Multivalent Random Walkers — A Model for Deoxyribozyme Walkers

Mark J. Olah and Darko Stefanovic

Department of Computer Science, University of New Mexico
MSC01 1130, 1 University of New Mexico, Albuquerque, NM 87131

Abstract. We propose a stochastic model for molecular transport at the nanoscale that describes the motion of two-dimensional molecular assemblies called multivalent random walkers (MVRWs). This walker model is an abstract description of the motion of multipedal molecular assemblies, called *molecular spiders*, which use deoxyribozyme legs to move over a surface covered with substrate DNA molecules, cleaving them to produce shorter product DNA molecules as they go. In this model a walker has a rigid inert body and several flexible enzymatic legs. A walker moves over a surface of fixed chemical sites. Each site has one of several molecular species displayed, and walker legs can bind to and unbind from these sites to move over the surface. Additionally, the enzymatic activity of the legs allows them to catalyze irreversible chemical changes to the sites, thereby permanently modifying the state of the surface. We describe a MVRW system as a continuous-time Markov process, where all state transitions in the process correspond to chemical reactions of the legs with the sites. We model the kinetics of the leg reactions by considering the constrained diffusion of the walker body and unattached leg. Through kinetic Monte Carlo simulations, we show that the irreversibility of the enzymatic action of the legs can bias the motion of walkers and cause them to move superdiffusively over significant distances.

1   Introduction

Nature at the nanoscale is different from our familiar macroscopic experience in many ways, the most fundamental of which is the stochastic character of motion and events. At this scale, all objects experience random collisions with molecules that transfer significant energy, effectively randomizing momentum and leading to diffusive motion. Diffusive motion is often not desirable as it becomes a limiting factor in the transfer of material and information in chemical computational systems. However, nanoscale walkers have the potential to move in purposeful, directed ways by expending energy to bias their otherwise diffusive motion, thus providing a mechanism for superdiffusive motion.

Recently a new class of molecular walker based on DNA has been synthetically constructed. These *molecular spiders* [11] consist of a rigid, inert body and several deoxyribozyme (i.e., catalytic single-stranded DNA) legs that act as enzymes and attach to and cleave complementary single-stranded DNA substrates (at a sepecific ribonucleotide impurity). When the substrates are arrayed as nanoscale tracks and paths on a surface the walker can move along such tracks by binding, cleaving, and unbinding

from the track sites [8]. As shown in Fig. 1, when the legs enzymatically modify a bound site by cleaving the substrate they leave behind a shorter product DNA sequence. The product remains complementary to the lower part of the leg. Thus, the legs can walk back over the product sites, albeit at a different rate than that for the substrates, and they can no longer modify the product sites.
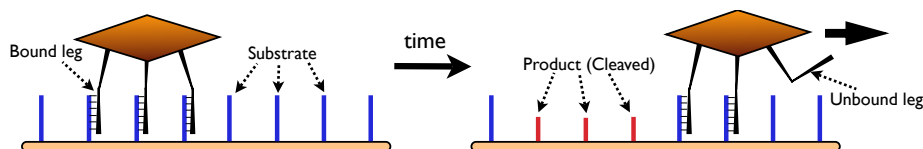

Fig. 1: A molecular spider moves over a surface covered with fixed chemical substrate sites as legs bind and unbind to the sites.

In order to understand how molecular spiders move, we have developed the multivalent random walker (MVRW) model. The model describes spiders in a 2-dimensional environment of chemical sites. The motion of the spiders is modeled as a continuous-time Markov process, where each transition in the Markov process corresponds to a chemical reaction between a leg and a surface-bound site.

In this work we describe the model in detail, and briefly discuss our Monte Carlo simulation methods. We explain that when there is a residency-time bias between modified and unmodified sites, the walker motion is biased in the direction of unmodified sites. Through simulation we show that this bias causes the walker to move superdiffusively, even in opposition to a force.

2 The Multivalent Random Walker Model

At the single-molecule level, chemical kinetics are stochastic in nature. Each individual reaction can be viewed as a transition between two different chemical states of the system as a whole. Accordingly, chemical systems at the single-molecule level can be modeled as continuous-time stochastic processes [9]. A key assumption in such stochastic models is that the system reaches a physical equilibrium (i.e., it is well mixed) in between successive chemical reactions. This makes the system Markovian, and it makes determining the rates of chemical reactions tractable, as the exact position and momentum of particles do not need to be part of the system state, and the state space of the system remains discrete. In the 1970's Gillespie popularized the use of Monte Carlo methods for numerical simulation of stochastic chemical kinetics [3].

Inspired by this approach to chemical kinetics, the MVRW model describes the motion of the molecular spider as a discrete-state, continuous-time Markov process where each transition corresponds to a chemical reaction. In our case, this is a reaction of a leg binding, unbinding, or cleaving sites on the surface, but under the restrictions imposed by the attachment of the legs to a common body. In between reactions, the walker and its legs are assumed to reach a physical equilibrium over all feasible positions. By computing the distribution of the spider's body location after each step, we can accurately
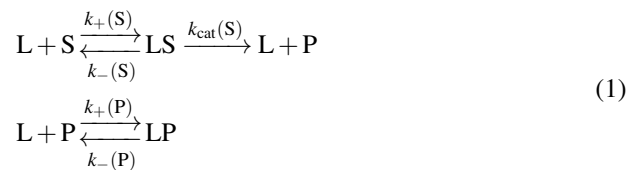
model the chemical reactions and how their rates are affected by the spatial constraints imposed by the spiders' geometry and the pattern of sites on the surface.

2.1 The State Space of the Walker and the Environment

In the MVRW model, the state of the Markov process is defined by the state of the walker and the state of the environment. Walkers are two-dimensional (2D), with a point body to which are attached k flexible legs. Each leg has length ℓ and a reactive site at the end called the *foot*. The walkers move in an environment of fixed chemical sites. The *environment* is defined by a (countable) set $S \subset \mathbb{R}^2$ of sites and a finite set Σ of species. Each site has a single species associated with it, but the species can be changed by the action of the walker legs. Thus, the state of the environment is defined by a mapping $\pi : S \to \Sigma$ that assigns a species to each site. The state of the walker is completely described by the state of its k feet. A foot is either attached to a site in S or is detached. No two feet may be attached to the same site. The state of the walker is represented by the number of detached legs $0 \le d \le k$, and the set $A \subset S$ of attached sites. Thus the state of the MVRW system is defined by the triple (π, d, A).

2.2 State Transitions

There are three types of state transitions corresponding to the three types of chemical reactions that can take place: binding (association), unbinding (dissociation), and catalytic transformation (cleavage). While the model can accommodate more general leg-site chemistries, we focus on the chemistry of the deoxyribozyme-based molecular spiders. In this system, there are two species $\Sigma = \{S, P\}$, a substrate and a product (cleaved) oligonucleotide. A leg (L) can reversibly bind to each species to form a leg-substrate (LS) or leg-product (LP) complex. Additionally, an LS complex can undergo catalysis, transforming the leg into a product before eventually unbinding. These reactions and the relevant rates are defined in Eq. 1. Note that we take $k_{\mathrm{cat}} = k_{\mathrm{cat}}(S)$ to encompass the rate of cleavage as well as the subsequent dissociation, and we assume this process is irreversible.

$$
\begin{aligned}
\mathrm{L} + \mathrm{S} &\underset{k_-(\mathrm{S})}{\overset{k_+(\mathrm{S})}{\rightleftarrows}} \mathrm{LS} \xrightarrow{k_{\mathrm{cat}}(\mathrm{S})} \mathrm{L} + \mathrm{P} \\
\mathrm{L} + \mathrm{P} &\underset{k_-(\mathrm{P})}{\overset{k_+(\mathrm{P})}{\rightleftarrows}} \mathrm{LP}
\end{aligned}
\tag{1}
$$

Of the three types of state transitions, dissociation and cleavage are both unimolecular reactions and, as in the Gillespie model of chemical kinetics, each individual LS or LP pair will dissociate or cleave according to the rates $k_-(S)$, $k_-(P)$, and k_{cat}. However, the association reactions are more complicated as they are bimolecular and their propensity depends on the likelihood of the leg being proximate to the chemical site, so that it may bind. This likelihood, in turn, depends on the position of the body and the unattached legs.
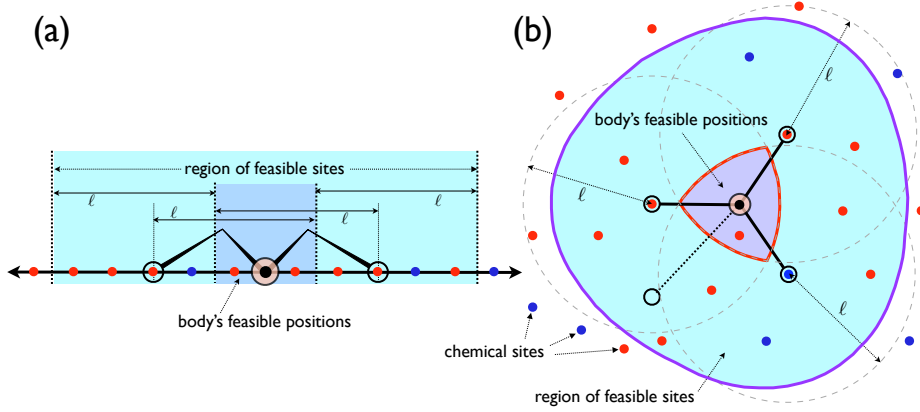
(a)

(b)

Fig. 2: The feasible body positions F for spiders in 1D (a) and 2D (b) are indicated in dark blue. We assume all legs have a maximum length ℓ and the body positions cannot violate these constraints. In light blue we show the region of feasible sites. These are the sites that can be reached by an unattached leg from some feasible body position.

2.3 The Equilibrium Body Distribution

The states of the environment and the walker have been defined to capture only the parts of the system that remain fixed in between reaction events, but that change after a reaction. Notice that the states are discrete, and that the body position is not part of the system state. In this way, the state transitions correspond directly to the chemical reactions and not to the physical motion of the walker body and legs. We assume that non-reactive processes, such as solvent collisions and molecular vibrations, occur on much faster timescales than the chemical reactions so that they come to an equilibrium in between state transitions.

First, let us consider the diffusion of the body. In between reaction events, the body will move in a constrained diffusion. Let random variable $\boldsymbol{B}$ give the 2D coordinates of the spider's body. We assume the legs are flexible with a maximum length ℓ and do not become tangled. Thus, the body will be constrained to be within distance ℓ from each site with an attached foot, so that $\mathbf{P}[\boldsymbol{B} = \boldsymbol{p}] = 0$ if there is any attached site $\boldsymbol{s} \in A$ such that $\|\boldsymbol{p} - \boldsymbol{s}\| > \ell$.

We call all values of $\boldsymbol{p}$ that satisfy $\|\boldsymbol{p} - \boldsymbol{s}\| \leq \ell$ for all $\boldsymbol{s} \in A$ the *feasible body positions*, as it is possible the body is in that position when a reaction finally occurs. The set of all feasible positions is denoted F, and is illustrated as dark blue in Fig. 2.

The exact distribution of $\boldsymbol{B}$ at equilibrium will be a Boltzmann distribution over the feasible positions $\boldsymbol{p} \in F$ according to the energy $E(\boldsymbol{p})$ at each of those positions,

$$\mathbf{P}[\boldsymbol{B} = \boldsymbol{p}] = p_{\boldsymbol{B}}(\boldsymbol{p}) = \frac{e^{-\beta E(\boldsymbol{p})}}{\int_F e^{-\beta E(\boldsymbol{p})} d\boldsymbol{p}}. \tag{2}$$

In Eq. 2, $\beta = 1/k_B T$, where k_B is Boltzmann's constant and T is absolute temperature. We are not concerned with temperature variation, so T will be constant.

2.4 Leg-site Interactions

The kinetics of the bimolecular reaction of leg-site binding is controlled by two factors: the probability of the leg being proximate to a site, and the probability that the leg and site molecules have enough energy to surmount a reaction energy barrier while they are proximate. These probabilities are controlled by the diffusion of the reactants and the activation energy barrier of the reaction [6]. Depending on which of these two processes is rate-limiting, there are two different types of kinetics for the leg-site binding reactions. If the energy barrier is relatively low, the leg is likely to react with one of the first few sites it comes in contact with. Thus, the leg will be more likely to react with sites closer to where it had previously been attached. Because the diffusion to new sites is the limiting factor in the reaction this situation is called *diffusion-limited*. On the other hand, if the energy barrier is higher, the probability of gaining enough kinetic energy from thermal fluctuations will be the limiting factor. This situation is called *reaction-limited*. The leg will diffuse around until there is enough energy to react, and because the leg is constrained to move over a small area, it will quickly reach an equilibrium distribution over sites.

At present we consider the reaction-limited case. Under these conditions, we can assume that the probability of a leg attaching to a site is proportional to the probability that the body is in a position that is less than distance ℓ from the site. We define a function

$$f_L(\boldsymbol{p}) = \begin{cases} 1 & \|\boldsymbol{p}\| \leq \ell \\ 0 & \text{otherwise} \end{cases},$$

that determines if a site at position $\boldsymbol{p}$ is feasible from the origin.

Now, because the body has a distribution over feasible positions, some sites can only be reached from a portion of the feasible body positions. For any site $\boldsymbol{s}$, we can define the probability for leg i being proximate to $\boldsymbol{s}$ when it reacts as

$$\mathbf{P}[i \text{ proximate to } \boldsymbol{s}] = \int_F p_B(\boldsymbol{p}) f_L(\boldsymbol{s} - \boldsymbol{p}) d\boldsymbol{p}. \tag{3}$$

Any site with non-zero probability of being reached is called a *feasible site* and this defines the region of feasible sites shown in Fig. 2 in light blue. Sites outside the feasible region have a rate of 0 for attachment to a leg. If a leg is proximate to a site, there is a a constant rate per unit time at which the particular leg-site binding will occur. This rate is a function of the diffusion rates, leg structure, and chemical free energy barriers, but as these are constants we ignore the details and just assume that the rate is $k_+(\pi(\boldsymbol{s}))$ when the site $\boldsymbol{s}$ has species $\pi(\boldsymbol{s})$. Then the rate of attachment for unattached leg i to feasible site $\boldsymbol{s}$ is $k_+(\pi(\boldsymbol{s}))\mathbf{P}[i \text{ proximate to } \boldsymbol{s}]$. Together with the much simpler rates for the uni-molecular dissociation and cleavage reactions, which are independent of body and leg diffusion, this enables us to model all of the reactions that lead to state transitions in the model.

2.5 Effect of Forces on Walkers

The MVRW model can also model the effect of forces on the body of the walker. This is an advantage of modeling the body's distribution as a Boltzmann distribution deter-

mined by the energy of the spider at each feasible position. Consider the original energy function $E(\boldsymbol{p})$. Under the effect of a force $\boldsymbol{f}$, the new energy of position $\boldsymbol{p}$ becomes,

$$\tilde{E}(\boldsymbol{p}) = E(\boldsymbol{p}) - \boldsymbol{f} \cdot (\boldsymbol{p} - \boldsymbol{p_0}). \tag{4}$$

Because the probability is determined by a Boltzmann distribution, the absolute value of the energy doesn't matter, so any $\boldsymbol{p_0}$ reference point will do for determining the energy of the positions.

The new energy $\tilde{E}$ will give a new equilibrium distribution whose probability mass is shifted in the direction of the applied force. The effect of forces on the body's equilibrium position, and the propensity for each of the feasible sites is shown graphically in Fig. 3.
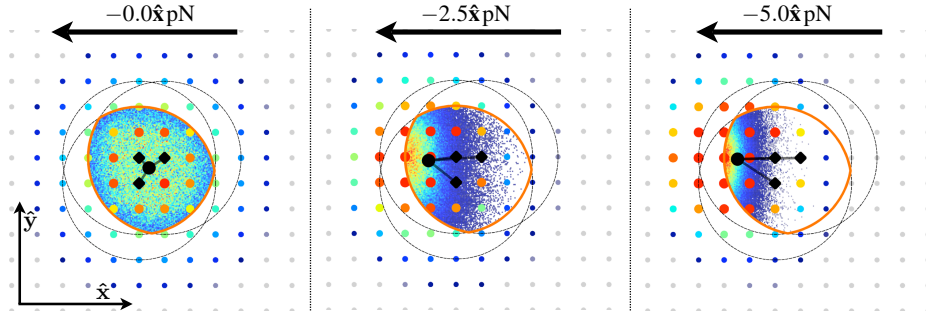

Fig. 3: The equilibrium body distribution for a walker under several different forces: $-0.0\hat{\boldsymbol{x}}$ pN, $-2.5\hat{\boldsymbol{x}}$ pN, and $-5.0\hat{\boldsymbol{x}}$ pN. As the force increases the energy of positions in the $+\hat{\boldsymbol{x}}$ direction become higher, and their probability decreases. The body is drawn at the distribution mean, and the color and size of sites indicates their effective rate for attachment reactions.

3 Simulation

The MVRW model takes the form of a discrete-space, continuous-time Markov process (CTMP). Let Ω be the set of states for the MVRW process, and recall that, as described in Section 2.1, each state $\omega \in \Omega$ can be defined as a triple (π, d, A). Then, given settings for the relevant model parameters and a suitable start state $\omega_0 = (\pi_0, d_0, A_0)$, the MVRW Markov process is described by $X(t)$, where for each $t \in [0, \infty)$, $X(t)$ is a random variable over Ω giving the distribution for the state of the process at that time. A full characterization of the Markov process would involve analytic estimates for the probability distributions $X(t)$. However, this is both infeasible and unnecessary for our purposes.

3.1 Monte Carlo Simulation

A more tractable way to analyze CTMP's is though the Monte Carlo approach. A *Monte Carlo simulation* generates a function $x : [0, t_{\max}] \to \Omega$, called a *realization* of $X(t)$. At each time t, $x(t)$ is a sample of the random variable $X(t)$.

Discrete-state Markov processes must jump instantaneously from one state to the next, hence such processes are often called *jump Markov processes* [4]. A jump Markov process can be described by a transition rate function Q, where $Q(\omega_1 \to \omega_2) \geq 0$ gives the rate of transition (jumping) from state ω_1 to state ω_2. This function, and an initial start state ω_0, completely determine the Markov process $X(t)$. For jump Markov processes, Monte Carlo simulation can be carried out exactly, because a realization $x(t)$ will be a piecewise constant function, consisting of jumps to a sequence of states $\{s_i\}$ at a sequence of jump times $\{t_i\}$, so that $x(t) = s_i$ for $t \in [t_i, t_{i+1})$.

There are two main uses for Monte Carlo simulations of Markov processes. The first is to estimate the equilibrium distribution of Markov processes with a limiting distribution. In this approach the state sequence $\{s_i\}$ becomes an unbiased sampling from a distribution that would otherwise be hard to sample from. The second use is to estimate the dynamic or kinetic properties of a Markov process as it evolves from its initial state. In this case we are interested in how an out-of-equilibrium Markov process behaves as it evolves according to the transition function Q.

To study the MVRW model we use both types of Monte Carlo simulations. At the timescales of chemical reactions we use the kinetic Monte Carlo algorithm to simulate the dynamics of the MVRW Markov processes, obtaining traces of individual spiders moving stochastically according to transition rates. In contrast, at the physical timescales we use the Metropolis-Hastings algorithm to sample from the equilibrium distribution of the body's position as it moves by constrained diffusion in the feasible region F.

3.2 The Kinetic Monte Carlo Algorithm

The kinetic Monte Carlo (KMC) method refers to a rejection-free method of generating exact realizations of a jump Markov process by starting at some fixed initial state and evolving the system state and time according to the transition rates of the model [13]. Let $X(t)$ be a Markov process over state space Ω with transition rate function Q. Given an initial state s_0, the KMC algorithm evolves the system state through time. After the n-th step of the algorithm, the system will be in state s_n at time t_n. The task of the KMC algorithm is to stochastically choose s_{n+1} and t_{n+1} according to the transition rates, Q. Let $Z = \{s' \in \Omega \mid Q(s_n \to s') > 0\}$ be the set of transitions from state s_n with non-zero rate. We assume that $|Z| = k$ is finite and non-zero, and thus we can enumerate it as $Z = \{z_i\}_{i=1}^k$, and define rates $\{r_i\}_{i=1}^k$, with $r_i = Q(s_n \to z_i)$. Let the total rate of all transitions be $R = \sum_{i=1}^k r_i$. This situation is illustrated in Figure 4a.

The probability of the process moving to state z_i at step $n+1$ is given by the ratio r_i/R. We can choose a next state $z^* \in Z$ by selecting a random number $\alpha \sim \text{Uniform}([0, R))$ and choosing $z^* = z_j$, where j is the smallest integer satisfying $\sum_{i=1}^j r_i > \alpha$. This process is depicted in Figure 4b.

Fig. 4: (a) At step n of the KMC algorithm, the system is in state s_n, and we must choose s_{n+1} from amongst the k possible next states $\{z_i\}_{i=1}^{k}$ according to their respective transition rates $\{r_i\}_{i=1}^{k}$. (b) We can select the next state with a single random number $\alpha \sim \text{Uniform}((0,R))$, where $R = \sum r_i$ is the total rate. This example shows the next state chosen to be z_2.

Finally, the algorithm decides how much time should elapse until the transition to z^*. From our current state, all of the possible transitions in Z occur stochastically with constant rate per unit time. Thus, the time τ_i until the transition to z_i will be exponentially distributed, $\tau_i \sim \text{Exp}(r_i)$. We are interested only in the probability distribution for the minimum of these times, $\tau^* = \min\{\tau_1, \ldots, \tau_k\}$. The exponential distribution has the convenient property that τ^* will also be exponentially distributed, as

$$\mathbf{P}[\min\{\tau_1, \ldots, \tau_k\} > t] = \mathbf{P}\left[\bigwedge_{i=1}^{k} \tau_i > t\right] = \prod_{i=1}^{k} \mathbf{P}[\tau_i > t] = \prod_{i=1}^{k} e^{-tr_i} = e^{-t\sum r_i} = e^{-tR}.$$

Thus, we see that $\tau^* \sim \text{Exp}(R)$. Sampling from the exponential distribution is particularly easy, as $\tau^* = -\ln\beta/R$ for $\beta \sim \text{Uniform}((0,1))$.

At this point, the KMC algorithm records the next state $s_{n+1} = z^*$ and the new time $t_{n+1} = t_n + \tau^*$, and then the process repeats until N simulation steps have been made.

3.3 Metropolis-Hastings Distributions

The MVRW model assumes that the body and unattached legs come to an equilibrium distribution in between reaction steps. In Section 2.4 we explain how the transition rates for binding reactions are computed given a probability distribution $p_{\boldsymbol{B}}(\boldsymbol{p})$ over body locations, and a probability distribution $p_L(d)$ for the leg's distance from the body. Overall, the rate at which unbound leg i binds to site $\boldsymbol{s}$ is given by Eq. 3.

With knowledge of $p_{\boldsymbol{B}}$, we can estimate the rates $r_{i \to \boldsymbol{s}}$ using Monte Carlo integration. If $\boldsymbol{P}_1, \ldots, \boldsymbol{P}_n \sim \boldsymbol{B}$ are samples from $\boldsymbol{B}$, they can be used as an unbiased estimator for a function f of the body's position [7],

$$\int_F f(\boldsymbol{p}) p_{\boldsymbol{B}}(\boldsymbol{p}) d\boldsymbol{p} = \left\langle \frac{1}{n} \sum_{i=1}^{n} f(\boldsymbol{P}_i) \right\rangle.$$

The distribution $p_{\boldsymbol{B}}$ is defined in Eq. 2. The denominator in this function,

$$Z = \int_F e^{-\beta E(\boldsymbol{p})} d\boldsymbol{p}, \tag{5}$$

is called the partition function, and is difficult to compute making sampling directly from p_B difficult. The Metropolis-Hastings (MH) algorithm [5, 10] allows p_B to be sampled without knowledge of Z.

The MH algorithm samples from p_B by starting with any Markov process on the distribution domain, $\mathbb{R}^2$, transforming that Markov process into an ergodic discrete-time Markov chain that has p_B as an equilibrium distribution. This Markov chain is defined by transition probabilities $\widetilde{Q}$ where $\widetilde{Q}(\boldsymbol{p}_1 \to \boldsymbol{p}_2) = Q(\boldsymbol{p}_1 \to \boldsymbol{p}_2)\alpha$, and

$$\alpha = \min\left\{1, \frac{p_B(\boldsymbol{p}_2)Q(\boldsymbol{p}_2 \to \boldsymbol{p}_1)}{p_B(\boldsymbol{p}_1)Q(\boldsymbol{p}_1 \to \boldsymbol{p}_2)}\right\}. \qquad (6)$$

The MH algorithm can simulate the Markov process under $\widetilde{Q}$ without ever constructing a rate table explicitly. Also, because the definition of α has p_B in the numerator and denominator, the partition function Z will cancel eliminating the need to compute it. Together these considerations make the MH algorithm an efficient and effective means of sampling from p_B.

The result of the MH algorithm is a sequence of values $\{\boldsymbol{p}_i\}_{i=0}^{N}$. At step i, the simulation has value $\boldsymbol{p}_i$ and it uses this to draw a candidate value $\boldsymbol{p}^* \sim q(\boldsymbol{p}) = Q(\boldsymbol{p}_i \to \boldsymbol{p})$. If Eq. 2 is written as $p_B(\boldsymbol{p}) = f(\boldsymbol{p})/Z$, then we calculate $f(\boldsymbol{p}^*)$ and with Eq. 6 get

$$\alpha = \min\left\{1, \frac{f(\boldsymbol{p}^*)Q(\boldsymbol{p}^* \to \boldsymbol{p}_i)}{f(\boldsymbol{p}_i)Q(\boldsymbol{p}_i \to \boldsymbol{p}^*)}\right\}.$$

With probability α we choose to accept the point and set $\boldsymbol{p}_{i+1} = \boldsymbol{p}^*$, otherwise we reject this candidate value and try again. We repeat this until we have generated N values. This procedure is illustrated in Figure 6. The sequence of values returned will include an initial period before the chain reaches equilibrium. The initial points are highly dependent on the starting value, and thus are not an unbiased sample. Typically these points are dropped. Depending on the nature of the distribution and the application a threshold can be set so that subsequent points are independent of the starting value with high probability [2].
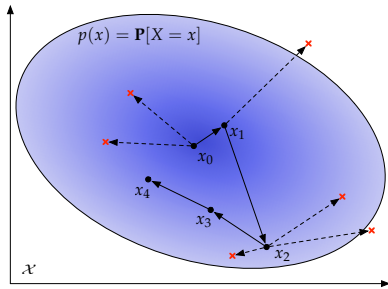

Fig. 6: The Metropolis-Hastings algorithm samples from probability distribution $p(x)$, by simulating a Markov chain with an equilibrium distribution equal to $p(x)$. The algorithm generates a sequence of points $\{x_i\}_{i=0}^{N}$ by using the current point to draw a new candidate point, and choosing to accept or reject that point with probability α. In this figure a red cross represents a rejected point, and a labeled black point represents an accepted point.

Parameter	Description
$k = 4$	Number of legs
$\ell = 2.5\,\text{nm}$	Length of each leg
$k_+(\text{S}) = k_+(\text{P}) = 1.0 \times 10^3\,\text{s}^{-1}$	On rate for leg binding
$k_-(\text{P}) = 1.0\,\text{s}^{-1}$	Off rate for products
$k_-(\text{S}) = 0.0\,\text{s}^{-1}$	Off rate for substrates
$k_{\text{cat}} \in \{1.0\,\text{s}^{-1}, 0.01\,\text{s}^{-1}\}$	Catalysis rate
$f \in \{0.00\,\text{pN}, 0.05\,\text{pN}, 0.10\,\text{pN}, 0.50\,\text{pN}\}$	Force in the $-\hat{\boldsymbol{x}}$ direction
$T = 300.0\,\text{K}$	Absolute temperature

Table 1: Parameters used in simulations.

4   Preliminary Results

We used our KMC algorithm to simulate 100 realizations of the MVRW Markov process for several different parameter values. The walkers were simulated until time $t_{\text{max}} = 3.0 \times 10^6\,\text{s}$. In our experiment the walker started at the origin on a semi-infinite track. The track is 3 sites wide and sites are on a $1.0\,\text{nm} \times 1.0\,\text{nm}$ grid, as shown in Fig. 7. The walkers experienced a force in the $-\hat{x}$ direction that essentially opposed the direction of highest substrate gradient. The simulations parameters are summarized in Table 1. The only parameters varied were the force and the catalysis rate. The experiments with $k_{\text{cat}} = 1.0\,\text{s}^{-1}$ have no effective difference between substrate and product, so walkers can only be expected to move diffusively (at least in the absence of any force). However, the walkers with $k_{\text{cat}} = 0.01\,\text{s}^{-1}$ will experience significantly slower detachment from substrates than from products, producing a residency-time bias.
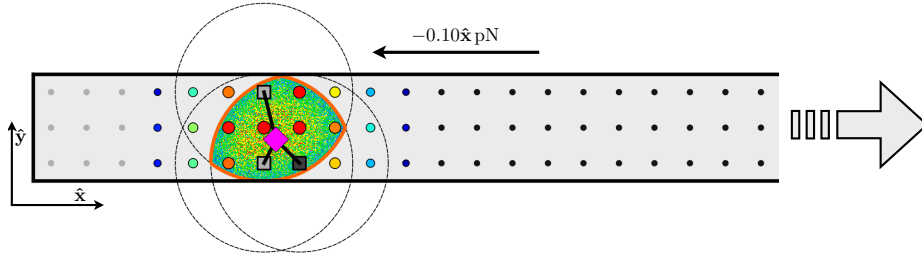

Fig. 7: An example configuration of a MVRW simulation. The walker has three attached legs. Light gray sites represent products, dark gray are substrates. Each attached leg forms a circular constraint on the body's position, defining the feasible region in orange. The body's distribution is shown within this feasible region, and the color and size of surrounding sites represents their effective rate for attachment reactions.

To quantify the diffusive properties of the walker we estimated moments of several random variables relevant to the walker motion. One of the defining characteristics of diffusive motion is that the mean-squared displacement, $\langle \|\boldsymbol{p}\|^2 \rangle$, of a walker increases

as a power law with exponent $\alpha = 1$. Eq. 7 defines various forms of anomalous diffusion when $0 \leq \alpha \leq 2$, where $d = 2$ is the dimension and D is the diffusion constant.

$$\langle \|\boldsymbol{p}(t)\|^2 \rangle = (2dD)t^{\alpha}, \begin{cases} \alpha = 0 & \text{stationary} \\ 0 < \alpha < 1 & \text{subdiffusive} \\ \alpha = 1 & \text{diffusive} \\ 1 < \alpha < 2 & \text{superdiffusive} \\ \alpha = 2 & \text{ballistic or linear} \end{cases} . \quad (7)$$

In Fig. 8(a) we show the mean-squared displacement of the walkers on a log-log plot where power laws are straight lines. We show reference lines for the power laws corresponding to diffusive and ballistic motion. Clearly, the walkers with lower k_{cat} experience significant periods of superdiffusive motion, while the walkers with $k_{\text{cat}} = 1.0\,\text{s}^{-1}$ move mainly diffusively.
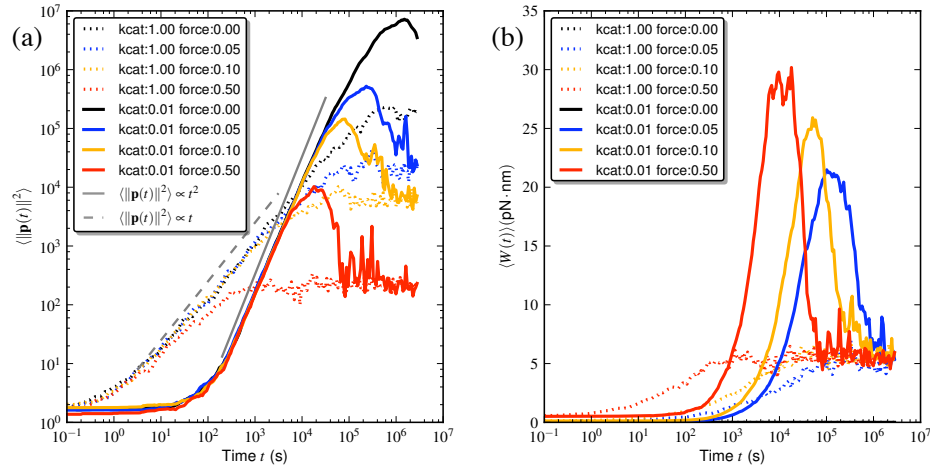

Fig. 8: (a) The mean squared displacement of walkers shows significant superdiffusion in walkers with small k_{cat}. (b) The work done by walkers against an opposing force. The walkers moving under zero force always do zero work. Note that $\langle W(t) \rangle \to 4.14\,\text{pN} \cdot \text{nm} = k_B T$, as $t \to \infty$.

If we consider the work done against the opposing force, there is a significant amount of work done on average for all of the walkers, but the maximum mean work is done by walkers with lower k_{cat}. In Sec. 5 we show that the lower values of k_{cat} act to bias the walker in the direction of uncleaved substrate, and this direction is essentially in opposition to the force exerted on the walkers, allowing them to use this bias to do work against the force. Eventually, however, all of the walkers move backwards into regions of product sites, and end up effectively diffusing like the walkers with $k_{\text{cat}} = 1.0\,\text{s}^{-1}$ (Fig 8(b)).

5   Mechanism of Superdiffusive Motion

The results of Sec. 4 show that spiders can move superdiffusively in the direction of new sites even in opposition to a force. Over significant spans of time, the walkers will have effectively done work against the force as their motion is biased by the chemical energy in the sites they cleave.

Molecular spiders operate by cleaving a substrate oligonucleotide, leaving behind a shorter oligonucleotide product – an irreversible reaction. A molecular spider starting on a substrate-covered surface is a system far from equilibrium, and consequently has the potential to do useful work as it relaxes towards equilibrium. Under the parameters of Table 1, there is as a *residency-time bias* between leg-substrate and leg-product bindings for the walkers with $k_{cat} < 1.0\,s^{-1}$, because the leg-substrate bindings are much longer lived than the leg-product bindings. When combined with a non-uniform local distribution of substrates, the slower unbinding from substrates causes the walker to be effectively biased in the direction higher substrate density.

When a leg binds to a site, it forms a constraint on the position of the body and the actions of the other legs until a dissociation reaction occurs. According to Eq. 1, the rate of detachment for a leg-substrate complex is $k_{cat}(S) + k_-(S)$, versus $k_-(P)$ for a leg-product complex. We define $r = (k_{cat}(S) + k_-(S))/k_-(P)$. If $r = 1$, there is effectively no difference between substrate and product; although the substrate sites are transformed to products, they do not affect the behavior of the walker. This is equivalent to a walker moving over an all-product surface – an equilibrium process. Thus, we can expect the walker to undergo normal diffusion when $r = 1$. Indeed, this is what we see in Fig. 8a, where the spiders with $k_{cat} = 1.0\,s^{-1}$ move diffusively with $\langle \|\boldsymbol{p}(t)\|^2 \rangle \propto t$.

However, when $0 < r < 1$, a leg-substrate bond lasts longer than a leg-product bond, and substrates effectively act like anchors. A leg attached to a substrate restricts the movement of the walker body and other legs until the substrate is cleaved, and the other legs are constrained to attach to feasible sites close to the attached leg. If a free leg attaches to a product, it will quickly detach and be free to attach again to another site. If there are any other substrates in the local environment, one of the other free legs will eventually find and attach to one. Thus, the legs are in some sense attracted to substrates, but not because they specifically seek out the substrates or prefer them to products. Instead, the bias is more subtle, caused by a combination of the residency-time bias and the collective constraints on the legs imposed by the connection to a common body. The legs eventually find the substrates simply because if they attach to a product, they will quickly end up detaching and randomly choosing a new attachment site again and again until they find a substrate. Note that this effect is only present when the walker has more than one leg and has $r < 1$, so both of these properties are critical for spiders to move superdiffusively.

This bias, however, also depends on the local availability of substrates. Once a leg attaches to a substrate, the site will eventually be irreversibly transformed into a product. Thus, while the legs (passively) seek out the substrates, they eventually will deplete the local substrate supply. For a small environment with a limited number of sites, substrates will all quickly be turned into products, at which point the system will be at equilibrium and the walker will move diffusively. However, with larger environments this march towards equilibrium takes a significant amount of time, and during this non-

equilibrium period there is potential for superdiffusive motion and for doing physical work against a force.
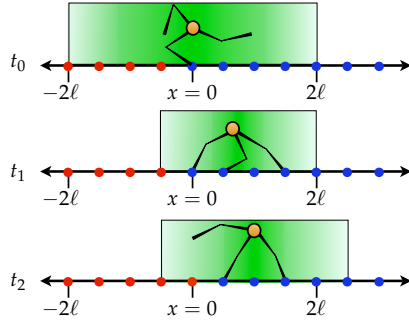

Fig. 10: A residency-time bias combined with a non-uniform local distribution of substrates can lead to a directional bias. There is a boundary at $x = 0$ between substrates (blue) and products (red). At time t_0 a single leg is attached to a substrate, and the other legs can attach to any feasible sites (shaded area). Because the leg-product pairs are short-lived, the legs are more likely to end up attached to substrates at time t_1. When the first leg detaches at time t_2, the equilibrium position and substrate boundary will move right.

Now, consider what happens when the local environment has a non-uniform distribution of substrates. Suppose, as in Fig. 10, the walker has a single leg attached to a substrate at site s with location $x = 0$. The local environment of feasible sites will then consist of all sites within two leg lengths (2ℓ) from $x = 0$. Suppose that all sites with position $x \geq 0$ are substrates and all sites with position $x < 0$ are products. Now consider what happens when the process is started. The initially attached leg will likely remain attached to the substrate for some time if $r < 1$. During this time the other $k - 1$ legs will be restricted to the feasible sites. Short lived product attachments mean that legs will end up preferentially attached to substrates by the time the first leg cleaves and detaches. At this point if most of the legs are on substrates, and all of the substrates are to the right, the spider's equilibrium body position will move right. At the same time, because the site at $x = 0$ is now a product, the boundary between the substrates and products also moves right. Thus, the walker is biased towards moving right, and simultaneously shifts the biasing-inducing substrate/product boundary rightward as well. As long as the walker stays attached to substrates by the boundary, it will tend to move along with the boundary, causing the walker to move ballistically in the direction of new substrates. However, there is still some probability that the walker detaches from all substrates and moves backwards over previously visited sites. In this case, the walker must move diffusively.
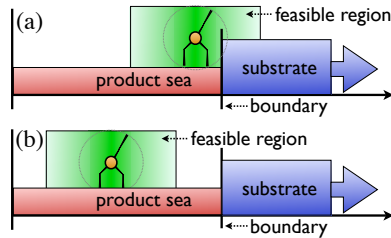

Fig. 12: (a) The walker in a boundary state B where it is attached to substrates on the boundary between visited and unvisited sites. The residency-time bias and non-uniform local distribution of substrates gives the spider an outward bias. (b) The walker in the diffusive state D where it moves over previously visited sites.

In previous work [12] we also observed significant periods of superdiffusive motion in the simpler one-dimensional molecular spider models of Antal and Krapivsky [1]. For these models, we explained this superdiffusive motion by showing that the Markov process can be viewed as consisting of two metastates: a boundary (B) state where the walker is on the boundary between cleaved and uncleaved sites, and a diffusive (D) state where the walker is moving over previously visited sites. The walker moves ballistically in the B state and diffusively in the D state, and the overall motion depends on how much time the walker spends in each of the metastates. Similarly, the initial superdiffusive motion in the MVRW model for $r < 1$ can be understood as the walker moving between a B and a D state as shown in Fig 12. The walker initially spends most of its time in the B state, moving ballistically away from the origin in the direction of unvisited sites, and in opposition to the force. However, the walker has a constant probability of falling off the boundary and into the D state where it moves diffusively over previously visited sites. In the D state, the force acts to bias the motion of the walker backwards, and as the size of the region of cleaved products (the *product sea*) grows, the spider takes increasingly long to return to the B state, and eventually becomes on average stationary at some equilibrium position with mean work $\langle W(t) \rangle = k_B T$, as observed in Fig 8a.

6 Discussion

Given the many potential nanoscale applications for molecular spiders, it is interesting to see that the MVRW model predicts that walkers move superdiffusively over significant times and distances, even in the presence of a force. This motion is not a product of differing k_+ rates, but is rather of a more subtle nature, emerging from the interaction of a residency-time bias, local substrate anisotropy, and constraints imposed by multiple legs attached to a single body. Walkers with $r < 1$ stay attached to substrate sites longer than to product sites. The presence of a non-uniform local distribution of substrates combined with the constrained diffusion imposed by the attached legs causes the walker to move in the direction of highest substrate density, leading to superdiffusive behavior when $r < 1$. This effect relies on the walker having multiple legs.

The ability of the MVRW model to stochastically incorporate the effect of force on the walker kinetics is due to the separation of time scales between the very fast physical motion and vibration of molecules and the much slower chemical reactions. Because the body and unattached legs come to an equilibrium before reattaching, we can model their motion together with the effect of a force using a Boltzmann distribution. This assumption means that the body and unattached leg positions need not be part of the MVRW model state, so our model remains discrete and can be simulated exactly.

Because of our choice to explicitly separate the timescales of the physical and chemical events, the MVRW model uses Monte Carlo simulation separately for both equilibrium and kinetic analysis of Markov processes. The MVRW model is a non-ergodic Markov process describing a system significantly out of equilibrium, and we use kinetic Monte Carlo techniques to observe the simulated stochastic evolution of a walker moving over the sites. This puts us in the position of a virtual experimenter, able to run simulated traces of the spider's motion and measure exactly any desired properties

of their motion. In contrast to this kinetic simulation, we use the Metropolis-Hastings algorithm to study the equilibrium distribution of the walker's body moving under the constrained diffusion as enforced by the attached legs.

The superdiffusive motion of walkers in the MVRW model can be understood through the decomposition of the process into a *B* metastate where the walker is on the boundary between substrates and products and is moving ballistically, and a *D* metastate where the walker is moving diffusively over product sites. From a practical standpoint, the duration of the superdiffusive effect and the magnitude of the work done against a force can be increased by designing walkers that are less likely to move from the *B* to *D* states. Future work will focus on the how the geometry of the walkers and their kinetic properties can be optimized to increase the amount of time they spend in the *B* state moving ballistically, hence maximizing their utility for faster than diffusion molecular transport and communication.

References

1. Antal, T., Krapivsky, P.L.: Molecular spiders with memory. Physical Review E 76(2), 021121 (2007)
2. Geyer, C.J.: Practical Markov chain Monte Carlo. Statistical Science 7(4), 473–483 (Nov 1992)
3. Gillespie, D.T.: A general method for numerically simulating the stochastic time evolution of coupled chemical reactions. Journal of Computational Physics 22, 403–434 (1976)
4. Gillespie, D.T.: Markov processes. Academic Press Inc., Boston, MA (1992)
5. Hastings, W.K.: Monte Carlo sampling methods using Markov chains and their applications. Biometrika 57(1), 97–109 (April 1970)
6. Henriksen, N.E., Hansen, F.Y.: Theories of Molecular Reaction Dynamics. Oxford University Press, New York, NY (2008)
7. Kalos, M.H., Whitlock, P.A.: Monte Carlo Methods. John Wiley & Sons, New York, NY (1986)
8. Lund, K., Manzo, A.J., Dabby, N., Michelotti, N., Johnson-Buck, A., Nangreave, J., Taylor, S., Pei, R., Stojanovic, M.N., Walter, N.G., Winfree, E., Yan, H.: Molecular robots guided by prescriptive landscapes. Nature 465, 206–210 (May 2010)
9. McQuarrie, D.A.: Stochastic approach to chemical kinetics. Journal of Applied Probability 4(3), 413–478 (1967)
10. Metropolis, N., Rosenbluth, A.W., Rosenbluth, M.N., Teller, A.H., Teller, E.: Equation of state calculations by fast computing machines. The Journal of Chemical Physics 21(6), 1087–1092 (1953)
11. Pei, R., Taylor, S.K., Stefanovic, D., Rudchenko, S., Mitchell, T.E., Stojanovic, M.N.: Behavior of polycatalytic assemblies in a substrate-displaying matrix. Journal of the American Chemical Society 39(128), 12693–12699 (2006)
12. Semenov, O., Olah, M.J., Stefanovic, D.: Mechanism of diffusive transport in molecular spider models. Physical Review E 83(2), 021117 (Feb 2011)
13. Voter, A.: Introduction to the kinetic Monte Carlo method. In: Sickafus, K., Kotomin, E., Uberuaga, B. (eds.) Radiation Effects in Solids. Springer (2007)