. For a given holomorphic function, it is not easy to find its inverse. But in our case the holomorphic function itself is not even given, so inverting it is practically impossible. Therefore, we will have to use a numerical approximation to the integral in eqn. 76.

Before we delve into more complex analysis and numerical integration, let us think back to the original problem. We want to know what the probability is,

for a random walk started somewhere inside the triangle, to exit through the x-axis. We could of course use a numerical simulation to find this probability. We simulate the random walk, see where it exits, and repeat that procedure many times. Now the main disadvantage of numerical simulations is that the result will always be an approximation, and never exact. But even if we try to solve the problem exactly, we still need to evaluate the integral numerically. Because it is inevitable to use a simulation somewhere, there is no point in continuing the exact analysis.

4.4.2 Validity of the triangle approximation

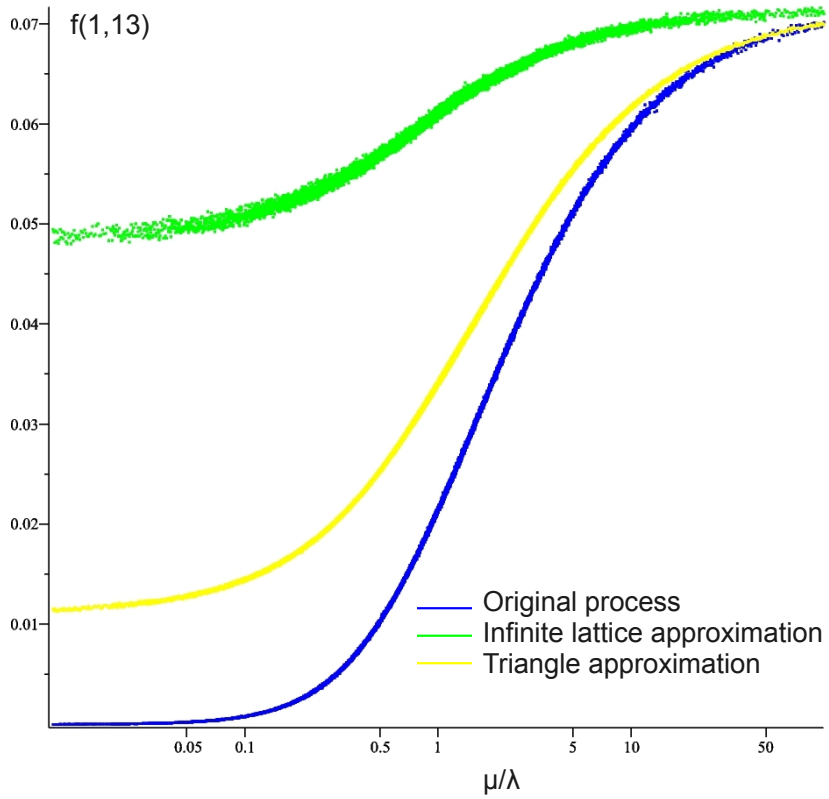

Figure 16: Graph of $f(1, 13)$, now also with the triangle approximation.

From figure 16 we can see that the triangle approximation is a lot better than the infinite lattice approximation. However, there is still some difference between the original process and the approximation. We will now explain why the triangle approximation is different from the original process, and why that difference is so small. Suppose that, in the original process, the defect makes a jump to the right. This means that, in the triangle approximation, x_t increases by 1 and y_t decreases by 1. Then both ends also make a jump to the right, so x_t decreases to its original value, as well as y_t. Afterwards, both x_t and y_t are back to normal, and nothing has changed. But that is not true. When the right end moves to the right, it gets closer to the end of the nucleosome. When it hits the end, there is a definite change in the dynamics of the system, because then the jump $y_t \to y_t + 1$ is prohibited.

There is only one conclusion we can draw from this previous argument. Knowing the value of x_t and y_t is not sufficient to determine the state of the

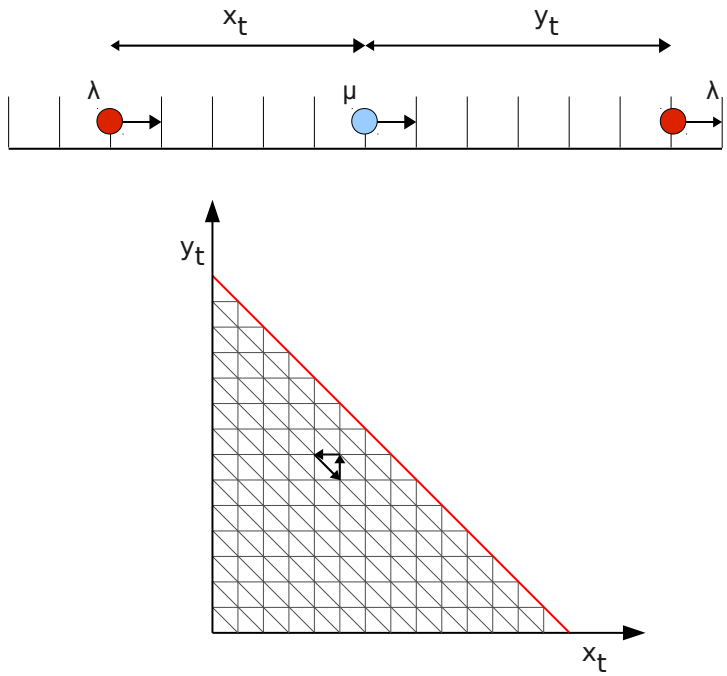

Figure 17: When both ends and the defect make a jump to the right, the triangle approximation makes a triangular loop. This means that the triangle approximation neglects any effects caused by such motion, and is therefore not perfect.

original process. So the triangle approximation is based on incomplete information, and therefore, it fails to describe the process exactly. However, to get a difference between the triangle approximation and the original Markov chain, we supposed that both ends make a jump. That means that such an event will occur with a probability of order λ^2. As λ is assumed to be small compared to μ, the effect of these events is negligible.

5 Multiple nucleosome Dynamics

Until now, we have only looked at the behaviour of a single nucleosome on a DNA chain. In short, this nucleosome makes a symmetric nearest neighbour random walk on the chain with some diffusion constant D. However, it makes no sense to study a single nucleosome in a vacuum. In a human cell, there are many other proteins with which the nucleosome interacts. Not in the least, the nucleosomes interact with each other.

Suppose two nucleosomes attached to a single DNA chain happen to diffuse

towards each other. These nucleosomes are large proteins, which cannot pass each other without letting go of the DNA chain. But that will not happen, because they are strongly attached to the DNA. So these nucleosomes hit, stay close for a while, and then they diffuse away. Now envision a DNA chain with many nucleosomes attached to it. All of these nucleosomes perform a symmetric random walk, all with the same diffusion constant, and they never interchange positions.

We will model the behaviour of nucleosomes on a DNA chain as a Markov process. The DNA is represented by a line of binding sites, which can be occupied by nucleosomes. If a site is occupied, there will be a region of 14 sites around it which cannot be occupied anymore. For the moment, let us be satisfied with requiring that there is never more than one nucleosome per binding site. This means the nucleosomes can get quite close to each other, but they cannot change their relative positioning. The entire state of the DNA chain and the nucleosomes can be described by a sequence of zeros and ones, where a 1 indicates that a certain site is occupied, and a 0 indicates it isn't. In principle, the DNA has only a finite number of base pairs, and this should be a finite sequence. In practice, we are working on a length scale of single base pairs, whereas a DNA chain can be millions to billions of base pairs long. Therefore, it is probably not very harmful to assume that the sequence is infinite. The dynamics of the nucleosomes are then described by the **symmetric exclusion process**. It is discussed in great detail by Liggett [9] and Seppäläinen [11].

Definition 7. *The symmetric exclusion process is the stochastic process with state space* $\{0,1\}^{\mathbb{Z}} = \{\eta : \mathbb{Z} \to \{0,1\}\}$ *and generator*

$$Lf(\eta) = \frac{1}{2}\sum_{x\in\mathbb{Z}}[f(\eta_{x,x+1}) - f(\eta)], \tag{77}$$

where $\eta_{x,y}$ *denotes the configuration after interchanging* $\eta(x)$ *and* $\eta(y)$*:*

$$\eta_{x,y}(z) = \begin{cases} \eta(z) & z \notin \{x,y\} \\ \eta(x) & z = y \\ \eta(y) & z = x \end{cases} \tag{78}$$

5.1 The asymmetric tagged particle process

The main question raised at the beginning of this article is: How can a DNA chain be transcribed, when there are so many nucleosomes attached to it? Part of the answer has already been given. The full information stored on the DNA is never visible, because the DNA is covered in nucleosomes. But these nucleosomes move around, and any specific site becomes available every once in a while. So an RNA polymerase can attach itself to the DNA at any specific point, but only after the site has cleared.

But what happens after the polymerase is attached? The polymerase wants to move forward, but every step it makes, it has to wait before the site in front of it is clear. Only when the nucleosome in front of the polymerase moves

forward, the polymerase can advance. And after that, the nucleosome can never go back to its previous position. So this nucleosome gets pushed forward by the polymerase in a special way. After a few steps, this nucleosome hits a second one, which will then have to move forward too. In this way, nucleosomes will accumulate in front of the polymerase, and the entire bulk slows down.

It is clear that whenever an RNA polymerase is attached to a DNA chain, it gets slowed down by a bulk of nucleosomes. It is not clear, however, how much and how fast it slows down. In the end, we will show that the average speed of the polymerase decreases like $v(t) \propto t^{-1/2}$. But first we have to adapt our model to include the polymerase.

To take the polymerase into account, we modify the symmetric exclusion process by enforcing that one of the particles only jumps forward, and never backwards. This special particle represents the polymerase, and as such, it will move at a different rate than the nucleosomes. Of course, the polymerase is still hindered by the nucleosomes, so it can only move if the site in front of it is unoccupied. We normalise time such that the rate of movement of the nucleosomes is 1. The rate at which the polymerase tries to move is called α.

Whenever the polymerase would move forward, we shift all other particles backwards instead. In this way, the dynamics are exactly the same, but the polymerase remains at its original position, which we assume to be the origin. Finally, it only matters what happens in front of the polymerase, so we discard all information about the particles behind it. We call this new process the **asymmetric tagged particle process**.

5.1.1 Invariant measures

Definition 8. *The asymmetric tagged particle process is the process with state space* $\Omega := \{\eta \in \{0,1\}^{\mathbb{N}} : \eta(0) = 1\}$ *and generator*

$$Lf(\eta) = \frac{1}{2}\sum_{x=1}^{\infty}[f(\eta_{x,x+1}) - f(\eta)] + \alpha(1-\eta(1))[f(\tau_1\eta) - f(\eta)], \tag{79}$$

where

$$\tau_1\eta(x) = \begin{cases} 1 & x = 0 \\ \eta(x+1) & x \geq 1 \end{cases} \tag{80}$$

Now that the process is accurately described, let us turn to the question at hand: At what speed does the polymerase move? Although the tagged particle stays fixed in the origin, the information about the relative position of the particle is not lost. We can simply count the number of shifts that have occurred. So the speed of the polymerase is equal to the rate at which shifts occur, which is $\alpha(1-\eta(1))$.

Within the model, it is obvious that there will be infinitely many particles attached to the DNA. If we start the process from a measure for which the particle number is almost surely infinite, then that will remain so for all time. If the process converges a limiting measure, that will be an invariant measure with infinitely many particles.

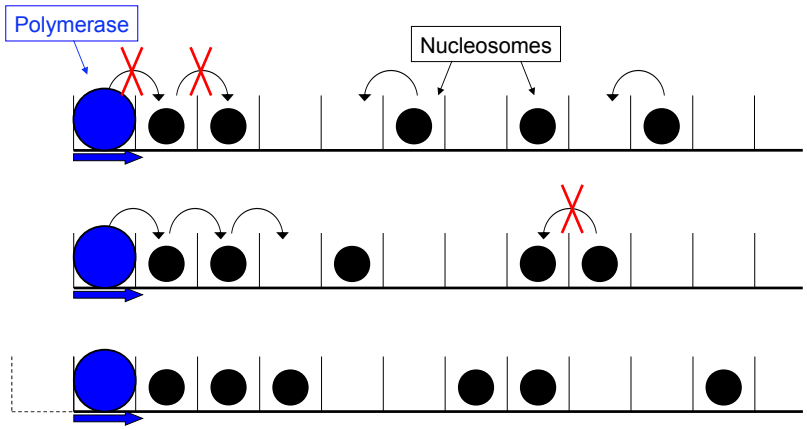

Figure 18: The Asymmetric Tagged Particle Process. The nucleosomes all make a symmetric random walk, and the polymerase moves forward with rate α. Whenever any jump would cause two particles to be at the same site, that jump is prohibited. Also, when the polymerase moves, the entire configuration is shifted backwards, and the first site is discarded.

Theorem 11. *The only invariant measure of the process such that* $\sum_{x=1}^{\infty} \eta(x) = \infty$ *almost surely is the Bernoulli product measure with parameter* 1*, i.e.*

$$\forall x : \nu_1\{\eta(x) = 1\} = 1. \tag{81}$$

Proof. The proof of this theorem will be easier if we switch to a different representation of the process. Let $\zeta_t(j)$ denote the number of unoccupied sites between the j-th and the $j+1$-th particle. Formally, for a given $\eta \in \Omega$, let $f_\eta(x) = \eta(x) \sum_{i=0}^{x} \eta(i)$. Then

$$\zeta(j) = \begin{cases} f_\eta^{-1}(j+1) - f_\eta^{-1}(j) - 1 & j > 0 \\ f_\eta^{-1}(1) - 1 & j = 0. \end{cases} \tag{82}$$

Remark that $f_\eta : \mathbb{N} \to \mathbb{N}$ is surjective because $\sum_{x=1}^{\infty} \eta(x) = \infty$. The transition rules in this representation take a different form than before. If the j-th particle makes a jump to the right, then $\zeta(j)$ decreases by one, and $\zeta(j-1)$ increases by one. Of course, this can only happen if the site in front of the particle is empty, in other words, if $\zeta(j) \neq 0$. If the polymerase moves, $\zeta(0)$ decreases by one, and that's all that happens.

The process in terms of ζ_t can also be described as a stochastic process. The state space of the process is $\mathbb{N}^{\mathbb{N}} := \{\zeta : \mathbb{N} \to \mathbb{N}\}$. In order to give the generator we need some definitions. $g(x) := 1_{x>0}$, ∂_x is the configuration with no particles except at site x, and summation is done pointwise. Furthermore,

$$L_s f(\zeta) := g(\zeta(0))\{f(\zeta - \partial_0) - f(\zeta)\}, \tag{83}$$

$$L_{x,y}f(\zeta) := g(\zeta(x))\{f(\zeta - \partial_x + \partial_y) - f(\zeta)\}, \tag{84}$$

The generator is then given by

$$L = \alpha L_s + \frac{1}{2}\sum_{x=0}^{\infty}[L_{x,x+1} + L_{x+1,x}]. \tag{85}$$

This representation of the process is called the **zero range process**.

Let μ be an invariant measure for this new process. This means that $\mu S(t) = \mu$ for all $t \geq 0$. So for all continuous functions $f : \mathbb{N}^{\mathbb{N}} \to \mathbb{R}$ we have

$$\int S(t)f d\mu = \int f d(\mu S(t)) = \int f d\mu. \tag{86}$$

This does not depend on t, so the derivative $\frac{d}{dt}\int S(t)f d\mu$ vanishes.

$$\int Lf d\mu = \int \frac{d}{dt}S(t)f d\mu = \frac{d}{dt}\int S(t)f d\mu = \frac{d}{dt}\int f d\mu = 0. \tag{87}$$

This equation holds for all continuous functions f, but in order to get something useful out of it, we will have to make a clever choice. This turn out to be $f_x(\zeta) = g(\zeta(x))$. Then we have that $L_s f_x(\zeta) = -\delta_{x,0}f_x(\zeta)$, and $L_{x,y}f_z(\zeta) = f_x(\zeta)(\delta_{y,z} - \delta_{x,z})$. If we introduce $p_x = P^\mu\{\zeta(x) \neq 0\} = \int f_x d\mu$, and use eqn. 87, we get

$$\forall x \geq 1 : 2p_x = p_{x+1} + p_{x-1}, \tag{88}$$

$$p_1 = (1 + 2\alpha)p_0. \tag{89}$$

The unique solution to this system is $p_x = (2x + 1)p_0$, but this solution is unbounded if $p_0 \neq 0$. So $p_0 = 0$, which means $p_x = 0$ for all x. So there are no holes at all in the system, which means that the configuration is entirely filled. □

5.1.2 Speed of the polymerase

We have argued that the speed of the polymerase is proportional to the probability that $\eta(1) = 0$. Consequently, as t tends to infinity, this probability reduces to zero and the polymerase slows down. Now we investigate how fast the speed of the polymerase decreases. The results in this section are similar to results obtained by Olla and Landim [20].

The exact behaviour of the system will of course depend on its starting distribution. When the polymerase attaches to the DNA chain, the nucleosomes have not felt its influence yet, so we assume they are in equilibrium. The equilibrium states for the symmetric exclusion process are Bernoulli product measures ν_ρ, defined by $\nu_\rho\{\eta(x) = 1\} = \rho$ for all x. So the starting distribution will be ν_ρ for some $0 < \rho < 1$.

We will work again in a continuum limit. As before, we rescale the space coordinate $x \to \frac{x}{N}$, while simultaneously scaling time by $t \to N^2 t$. We then

suppose that the measures

$$\pi_t^N := \frac{1}{N} \sum_{x \in \mathbb{Z}} \eta_{N^2 t}(x) \delta \left(\frac{x}{N} \right) \tag{90}$$

converge in probability to a macroscopic profile $\Psi(x,t)dx$. Here δ denotes the Dirac measure, and dx the Lebesgue measure. This macroscopic profile will then obey the differential equation

$$\frac{\partial \Psi}{\partial t} = \frac{1}{2} \frac{\partial^2 \Psi}{\partial x^2} - v(t) \frac{\partial \Psi}{\partial x}, \tag{91}$$

with boundary conditions $\Psi(0,t) = 1, \Psi(x,0) = \rho$, and

$$v(t) := \alpha \frac{\partial \Psi}{\partial x}(0,t). \tag{92}$$

We do not attempt to prove these facts here, because the proof is rather long and complicated. In [8], an analogous result is proven for other types of exclusion processes. We expect that a similar technique can be used to prove the convergence in this case. So let us suppose that the continuum limit works, and try to solve the differential equation (eqn. 91).

Before we begin, let us remark that the function $v(t)$ can indeed be associated to the speed of the polymerase $\alpha(1 - \eta_t(1)$. To be more precise, if the system starts from a Bernouilli measure with parameter $0 < \rho < 1$, and we define $\tilde{v}_t = \alpha(1 - \eta_t(1)$, then

$$\frac{1}{t} \int_0^t \tilde{v}_s ds \sim v(t), \tag{93}$$

as $t \to \infty$.

Theorem 12. *There exists a constant C such that*

$$v(t) = \frac{C}{\sqrt{t}}, \tag{94}$$

and

$$\operatorname{erfc}(C\sqrt{2}) C e^{2C^2} = \sqrt{\frac{2}{\pi}} \alpha (1 - \rho), \tag{95}$$

where erfc *denotes the complementary error function:* $\operatorname{erfc}(x) := \frac{2}{\sqrt{\pi}} \int_x^\infty e^{-y^2} dy$

Proof. We will use the method of separation of variables. If we can find new coordinates $y(x,t)$ and $\tau(x,t)$ such that

$$\frac{\partial t}{\partial y} = 0, \qquad \frac{\partial t}{\partial \tau} = 1,$$
$$\frac{\partial x}{\partial y} = 1, \qquad \frac{\partial x}{\partial \tau} = -v(t),$$

then the differential equation in terms of y and τ simplifies to the heat equation

$$\frac{\partial \Psi}{\partial \tau} = \frac{1}{2}\frac{\partial^2 \Psi}{\partial y^2}. \tag{96}$$

From the conditions on y and τ it follows that $\tau = t$ and $y = x + \int v(t)dt$. Here we will adopt a somewhat circular procedure. We assume that $v(t) = Ct^{-1/2}$ for some C, and use that to find a solution to the differential equation (eqn. 91), with the right boundary conditions. From that solution, we can calculate $v(t)$ by using eqn. 92, and see whether it is consistent with our assumption. It will turn out that the consistency condition is indeed satisfied, and we can even determine the constant C. So then we have found a solution to the differential equation

$$\frac{\partial \Psi}{\partial t} = \frac{1}{2}\frac{\partial^2 \Psi}{\partial x^2} - \alpha \frac{\partial \Psi}{\partial x}\Big|_{x=0}\frac{\partial \Psi}{\partial x} \tag{97}$$

This does not prove that $v(t)$ has the right form, because there could be other solutions to the differential equation. But, most of the time, the solution to a differential equation is uniquely determined by its boundary conditions. So we stop looking for another solution. With this idea in mind, we return to the calculations. Assume that $v(t) = Ct^{-1/2}$, so that $y = x + 2C\sqrt{t}$. A complete set of solutions of the heat equation is given by

$$\Psi_a(y,t) = \frac{1}{\sqrt{2\pi t}} e^{-\frac{(y-a)^2}{2t}}. \tag{98}$$

We can substitute y and then integrate over a to find

$$\Psi(x,t) = \rho + (1-\rho)\frac{\operatorname{erfc}\left(\frac{x}{\sqrt{2t}} + C\sqrt{2}\right)}{\operatorname{erfc}(C\sqrt{2})}. \tag{99}$$

We can see this solution satisfies all the boundary conditions

$$\lim_{t\to 0} \Psi(x,t) = \rho, \tag{100}$$

$$\lim_{x\to 0} \Psi(x,t) = 1, \tag{101}$$

$$\lim_{t\to \infty} \Psi(x,t) = 1, \tag{102}$$

$$\lim_{x\to \infty} \Psi(x,t) = \rho. \tag{103}$$

$$\tag{104}$$

And furthermore that

$$\frac{\partial \Psi}{\partial x}(0,t) = -\sqrt{\frac{2}{\pi}}(1-\rho)\frac{e^{-2C^2}}{\operatorname{erfc}(C\sqrt{2})} \cdot \frac{1}{\sqrt{t}} \tag{105}$$

So the consistency equation (eqn. 92) is satisfied if and only if $\operatorname{erfc}(C\sqrt{2})Ce^{2C^2} = \sqrt{\frac{2}{\pi}}\alpha(1-\rho)$. □

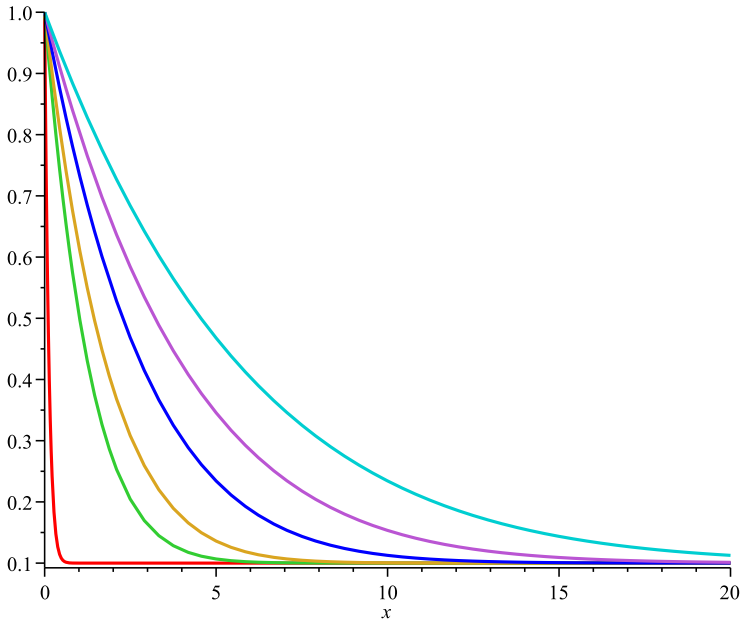

Figure 19: Graph of $\Psi(x,t)$ for $t = 0, 10, 20, 50, 100, 200$. As $t \to \infty$, the density of nucleosomes at the beginning increases to 1.

Now something very strange happens. The theorem states that $v(t)$ is inversely proportional to the square root of time, and that the proportionality constant satisfies some consistency equation. But this consistency equation is not really what we expect of it. The function on the left hand side is strictly increasing and continuous, so it is injective. Therefore, the constant C is uniquely determined by α and ρ. The function is not surjective, however, as

$$\lim_{C\to\infty} \operatorname{erfc}(C\sqrt{2})Ce^{2C^2} = \sqrt{\frac{1}{2\pi}}, \tag{106}$$

which implies that the solution only works if $2\alpha(1-\rho) < 1$.

It is not clear why our calculation fails if α becomes larger than $\frac{1}{2(1-\rho)}$. Suppose we consider the asymmetric tagged particle process, with $\alpha \gg 1$. This means that the polymerase is continuously trying to move forward. Once the first nucleosome makes a step to the right, the polymerase immediately follows, so it can never turn back. Then the first nucleosome effectively takes on the role of the polymerase, but it moves forward with rate $\frac{1}{2}$, instead of α. We expect that, if $\alpha \to \infty$, the density evolves completely like it would if $\alpha = \frac{1}{2}$ (at least for large x).

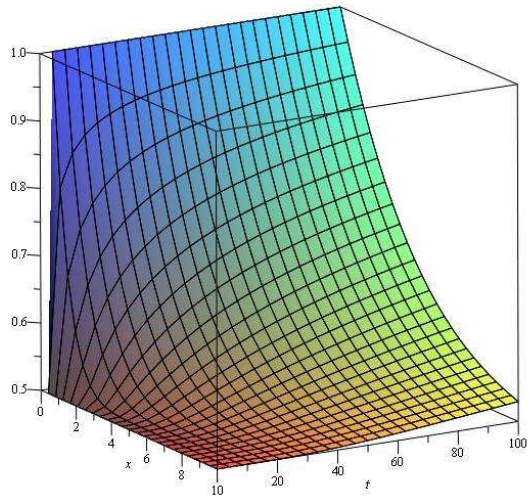

Figure 20: 3-dimensional graph of $\Psi(x,t)$. The limits in eqn. 100 are clearly visible.

5.2 A more realistic process

By now we know that after a polymerase has attached to a DNA chain, it gets hindered by the nucleosomes on that DNA chain so much that it is practically unable to move forward. The speed of the polymerase decreases like $t^{-1/2}$, so its position is proportional to $t^{1/2}$.

Our model predicts that a polymerase is unable to transcribe a DNA chain, because of the nucleosomes. If this were true, all of the information stored on the DNA could not be read. Therefore, the model has to be incorrect. That is, we have to include another ingredient to make it work.

At this point, we would like to describe an experiment conducted by ten Heggeler-Bodier ([18]). There a DNA chain with nucleosomes attached to it was put into solution along with RNA polymerase. The polymerase attached to the chain, as it should, and then it started to move. While it was moving forward, the nucleosomes in front of it detached from the DNA. Curiously, the nucleosomes did not reassemble after the polymerase had moved through. Then the same experiment was repeated, but with some cell nucleus extract added to the solution. The same thing happened, but now the nucleosome did reassemble

on the DNA chain.

This experiment is relevant to us because it shows that the polymerase is able to push nucleosomes forward, and even disassemble them from the DNA, without the help of other proteins. In order to see how this can happen, we need to go back to single nucleosome dynamics. Let us focus our attention to the first nucleosome after the polymerase. When the first stretch of DNA on the side of the polymerase unwraps, those 10 base pairs become available to the polymerase. If the polymerase moves forward before the DNA stretch rewraps, the nucleosome has effectively lost a binding site. If this happens fourteen times in a row, the nucleosome will have completely disassembled, and it falls off. Of course, the rate with which a DNA stretch rewraps is a lot larger than the rate with which it unwraps, especially for the inner binding sites. But in this case the DNA stretch cannot rewrap anymore, so it is only a matter of time before the nucleosome falls off.

There are quite a few things necessary for a nucleosome to disassemble. First of all, the first binding site has to open. Then the polymerase has to move before it closes, which will not always be the case. And then this process has to repeat itself fourteen times. So the rate at which the first nucleosome falls off will be very small, but nevertheless, it will be nonzero.

Let us now turn to calculations. We start with defining the new generator, then we calculate the equilibrium distribution.

Definition 9. *The generator of the new process is given by*

$$Lf(\eta) = \sum_{x=1}^{\infty}[f(\eta^{(x,x+1)}) - f(\eta)] + \alpha(1-\eta(1))[f(\tau_1\eta) - f(\eta)] + \varepsilon[f(\eta^\dagger) - f(\eta)]. \tag{107}$$

We use the same notations as before, and

$$\eta^\dagger(x) := \begin{cases} 0 & x = 1 \\ \eta(x) & x \neq 1 \end{cases} \tag{108}$$

Furthermore, ε is assumed to be small compared to α, and 1.

Lemma 10. *If μ is an invariant measure for this new process, then*

$$\mu\{\eta(1) = 1\} < 1. \tag{109}$$

Proof. As before, let $f_x(\eta) := \eta(x)$, and let μ denote the equilibrium distribution. Again, it has to hold that $\int Lf_x d\mu = 0$. Let us now define $p_x = \int f_x d\mu$, and $c(x) = \mathrm{cov}_\mu(\eta(x), \eta(1))$. If we now plug in the explicit form of the generator, we get

$$\forall x \geq 2 : p_{x+1} + p_{x-1} - 2p_x + \alpha(1-p_1)(p_{x+1} - p_x) - \alpha(c(x+1) - c(x)) = 0, \tag{110}$$

$$p_2 - p_1 + \alpha(p_2(1-p_1) + c(2)) - \varepsilon p_1 = 0. \tag{111}$$

Suppose now that $p_1 = 1$. Then $\eta(1) = 1$ almost surely, and consequently $c(x) = 0$ for all $x \geq 2$. If we then look at the second equation, we get that $p_2 = 1 + \varepsilon > 1$. That is a contradiction, so p_1 cannot be exactly 1. □

This lemma is actually great news, because it proves that the speed of the polymerase will not decrease to zero in the new model. There will still be nucleosomes accumulating in front of the polymerase, but the size of this accumulation will remain bounded. As a last order of business, let us estimate the speed at which the polymerase moves in this new description.

Theorem 13. *If the system starts out from the Bernoulli measure ν_ρ,*

$$\lim_{t\to\infty} v(t) = \frac{\varepsilon}{\rho} + O(\varepsilon^2). \tag{112}$$

Proof. From earlier observations, we know that $v(t) = \alpha(1 - P\{\eta_t(1) = 1\})$, which implies that $v := \lim_{t\to\infty} v(t) = \alpha(1 - p_1)$. We would like to calculate $\beta := 1 - p_1$ from equations 110 and 111, but there is a contribution $c(x)$ which may be problematic. But we know that the covariance of two random variables is smaller in magnitude than the product of their respective standard deviations. We expect p_1 to be close to unity, so the variance of $\eta(1)$ will be very small. Therefore, $c(x)$ will be negligible, so we assume $c(x) = 0$ for all $x \geq 2$. Then eqn. 110 reduces to a linear recurrence relation

$$p_{x+1} + p_{x-1} - 2p_x + \alpha\beta(p_{x+1} - p_x) = 0. \tag{113}$$

for all $x \leq 2$. This equation has linearly independent solutions $p_x = 1$ and $p_x = (1 + \alpha\beta)^{-x}$. The full solution is a linear combination of those:

$$\forall x \geq 1 : p_x = \rho + (1 - \rho - \beta)(1 + \alpha\beta)^{-x+1}. \tag{114}$$

Remark that $\lim_{x\to\infty} p_x = \rho$, so ρ is indeed the density of nucleosomes at infinity. The other equation, eqn. 111 allows us to determine β, which results in

$$\beta = \frac{\varepsilon}{\alpha\rho + \varepsilon} = \frac{\varepsilon}{\alpha\rho} + O(\varepsilon^2), \tag{115}$$

and finally,

$$v = \alpha\beta = \frac{\varepsilon}{\rho} + O(\varepsilon^2). \tag{116}$$

□

The most remarkable thing about theorem 13 is that the asymptotic speed of the polymerase does not depend on α. That is, as long as α is much larger than 1. This is not illogical. If the polymerase pushes harder on the nucleosomes, there will be a larger accumulation, and the polymerase gets dragged more. In the end, the effects balance out precisely.

In a way, this is the only result that could have come out of the analysis. The polymerase has to push the nucleosomes forward, or it has to disassemble them. The symmetric exclusion process admits no flux of particles, so the polymerase has to detach all the nucleosomes it encounters. The speed at which it detaches nucleosomes is ε, and the number it encounters is ρ per unit length. So the time it needs to advance a unit length is $\frac{\rho}{\varepsilon}$, or equivalently, its speed is $\frac{\varepsilon}{\rho}$.

6 Concluding Remarks

By now we understand quite well what happens when a DNA chain is transcribed by a polymerase. The nucleosomes on the DNA wiggle around, until the promotor site for the polymerase is free, and the polymerase binds to the DNA. It then starts to move forward, carrying some of the nucleosomes in front of it along. As it progresses, more and more nucleosomes will block its path, and the polymerase slows down. Eventually, the first nucleosome gets pushed so hard that it falls off the DNA. Then the polymerase can advance, until it encounters the second nucleosome, and the entire story repeats itself.

There are two crucial ingredients for this to happen. First of all, the sliding. If the nucleosome were not sliding, the promotor site of the polymerase would never be totally free, and nothing would happen. Second, there is the breathing. The first nucleosome in front of the polymerase has to be breathing, otherwise it would not fall off the DNA. Then the polymerase gets dragged almost to a full stop.

We have discussed both these processes, but there is a lot we did not discuss. In the single nucleosome part, we used a very simplified model of the breathing process, which can definitely be improved on. Instead of assuming that the rate with which a nucleosome unwraps and rewraps are the same, a better approximation would be to assume that the rate of unwrapping is always a certain factor smaller than the rewrapping rate. Also, in the multiple nucleosome part, we ignored the size of the nucleosomes, which means we also ignored all excluded volume effects.

Perhaps the most important factor that we left out is the DNA itself. The binding and bending energies of a DNA chain are strongly dependent on its sequence ([23, 16, 21]). This means that the nucleosomes do not make a symmetric random walk along the DNA, but a random walk governed by a potential landscape. Therefore, we should also consider the symmetric exclusion process in a potential landscape. Although this is a very interesting process to study, both from a physical as well as a mathematical point of view, doing so would take us too far for this thesis. That might be a good topic for future study.

7 Appendix

Definition 10. *In the following chapter, x_t will denote a Markov process on X, which is a compact metric space with a sigma-algebra generated by Borel subsets. The semigroup associated to this process is called $S(t)$, and the generator L.*

7.1 Relation between hitting times and Dirichlet problems

Definition 11. *A **hitting time** for a Markov process x_t is a time τ of the form $\tau_A = \inf\{t \geq 0 | x_t \in A\}$, where A is a Borel set.*

Lemma 11. *If x_t is a Feller process and $A \subset X$ a closed set, then the hitting time τ_A is a (not necessarily bounded) stopping time.*

Proof. If $\tau_A = t$, there exists a decreasing sequence of times $t_n \geq t$ such that $x_{t_n} \in A$ for all n and $t_n \to t$ as $n \to \infty$. The trajectory of a Feller process is almost surely right continuous. This means that $x_t = \lim_{n\to\infty} x_{t_n}$, and because A is closed, $x_t \in A$. So the question whether $\tau_A \leq t$ is determined by the trajectory up to time t. □

Lemma 12. *Let $\tau = \tau_A$ be a hitting time for a Feller process x_t, with $A \subset X$ closed. Let L be the generator of x_t, and $D(L)$ its domain. Then the stopped process $y_t = x_{\min(t,\tau)}$ is also a Markov process, and its generator L' satisfies*

$$L'g(x) = \begin{cases} Lg(x) & : x \in A \\ 0 & : x \notin A \end{cases} \tag{117}$$

for all $g \in D(L)$.

Proof. First of all, remark that the evolution of the stopped process after time t depends only on x_t and whether $\tau \leq t$. For a hitting time, however, the question whether $\tau \leq t$ is also completely determined by x_t. So the future of the process only depends on the position at time t, or in other words, it is Markov.

It is trivial to show that $L'g(x) = 0$ if $\tau = 0$. Therefore, we will assume $\tau > 0$. By definition of the generator we have that:

$$L'g(x) - Lg(x) = \lim_{t\downarrow 0} \frac{E[g(x_{\min(t,\tau)}) - g(x_t)]}{t}. \tag{118}$$

By general theory, we know that $M_t = g(x_t) - g(x_0) - \int_0^t Lg(x_s)ds$ is a martingale. Because τ is a stopping time for the process, $\tau' = \min(\tau, t)$ is also a stopping time for any fixed $t > 0$. This τ' is bounded above by t, so we may apply the optional stopping theorem to obtain:

$$E[g(x_{\tau'}) - g(x_t)] = E[\int_{\tau'}^{t} Lg(x_s)ds]. \tag{119}$$

Therefore, we can estimate

$$|E[g(x_{\tau'}) - g(x_t)]| \leq E[t - \tau'] \sup_{x\in X} |Lg(x)|. \tag{120}$$

The generator L maps $D(L)$ into $C(X)$, so Lg is continuous. This means Lg is a continuous function on a compact space, which is necessarily bounded. Furthermore, by using the law of total expectation, we can see that

$$\frac{E[t - \tau']}{t} = P(\tau < t)\frac{E[t - \tau | \tau < t]}{t}. \tag{121}$$

In the limit $t \downarrow 0$, the probability $P(\tau < t)$ decreases to 0, while the second term is bounded. In conclusion, we have proved that

$$\frac{E[g(x_{\min(t,\tau)}) - g(x_t)]}{t} \text{ as } t \to 0, \tag{122}$$

which is exactly what we need. □

Theorem 14. *Let $A, B \subset X$ be disjoint closed subsets of X, and $\tau = \tau_{A\cup B}$. It is assumed that A and B are such that $P(\tau < \infty) = 1$. Denote with $f(x)$ the probability that the process, starting from $x \in X$, reaches the set A before B, i.e.*

$$f(x) = P^x\{x_\tau \in A\}. \tag{123}$$

Then, the following hold:

$$\forall x \in A: \quad f(x) = 1, \tag{124}$$

$$\forall x \in B: \quad f(x) = 0, \tag{125}$$

$$\forall x \in X\backslash(A \cup B): \quad Lf(x) = 0. \tag{126}$$

Proof. The first two conditions are trivial, the third requires some work. Consider the stopped process $y_t = x_{\min(t,\tau)}$, and note that it holds that

$$f(x) = P^x\{x_\tau \in A\} = E^x[1_A(x_\tau)] = \lim_{t\to\infty} E^x[1_A(x_{\min(t,\tau)})] = \lim_{t\to\infty} E^x[1_A(y_t)]. \tag{127}$$

If we apply the generator of the stopped process to f, we see that

$$L'f(x) = \lim_{t\to\infty} L'S'(t)1_A(x) = \lim_{t\to\infty} \frac{d}{dt}S'(t)1_A(x) = 0. \tag{128}$$

As shown in the previous lemma, the generator L' of the stopped process satisfies

$$L'g(x) = \begin{cases} Lg(x) & : x \notin A \cup B \\ 0 & : x \in A \cup B \end{cases} \tag{129}$$

for all $g \in D(L)$. Therefore, the condition $L'f(x) = 0$ is automatically met for $x \in A \cup B$, but for $x \in X\backslash(A \cup B)$ it means that $Lf(x) = 0$. □

Remark 1. *The proof above is actually not correct. We cannot apply the generator to an indicator function, because that function is not continuous, in general. However, given the correct regularity conditions on $S(t)$, A and B, this does not pose a real problem. In our case, these regularity conditions will always be satisfied. A more rigorous version of this theorem is proven in [13].*

7.2 Dirichlet problems for finite Markov chains

A very important class of Markov processes are Continuous time Markov chains with finite state space. The generator then takes on a relatively simple form, which makes the Dirichlet problem easier to solve.

Definition 12. *In this section, x_t will be a reversible Markov chain with finite state space X and transition rates $r_{x\to y}$. The generator of this process is given by*

$$Lf(x) := \sum_{y\in X} r_{x\to y}[f(y) - f(x)], \tag{130}$$

for all $f : X \to \mathbb{R}$. Now let π denote the stationary distribution of this process, and define $d_x := \#\{y : r_{x\to y} \neq 0\}$. We will assume that the rates are normalized such that $\forall x \in X : \sum_{y\in X} r_{x\to y} = \frac{d_x}{\pi(x)}$, and that $d_x \neq 0$ for all $x \in X$. Note that rescaling the rates results in another process viewed in continuous time, but with the same jump process.

Definition 13. *The **edge weights** are given by $w_{xy} := \pi(x)r_{x\to y}$. From this definition it immediately follows that*

$$\forall x, y \in V : w_{xy} = w_{yx}, \text{ and} \tag{131}$$

$$\forall x \in V : \textstyle\sum_{y\in V} w_{xy} = d_x. \tag{132}$$

We can now define the graph of X by $x \sim y \iff w_{xy} \neq 0$.

Definition 14. *For any $S \subset X$, let F_S denote the real vector space of all functions from S to $\mathbb{R}$. Remark that the generator L now acts as a linear map from f_X to itself. For $S' \subset S$, and $f \in F_S$, $f_{S'}$ denotes the function f restricted to S'. Also, for a matrix A defined on S, we define $A_{S'}$ as the submatrix consisting of those elements $A(x, y)$, where $x, y \in S'$.*

Definition 15. *For a given $S \subset X$, let δS denote the set of all vertices not in S, but adjacent to S: $\delta S = \{x \notin S : \exists y \in S : x \sim y\}$. The **Dirichlet problem** is to find, for a given $S \subset X$ and $\sigma : \delta S \to \mathbb{R}$, a function $f \in F_{S\cup\delta S}$ such that:*

$$(Lf)_S = 0, \tag{133}$$

$$f_{\delta S} = \sigma. \tag{134}$$

Remark 2. *If S is not connected, we can solve the Dirichlet problem for each of its connected components, then put those solutions together. Therefore, we will from now on assume that S is connected.*

Remark 3. *We can rewrite the generator to see that $(Lf)_S = 0 \iff (\Delta f)_S = 0$, where*

$$\Delta(x, y) = \begin{cases} 1 - \frac{w_{xx}}{d_x} & : x = y \\ -\frac{w_{xy}}{d_x} & : x \neq y \end{cases} \quad \forall x, y \in S \cup \delta S. \tag{135}$$

Note that this matrix is in general not symmetric. However, it is similar to a symmetric matrix $\mathbb{L} = T^{1/2}\Delta T^{-1/2}$, where $T(x, y) = 1_{x=y}d_x$. This matrix is called the **normalized Laplacian** of S. Of course, any submatrix of $\mathbb{L}$ is also symmetric. Therefore, there exists an orthonormal basis of eigenfunctions of $\mathbb{L}_S$.

Theorem 15. *Let $\{(\phi_i, \lambda_i), i \in \mathcal{I}\}$ be such an orthonormal eigensystem of $\mathbb{L}_S$. The solution to the Dirichlet problem is then given by:*

$$f(x) = \sum_{i\in\mathcal{I}} \frac{1}{\lambda_i} \sum_{\substack{z\in S \\ z\sim y\in\delta S}} w_{yz}\phi_i(z)\sigma(y)d_z^{-1/2}d_x^{-1/2}\phi_i(x). \tag{136}$$

Proof. We need to solve $(\mathbb{L}g)_S = 0$, where $g = T^{1/2}f$. Observe now that for all $x \in S$:

$$(\mathbb{L}g)_S(x) = \sum_{y \in S \cup \delta S} \mathbb{L}(x,y)g(y) = \mathbb{L}_S g_S(x) + \sum_{y \in \delta S} \mathbb{L}(x,y)g(y) = 0. \tag{137}$$

So we need to solve the equation $\mathbb{L}_S g_S = \alpha$, where

$$\alpha(x) = -\sum_{y \in \delta S} \mathbb{L}(x,y)g(y) = \sum_{y \in \delta S} d_x^{-1/2} w_{xy} \sigma(y) \tag{138}$$

The matrix-tree theorem [6] states that the number of spanning trees of S equals $\det \mathbb{L}_S \times \prod_{x \in S} d_x$. Because S is assumed to be connected, there is at least one spanning tree, so the determinant of $\mathbb{L}_S$ is nonzero. In other words, $\mathbb{L}_S$ is invertible. If we use that $\mathbb{L}_S = \sum_{i \in \mathcal{I}} \lambda_i \phi_i {\phi_i}^T$, we can compute

$${\mathbb{L}_S}^{-1}\alpha = \sum_{i \in \mathcal{I}} \frac{1}{\lambda_i} \phi_i {\phi_i}^T \alpha, \tag{139}$$

$$f_S = T^{-1/2} g_S = \sum_{i \in \mathcal{I}} \frac{1}{\lambda_i} T^{-1/2} \phi_i {\phi_i}^T \alpha. \tag{140}$$

If we work out this last expression, we see that :

$$f_S(x) = \sum_{i \in \mathcal{I}} \frac{1}{\lambda_i} \sum_{y \in S} \sum_{z \in \delta S} \sigma(y) w_{yz} \phi_i(z) d_y^{-1/2} d_x^{-1/2} \phi_i(x) \tag{141}$$

□

8 Acknowledgements

I would like to thank my advisors, prof. Helmut Schiessel and prof. Frank Redig, for helping throughout the whole project. Also, I would like to thank Marcel de Jeu, Onno van Gaans, and Johan van Leeuwarden for their assistance with some parts that I failed to understand myself.

References

[1] Billingsley, *Probability and measure*. John Wiley & sons, 3rd edition, 1995.

[2] Bingham, Goldie, Teugels, *Regular Variation*. Cambridge University Press, 1987.

[3] Conway, *Functions of One Complex Variable 1* Springer-Verlag, 1978.

[4] Driscoll, Trefethen, *Schwarz-Christoffel Mapping* Cambridge university Press, 2002.

[5] Ethier, Kurtz, *Markov Processes: Characterization and Convergence* John Wiley & sons, 2005.

[6] Fan Chung, *Spectral Graph Theory*. American Mathematical Society, 1997.

[7] van Kampen, *Stochastic Processes in Physics and Chemistry*. Elsevier, 3rd edition, 2007.

[8] Kipnis, Landim, *Scaling Limits of Interacting Partical Systems* Springer-Verlag, 1999.

[9] Liggett, *Interacting Particle Systems*. Springer-Verlag, 2005.

[10] Schiessel, *Biophysics for beginners*. in preparation.

[11] Seppäläianen, *Translation Invariant Exclusion Processes* in preparation.

[12] Spitzer, *Principles of Random Walk*. Springer-Verlag, 2nd edition, 1976.

[13] Stroock, *Probability theory and analytic view*. Cambridge University Press, 1993.

[14] Davey, Sardent, Luger et al. *Solvent Mediated Interactions in the Structure of the Nucleosome Core Particle at 1.9 Resolution*. Journal of Molecular Biology, 2002.

[15] Fan Chung, Yau, *Discrete Green's Functions*. Journal of Combinatorial Theory A, 2000.

[16] Flaus, Richmond, *Positioning and Stability of Nucleosomes on MMTV 3HLTR Sequences*. Journal of Molecular Biology, 1998.

[17] Godréche, Luck, Evans et al. *Spontaneous symmetry breaking: exact results for a biased random walk model of an exclusion process*. Journal of physics A, 1995.

[18] ten Heggeler-Bodier,Schild-Poulter et al. *Fate of Linear and supercoiled multinucleosomic templates during transcription* European Molecular Biology Organization Journal, 1995.

[19] Koopmans, Buning, Schmidt, van Noort, *spFRET Using Alternating Excitation and FCS Reveals Progressive DNA Unwrapping in Nucleosomes*. Biophysical journal, 2009.

[20] Landim, Olla, Volchan, *Driven tracer Particle in one Dimensional Symmetric Exclusion* Communications in Mathematical Physics, 1998.

[21] Lowary, Widom, *New DNA Sequence Rules for High Affinity Binding to Histone Octamer and Sequence-directed Nucleosome Positioning* Journal of Molecular Biology, 1998.

[22] Polach, Widom, *Mechanism of protein access to specific DNA sequences in chromatin: a dynamic equilibrium model for gene regulation.* Journal of Molecular Biology, 1995.

[23] Schiessel, *The physics of chromatin.* Journal of physics of Condensed Matter, 2003.