Supporting Material

Richard A. Neher and Ulrich Gerland

June 15, 2005

Waiting Time Distributions

Two Random Walker Model

We consider two random walkers in one dimension confined by two reflecting boundaries $M+1$ sites apart. Since we want to model the process of mutation opening preceding the sliding stage, we seek the distribution of times until encounter of both walkers, given they started at opposite boundaries. Their motion is equivalent to the motion of one walker on a triangular piece of the two dimensional square lattice. The 2D walker on site (m, n) corresponds the state, where the left 1D walker is m steps from the left boundary and the right 1D walker n steps from the right boundary (see Fig. 3, main text). The 2D walker is reflected at the lines $m = 0$ and $n = 0$. The line, where both coordinates add up to $M - 1$ corresponds to the cases, when both walkers in 1D meet and is therefore an absorbing boundary for the 2D walker.

The case, where the rates, at which the walker moves away and towards a boundary (k_{in} and k_{out}) are independent of the site, has been solved by Schwarz and Poland (1) using the methods of image charges.

The quantity we are interested in is the distribution of the time of the first encounter of the two random walkers in 1D, or equivalently the lifetime distribution $P(\tau)$ of the random walker on the triangle. A walker sitting on any site (m, n) with $m = M - 2 - n$ can hop on two absorbing sites with rate k_{in}. The distribution of τ is therefore given by

$$P(\tau) = 2k_{in} \sum_{n=0}^{M-2} \mathcal{P}(n, M-2-n; \tau), \tag{1}$$

where $\mathcal{P}(n,m;\tau)$ is the probability of finding the walker on site (n,m) at time τ, given it started at site $(0,0)$. In the following we derive approximations of the solution by Schwarz and Poland.

Unbiased Hopping

When the walker has no bias, e.g. $k_{in} = k_{out} = k$, $\mathcal{P}(n,m;\tau)$ is given by a sum of $4M^2$ terms. The solution by Schwarz and Poland can be rearranged to

$$\begin{aligned}\mathcal{P}(n,m;\tilde{\tau}) =& \frac{1}{M^2}\sum_{r,s=1}^{2M} e^{-2\tilde{\tau}M^2\left(2-\cos\frac{\pi r}{M}-\cos\frac{\pi s}{M}\right)}(1-(-1)^{r+s}) \\ & \cos\frac{\pi(2n+1)r}{2M}\cos\frac{\pi r}{2M}\cos\frac{\pi(2m+1)s}{2M}\cos\frac{\pi s}{2M},\end{aligned} \tag{2}$$

where the time variable has been rescaled as $\tau = \tilde{\tau}M^2/k$. Only terms, where the argument of the cosines in the exponent are close to 0 or 2π, contribute significantly when $\tilde{\tau} > 1/M^2$. After shifting the summation interval to $r,s = -M\ldots M-1$, significant terms are those the r,s close to 0. We can expand cosines with arguments $\frac{\pi r}{M}$ or $\frac{\pi s}{M}$ and keep only the first non-vanishing contribution.

$$\mathcal{P}(n,m;\tilde{\tau}) \approx \frac{1}{M^2}\sum_{r,s=-\infty}^{\infty} e^{-\tilde{\tau}\pi^2\left(r^2+s^2\right)}(1-(-1)^{r+s})\cos\frac{\pi(2n+1)r}{2M}\cos\frac{\pi(2m+1)s}{2M}. \tag{3}$$

The range of summation can be safely extended to $\pm\infty$, as terms with big r,s are exponentially small. Plugging this approximation into Eq. 1 yields, after some algebra, using similar approximations as above,

$$P(\tilde{\tau}) \approx \frac{2}{M^2}\sum_{r,s=-\infty}^{\infty} e^{-\frac{\tilde{\tau}\pi^2}{M^2}\left(r^2+s^2\right)}\frac{(1-(-1)^{r+s})(r^2+s^2)}{r^2-s^2} \tag{4}$$

Since only those terms with odd $r+s$ contribute, we change the summation variables to $2v = r+s-1$ and $2w = r-s-1$.

$$P(\tilde{\tau}) \approx \frac{4}{M^2}\sum_{v=-\infty}^{\infty} e^{-\frac{\tilde{\tau}\pi^2}{2}(2v-1)^2}\frac{(-1)^v}{2v-1}\sum_{w=-\infty}^{\infty} e^{-\frac{\tilde{\tau}\pi^2}{2}(2w-1)^2}(-1)^w(2w-1) \tag{5}$$

From this expression, we find a parameter-free lifetime distribution

$$\tilde{P}(\tilde{\tau}) = M^2P(\tilde{\tau}) = -\frac{16}{\pi^2}\frac{\partial}{\partial\tilde{\tau}}Q(\tilde{\tau})^2, \tag{6}$$

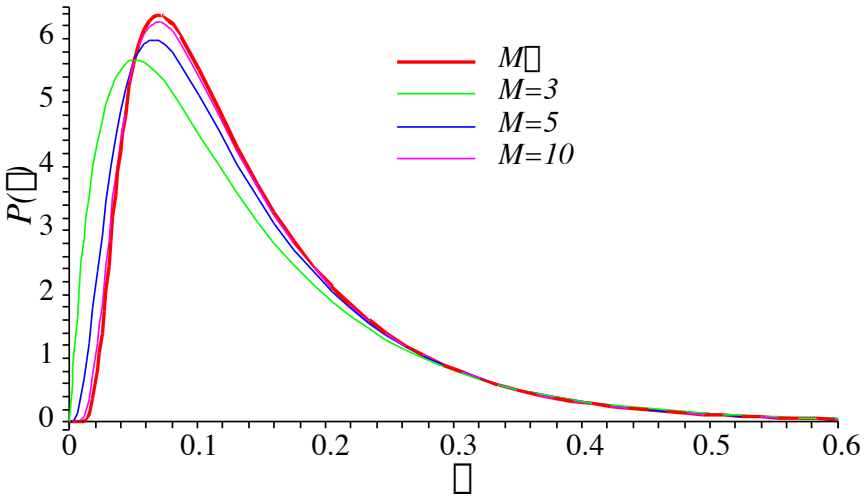

Figure 8: The lifetime distributions for $M = 3, 5, 10$ and the approximation for large M. The time axis is rescaled by M^2.

where $Q(\tilde{\tau})$ is given by

$$Q(\tilde{\tau}) = \sum_{n=1}^{\infty} \frac{(-1)^n e^{-\frac{\pi^2(2n-1)^2\tilde{\tau}}{2}}}{2n-1} \tag{7}$$

The approximations involved are justified for large M. However, even for small systems the agreement is excellent, as illustrated in Fig. 8.

Biased Hopping

When the rates k_{in} and k_{out} are different, there is no compact analytical expression for $\mathcal{P}(n, m; t)$. However, the longterm behaviour of a such a biased random walker is easily understood. If k_{in} is bigger than k_{out}, the walker approaches the absorbing boundary steadily. In the opposite case, the walker will stay close to the origin and only rare excursions will lead to absorption. Quantitatively, the hopping of the random walker on the triangle is well approximated by suitably chosen one-dimensional representation. To that end, we consider the probability to find the walker on the line ν steps away from the origin.

$$P(\nu; \tau) = \sum_{m=0}^{\nu} \mathcal{P}(m, \nu - m; \tau) \tag{8}$$

This amounts to projecting the motion of the random walker onto the symmetry axis of the triangle. The time derivative of this quantity is very similar to a one dimensional hopping process.

$$\begin{aligned}\partial_\tau P(\nu;\tau) = &-2(k_{in}+k^-)P(\nu;\tau) + 2k_{in}P(\nu-1;\tau) + 2k_{out}P(\nu+1;\tau) \\ &+ k_{out}\left[\mathcal{P}(0,\nu;\tau) + \mathcal{P}(\nu,0;\tau) - \mathcal{P}(0,\nu+1;\tau) - \mathcal{P}(\nu+1,0;\tau)\right]\end{aligned} \tag{9}$$

The contributions from the boundary terms in the second line depend on the ratio of k_{in} and k_{out}. When $k_{in} \gg k_{out}$ the walker rapidly approaches the absorbing boundary. The probability $\mathcal{P}(0,\nu;\tau)$ of finding the walker on the reflecting boundary is small, as it is unlikely to make equally many steps with high rate and a low rate. In this case the boundary terms can be neglected entirely, so that the process reduces entirely to a 1D first passage problem. Using standard methods described in ref. (2), one finds, that the mean first passage time

$$\langle\tau\rangle = \frac{M-1}{2k_{in}-2k_{out}} - k_{out}\frac{1-\left(\frac{k_{out}}{k_{in}}\right)^{M-1}}{2(k_{in}-k_{out})^2}, \tag{10}$$

increases linearly with the number of mutations M.

In the opposite limit, when $k_{in} \ll k_{out}$, $\langle\tau\rangle$ increases as $\left(\frac{k_{out}}{k_{in}}\right)^{M-1}$ with M. In this case equilibration along the line $n = \nu - m$ is fast compared to the lifetime of the walker and $\mathcal{P}(m,\nu-m;\tau)$ is almost independent of m. Setting all terms $\mathcal{P}(m,\nu-m;\tau)$ equal results in a 1D hopping process with site dependent rates. The mean first passage time of this process can be calculated in much the same way, yielding $\langle\tau\rangle \sim \left(\frac{k_{out}}{k_{in}}\right)^{M-1}$ with polynomial corrections.

In summary, we find that, depending on whether the walkers have an inward bias, an outward bias or no bias, the mean lifetime scales linearly, exponentially or quadratically with time. Since the force, at which k_{in} and k_{out} are equal, separates regimes, where the waitingtime increases exponentially with M from linear scaling, we call it critical force $\tilde{f}_c$ in the presence of mutations. The force $\tilde{f}_c$ converges towards the critical force f_c in the limit of no mutations.

Measuring Hopping Rates

So far, we have been concerned with the waitingtime distribution given a certain set of rates, at which mutations open or close. These rates depend on the applied force

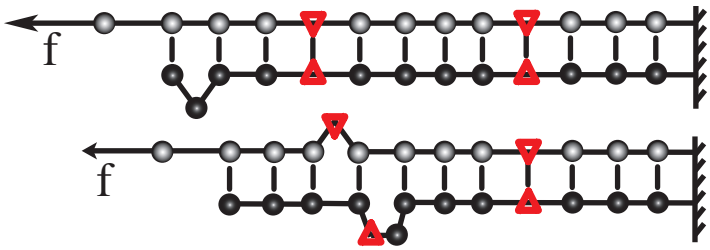

Figure 9: Simplified system to measure the opening and closing rates of mutations. The simulation starts from the ground state with all bases bound. We determine the first-passage time distributions of the opening of the rightmost mutation and fit these to a the first passage time distribution of a random walker.

and on the distance between consecutive mutations and have to be determined in simulations.

As long as there are at least two mutations bound, the dynamics of the opening and closing of mutations at one end is independent of the other end. To measure the rates for a given pair of force and mutation density, we used a simplified system, where a dsDNA with equidistant mutations is fixed on the right hand side and a force is applied to the first base of the upper strand (see Fig. 9). This simplified system is useful, as finite size effects are smaller when one walker crosses M mutations as when two walkers cross $M/2$ mutations each. Furthermore, subtleties of the mutual annihilation process do not enter the measurement. We measure the distribution of the time it takes to open the rightmost mutation for the first time and fit this distribution to the lifetime distribution of a random walker in one dimension between reflecting and absorbing boundary conditions. The rates k_{in} and k_{out} are fit parameters. This is done for a range of forces and mutation densities and the critical force $\tilde{f}_c$ for a certain mutation density can be extract from the crossing of k_{in} and k_{out}. To further pin down $\tilde{f}_c$, we generated data for many force values slightly above and below $\tilde{f}_c$ and fitted a linear relation for each rate to all data sets simultaneously. The crossing of the two resulting lines yield a robust estimate of $\tilde{f}_c$. Using a system of $N = 240$ basepairs, energy parameters $\varepsilon_{\mathrm{b}} = 1.11\mathrm{kT}, \varepsilon_{\ell} = 2.8\mathrm{kT}$ and different number of equidistant mutations, we determined $\tilde{f}_c$ over broad range of mutation densities. The results are shown in Fig. 7(c) in the main text.
To check the reliability of the estimation of $\tilde{f}_c$, we simulated waitingtime distributions by applying the force to both ends of the DNA and fitted the two random walker model to the waitingtime distribution. The force, where k_{in} and k_{out} coincide, reproduces the previously determined force $\tilde{f}_c$. Furthermore, fitting the critical

distribution (one fit parameter) to the waitingtime distribution with yields best fits for $f \approx \tilde{f}_c$. The absolute value of the rates shows slight dependencies on the length of the system (see below) and varies for fits to different setups.

Caveats of the Model

Equilibration of the loopdensity is only possible by propagation of loops from the end beyond a newly broken mutation, or in other words by sliding the unstretched strand some distance Δd inward. The sliding velocity, however, is inversely proportional to the length of the strand. Therefore, equilibration will slow down breaking of mutations for supercritical forces deep inside the double strand and the linear dependence of the waitingtime on the number of mutations will not persist for very large systems.

It is clear from the microscopic mechanism leading to breaking and opening of mutations (see main text and Fig. 10) that the rates k_{in} and k_{out} depend on the force f. The rate k_{out} also depends on the mutation density, since a great distance between mutations corresponds to a large entropy barrier for mutation closing, and hence a smaller closing rate k_{out}. The microscopic opening rate k_{in} is expected to be more or less independent of the mutation density. When looking at the opening and closing dynamics of an individual mutation, this is what we observe. However, the equilibration of loop densities after an opening or closing event takes some time. Therefore, successive microscopic opening and closing events are not entirely uncorrelated, which makes an unambiguous definition of the microscopic rates difficult. These correlations die out very quickly and it is still possible to describe the observed lifetime distribution with an uncorrelated random walker. The effective rates describing this motion both depend on mutation density and the applied force.

References

1. Schwarz, M., and D. Poland. 1975. Random walk with two interacting walkers. *J. Chem. Phys.* 63:557–568.

2. Gardiner, C. W. 1983. Stochastic Methods. Springer-Verlag.

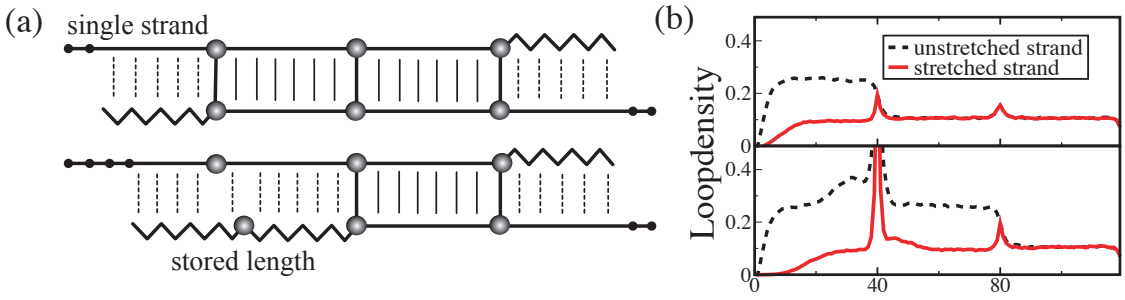

Figure 10: Left: Illustration of how the density of loops on the strands depends on the state of the mutated bases in the sequence. In between bound mutations loops are rare, as the formation of a loop costs initiation energy and shortens the system. The same applies to the stretched strands outside bound mutations. The only part, where a significant number of loops can be found, is the unstretched strand outside the bound mutations. When a mutation is broken, loops move across the mutation on the unstretched strand and locally both strands are shifted against each other. Thereby, the bases that previously formed the mutated basepair become permanently separated and the single strand part on the stretched strand grows. Right: To support the cartoon-like picture of part (a), we measured the time averaged loopdensity, conditioned on a certain mutation state. Mutations are located at base 40 and 80, the parameters are $\varepsilon_{\mathrm{b}} = 1.11\mathrm{kT}$, $\varepsilon_{\ell} = 2.8\mathrm{kT}$ and $f = 10.7\mathrm{pN}$. We consider only opening of mutation from the left, i.e. the rightmost base is kept fixed, as in Fig. 9. When all mutations are bound (upper panel), the loopdensity is high only on the unstretched strand to the left of the mutation at position 40. When this mutation is broken (lower panel), loops can spread from the left end to the mutation at position 80, yielding a fairly constant density interrupted only by the permanent loop at the position of the mutated base. The hump to the left of the broken mutation on the unstretched strand and to the right of the broken mutation on the stretched strand indicate the position of the mutated base on the opposite strand. A loop already present on one strand renders unbound bases on the other strand more likely, as no additional loop initiation has to be paid. The vanishing loopdensity at the end of the stretched strand indicates unbound ssDNA. Observe, that this is the longer, the more mutations are broken.