

The Effects of Multivalency and Kinetics in Nanoscale Search by Molecular Spiders

Extended Abstract

Oleg Semenov¹, Darko Stefanovic^{*1,2}, and Milan N. Stojanovic^{3,4}

¹ Department of Computer Science, University of New Mexico

² Center for Biomedical Engineering, University of New Mexico

³ Division of Experimental Therapeutics, Department of Medicine, Columbia University

⁴ Department of Biomedical Engineering, Columbia University

Abstract. Molecular spiders are nanoscale walkers made with catalytic DNA legs attached to a rigid body. They move over a surface of DNA substrates, cleaving them and leaving behind product DNA strands, which they are able to revisit. The legs cleave and detach from substrates more slowly than they detach from products. This difference in residence time and the presence of multiple legs make a spider move differently from an ordinary random walker. The number of legs, and their lengths, can be varied, and this defines how a spider moves on the surface, i.e., its gait. In this work we define an abstract model of molecular spiders. Using Kinetic Monte Carlo simulation, we study how efficiently spiders with various gaits are able to find specific targets on a two-dimensional lattice. Multi-legged spiders with certain gaits can find the targets faster than regular random walkers. The search performance of spiders depends on both their gait and the kinetic rate r describing the relative substrate/product “stickiness”. Spiders with gaits that allow more freedom for leg movements find their targets faster than spiders with more restrictive gaits. For every gait, there is an optimal value of r that minimizes the time to find all target sites.

1 Introduction

We are developing synthetic nanoscale walkers, called molecular spiders [1, 2], which are able to move across a surface, propelled by the multivalent chemical interactions of their multiple legs with the surface (Section 2). Molecular spiders may find use in biomedical applications, such as searching for clinically relevant targets on the surface of a cell. Here we present simulation-based results on the efficiency of concurrent search for multiple targets by multiple molecular spiders. Spiders and their targets are simulated on a finite two-dimensional grid of chemical sites (Section 3), which models the DNA origami surface [3, 4] on which we will eventually carry out laboratory experiments.

Molecular spiders’ legs are made of catalytic single-stranded DNA. These molecules are not known in nature, though they easily might have evolved in nature. Being both catalytic and information-carrying, they are candidate components for a synthetic biochemical artificial life. When we study them in simulation, we are not only taking inspiration from natural life, as in the field of artificial life, but we are also prototyping

how the natural world can be refashioned and engineered, which one might call *real artificial life*.

The salient kinetic parameter governing the *walking behavior* of a molecular spider is the ratio r between the residence time of a spider’s leg on a previously visited site (spent product) and the residence time on a new site (fresh substrate) [5–8]; r is at most 1 because substrates are generally “stickier” than products. As r is decreased, the motion exhibits an increasingly strong superdiffusive transient [7].

The conclusion of our *searching behavior* simulations (Section 4) is that the efficiency of search strongly depends on r and the configuration of the spiders. Among the studied spider configurations we found that those spiders whose legs have more freedom of movement find their targets faster than spiders with more restrictive rules of leg movement. Also, for the studied finite surface, we found that the efficiency of search does not vary monotonically with r ; indeed, in each of the several configurations of multi-legged spiders we simulated, there was a distinct optimal r value. This behavior is a consequence of the interplay between the effects of the spiders’ having multiple legs and the kinetic bias (Section 5).

2 Molecular Spiders

Cells in nature accomplish many of their complex tasks using self-assembled filament tracks and (linear) molecular motors that walk directionally along the filaments [9–14]. These *natural* protein motors solve the problem of efficient molecular cargo transport across the cell. Recent advances in single-molecule chemistry have led to *synthetic* molecular motors, including molecular assemblies that walk over surfaces, typically following fabricated or self-assembled tracks [2, 15–22]. Among these are *molecular spiders*, DNA-based autonomous synthetic molecular motors [1, 2].

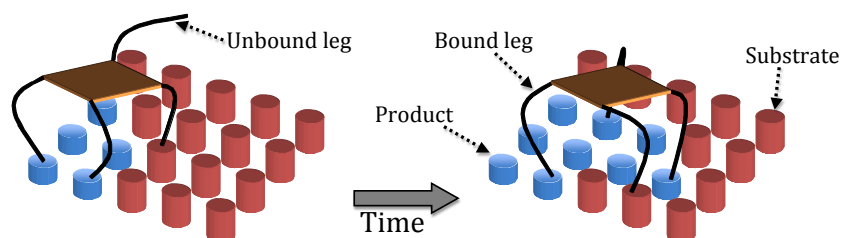


Fig. 1: A molecular spider moves over a surface covered with fixed chemical substrate sites as its legs bind and unbind to the sites.

A molecular spider consists of an inert body to which are attached flexible enzymatic legs (Figure 1, left). We have reported spiders with up to six legs, using a streptavidin or streptavidin dimer scaffold for the body [1]. Each leg is a deoxyribozyme—an enzymatic sequence of single-stranded DNA that can bind to and cleave a complementary strand of a DNA substrate. The hip joint between the body and a leg is a flexible

biotin linkage. When a molecular spider is placed on a surface coated with the single-stranded DNA substrate, its legs bind to the substrate. A bound leg can either detach from the substrate without modifying it, or it can catalyze the cleavage of the substrate, creating two product strands. The cleavage occurs at a designed ribonucleobase position within the otherwise DNA substrate. (The 8-17 enzyme we use for the legs was originally selected to cleave RNA [23].) Upon cleavage the two product strands eventually dissociate from the enzyme leg. The “lower” product remains bound to the surface. Because the lower product is complementary to the lower part of the spider’s leg, there is a residual binding of the leg to the product; this binding is typically much weaker than the leg-substrate binding and thus much shorter-lived. The “upper” product remains in solution, so there can be some product rebinding. In laboratory experiments this effect is minimized with a flow setup; in our model we neglect it.

Surface-plasmon-resonance experiments [1] show that a spider moves in a highly processive manner, cleaving thousands of substrates before eventually detaching from the surface. We conclude that it moves in the direction of fresh substrates, leaving behind a trail of products (Figure 1). But experimental observation of the motion has been limited by the small scale of a spider; although single-molecule fluorescence studies have been used to some effect [2], it has not yet been possible to establish the spider’s movement and gait with certainty, nor to track the substrate cleavage site-by-site in real time. It has therefore been necessary to approach the problem using detailed modeling studies, through mathematical analysis and computer-based simulations [5–8, 24–26]. Molecular spiders, viewed as random walkers, have been modeled at different levels of abstraction, and with various parameter settings. Their asymptotic behavior is diffusive, just as with ordinary random walkers. However, and more important for the connection with laboratory experiments, superdiffusive behavior is observed in the transient, and it lasts for significant amounts of time, over which a spider covers significant distances. We showed [7] that in the presence of a residence-time bias between substrates and products this behavior can be explained by the spider’s switching between two states—being on the boundary of the fresh substrates, and being in the sea of already cleaved products. A spider on the boundary extracts chemical energy from the landscape, moves preferentially towards fresh substrates, and thus carries the boundary along. A spider that has stepped back into the products wanders aimlessly, i.e., diffuses.

Depending on the level of detail captured by the model, many kinetic parameters can be used to characterize the motion [8]. The basic parameter, however, already present in the most abstract model [5], is the chemical kinetic ratio r mentioned above. Because all our models are at heart continuous-time Markov processes, r is defined as the ratio between the transition rate out of a leg-on-substrate state and that out of a leg-on-product state. In this paper, we only use this abstract model, and so r is the only kinetic parameter.⁵

Once we have characterized how well the spiders are able to *walk*, the question becomes what tasks spiders can usefully *do*, such as carrying cargo molecules, following predesigned tracks, and searching for targets in unstructured landscapes. The latter is

⁵ Elsewhere we study additional kinetic details using more elaborate and more computationally expensive models [8]. These models do permit useful characterization of mechanical motor properties, but they do not alter the basic walking behavior.

the topic of the present study. We describe models of future experiments wherein several identical molecular spiders will search for multiple specific target sites located on a finite surface made using the DNA origami technique. We are interested in the time it takes for *all* the spiders to find their targets.

3 Abstract Model of Molecular Spider Motion and Search

Motion. In our model, a spider walks on a two-dimensional rectangular lattice representing the chemical sites (initially substrates). There is complete exclusion: when one of the spider’s k legs detaches from a site, it moves to an unoccupied site in a certain neighborhood of its current location; sites occupied by another leg (whether of the same spider or of a different one) are excluded. The concrete definition of a neighborhood is different for different instances of the model, representing different spider gaits. Figure 2 illustrates the neighborhoods n we use in our model instances: a black circle is the current leg site and the surrounding white circles are sites accessible within one step. Because any leg can rebind immediately to the site which it just left, the current site is also considered to be a part of the neighborhood.

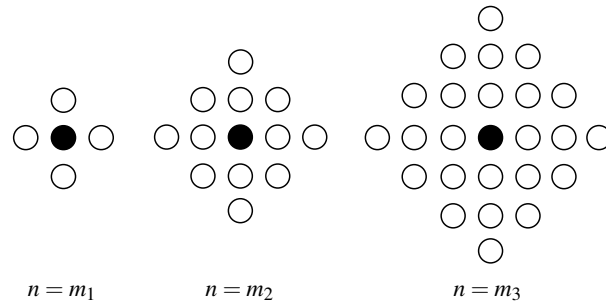


Fig. 2: Neighborhoods studied. By exploring a variety of neighborhood sizes and shapes in this abstract model, we hope to guide the choice of parameters for physical spiders in future laboratory experiments, primarily the leg length (which can be adjusted using spacers).

Another constraint on the leg movement is the maximum distance S between any two legs of a spider. The distance S can be measured differently in different models. In the following, we use two distance metrics, the Manhattan (L_1) distance S_m and the Chebyshev (L_∞ ; maximum) distance S_z .

Together, the parameters n , S , and k define the gait of a spider. Within the constraints of the gait, a spider’s motion is governed by the chemical kinetics of its legs, which we model as a continuous-time Markov process. Each leg independently interacts with the chemical site it is on; at the high level of abstraction of this model, the interaction is completely described by a single transition rate. A leg detaches at rate 1 from a product, and at rate r from a substrate, where $r \leq 1$. When a leg leaves a substrate, that

substrate is transformed into a product. At the high level of abstraction of the model, the reattachment of the leg is instantaneous. There is no directional bias in the model: if in the current state there are several moves that do not violate any restrictions, the leg that is moving chooses any one of them with equal probability.

Search. In our search model, the lattice is of a finite fixed size, 22 by 32; the numbers may seem arbitrary, but they reasonably describe the DNA-origami tiles used in past molecular spider experiments [2]. We use three searching spiders, initially in one corner of the lattice. The search targets are the three special *trap* sites, in the three opposite corners. We assume that a leg that attaches to a trap remains forever bound to it.⁶ Furthermore, when a spider’s leg is thus trapped, *all* its legs cease moving. All remaining sites initially are ordinary cleavable substrates. Because there are as many targets as spiders, eventually each spider reaches a target. When all target sites have been found all motion stops (in this crude abstraction).

The chemical kinetics parameter r and the spiders’ gait parameters n , S , and k are the variable parameters of the model, and they will influence how fast the spiders move and search for targets. In the following section we begin our exploration of this parameter space.

Combining the states of all the spiders and the surface gives a continuous-time Markov process for our model. We use the Kinetic Monte Carlo method [27] to simulate multiple trajectories of this Markov process. The simulation stops when all spiders are trapped, i.e., the search is complete, and then the simulated time is recorded. This time is an observation of a first-passage-time random variable. The results below will show the mean first passage time estimated from our traces.

4 Simulation Results

We carried out simulations using the following seven spider gaits: (a) $k = 2$, $S_z = 1$, $n = m_1$; (b) $k = 2$, $S_m = 2$, $n = m_3$; (c) $k = 2$, $S_m = 3$, $n = m_2$; (d) $k = 3$, $S_m = 2$, $n = m_3$; (e) $k = 3$, $S_m = 3$, $n = m_2$; (f) $k = 4$, $S_m = 2$, $n = m_3$; (g) $k = 4$, $S_m = 3$, $n = m_2$; and a simple random walker, which can be viewed as a spider with parameters (h) $k = 1$, $S_m = 1$, $n = m_1$. These gaits were chosen to correspond to different physically realistic molecular spiders, but it must be admitted that the space of possible plausible gaits is much larger, and we must defer its exploration to a future study. Table 1 describes the chosen gaits, along with the initial spider positions and the target site positions, shown graphically. Initial positions for the gaits with equal number of legs k are equivalent, and thus Table 1 groups the gaits by k . Black, green, and violet circles represent the initial leg positions of the first, second, and third spider respectively; gray circles represent ordinary substrates; red circles represent the three target traps. The targets are non-specific, that is, any spider can be trapped by any target.

⁶ In the laboratory, an uncleavable, pure-DNA substrate has been used [2]) for the purpose. In envisaged applications, the targets presented on the cell surface will not necessarily be DNA strands. To bind to non-DNA targets, in addition to the legs the spider may carry an “arm”, an aptamer molecule that specifically binds to the target.

Table 1: The initial state of the system and the spider gait for each simulated configuration. Configurations (a)–(g) are multi-legged spiders and configuration (h) is the control one-legged spider.

<p>(a) $k = 2, S_z = 1, n = m_1$ (b) $k = 2, S_m = 2, n = m_3$ (c) $k = 2, S_m = 3, n = m_2$</p>	<p>(d) $k = 3, S_m = 2, n = m_3$ (e) $k = 3, S_m = 3, n = m_2$</p>
<p>(f) $k = 4, S_m = 2, n = m_3$ (g) $k = 4, S_m = 3, n = m_2$</p>	<p>(h) $k = 1, S_m = 1, n = m_1$</p>

Simulation results are shown comprehensively in Figure 3, and also individually for each gait in Figure 4 to reveal additional detail.

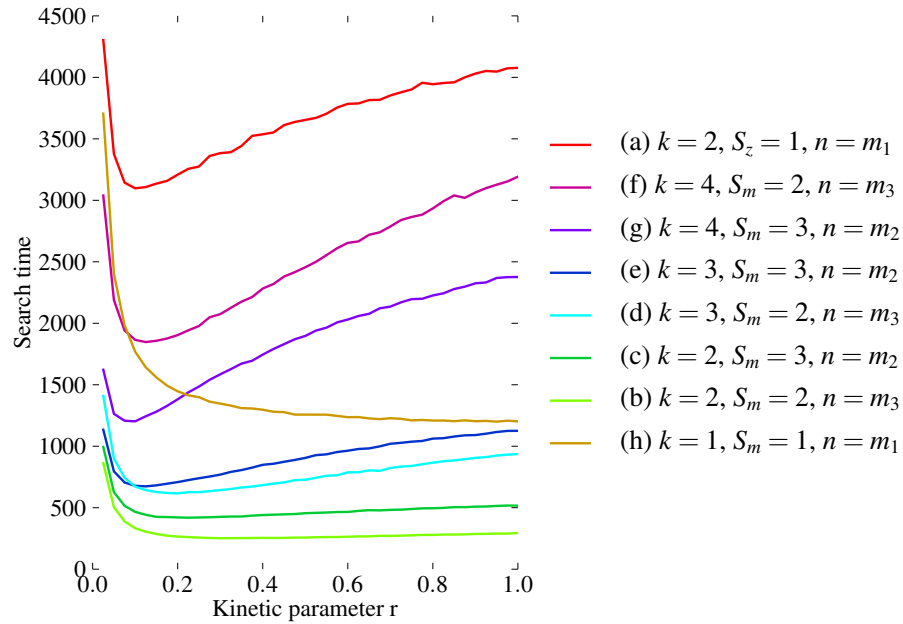


Fig. 3: Search time as a function of the kinetic parameter r for the eight configurations simulated. (a)–(g) are multi-legged spiders; (h) is a one-legged spider.

4.1 The effect of the gait

In Figure 3 curves corresponding to the studied gaits have different shapes and are vertically separated. Thus, the gaits of the spiders greatly influence their performance, and spiders with particular gaits can be faster than regular random walkers. For the simulated surface, spiders with gaits that allow more freedom for the legs to move (i.e., when a leg is moving to a new site it has more sites to choose from) achieve better performance. Two- and three-legged spiders with gaits (b), (c), (d), and (e) are the fastest among those simulated; they either have the larger neighborhood $n = m_3$ or the longest possible distance between legs $S_m = 3$, which makes these gaits the least restrictive. Spiders with gait (a) $k = 2, S_z = 1, n = m_1$ are the slowest, despite having the same number of legs $k = 2$ as the best performing spiders, with gait (c): gait (a) is much more restrictive, and with its parameters $S_z = 1$ and $n = m_1$ it gives a small candidate set of new positions for the legs. Gaits (f) $k = 4, S_m = 2, n = m_3$ and (g) $k = 4, S_m = 3,$

$n = m_2$ are also slower than gaits (b) through (e): they have the same parameters S_m and n , but the addition of an extra leg reduces the choice of sites for a moving leg (because the new position must be within a certain distance from each of the legs that remain attached), which leads to slower performance.

4.2 The effect of the kinetics

We now examine the influence of the kinetics (i.e., the difference between the substrates and the products displayed on the surface, which is amenable to adjustment in the laboratory) on the search performance of the spiders. We used 40 different values of the chemical kinetics parameter r in the range from 0.025 (heavy substrate-product bias) to 1.0 (no bias) in increments of 0.025. Results are shown in the eight panels of Figure 4. The kinetic parameter r significantly affects the performance of the spiders. One-legged spiders always have better performance when r is bigger—this observation is similar to the results of the study of one-legged spiders in one dimension [6]. Thus, the presence of memory on the surface in the form of substrates and products does not improve the performance of a monovalent random walker. For multi-legged spiders each gait has an optimal r value that minimizes the search time; this behavior is similar to two-legged spiders in one dimension [7].

In previous work [7] we found that when $r < 1$, spider ensembles go through three different regimes of motion—initially, spiders move slowly; then they start moving faster and achieve better performance than regular diffusion; and, finally, in the time limit they slow down and move as regular diffusion. For lower r values the initial slow period is longer than for higher r values, but then later the superdiffusive period is longer and faster for smaller r values. For travel over shorter distances the length of the initial period is more important than for longer distances and thus smaller r values can result in higher first passage times. For travel over longer distances the superdiffusive period is important and smaller r values give better results. Thus for every particular distance there is an optimum value r that minimizes the mean first passage time. Figure 5 shows the mean first passage time for a single two-legged spider in one dimension. For example, for the distance of 50 sites the spider with $r = 0.1$ is faster than the other (sampled) r values; but for the distance of 20 sites the $r = 0.5$ spider is faster. Also, although spiders with $r = 0.01$ and $r = 0.005$ are the slowest for distances of 50 sites and less, they eventually overtake all other simulated spiders at greater distances. Thus, similarly to these results for two-legged spiders in one dimension [7], we speculate that for bigger search lattices the optimum r values will decrease.

5 Discussion

The simulations were performed for a relatively small lattice and the gaits that are currently the fastest may not turn out to be so for bigger lattices. Also, it appears that r values have more influence on search time for some gaits than for others: gaits (f) $k = 4$, $S_m = 2$, $n = m_3$ and (g) $k = 4$, $S_m = 3$, $n = m_2$ get the most improvement in performance when r decreases from 1 to its optimal value, and gaits (b) $k = 2$, $S_m = 2$,

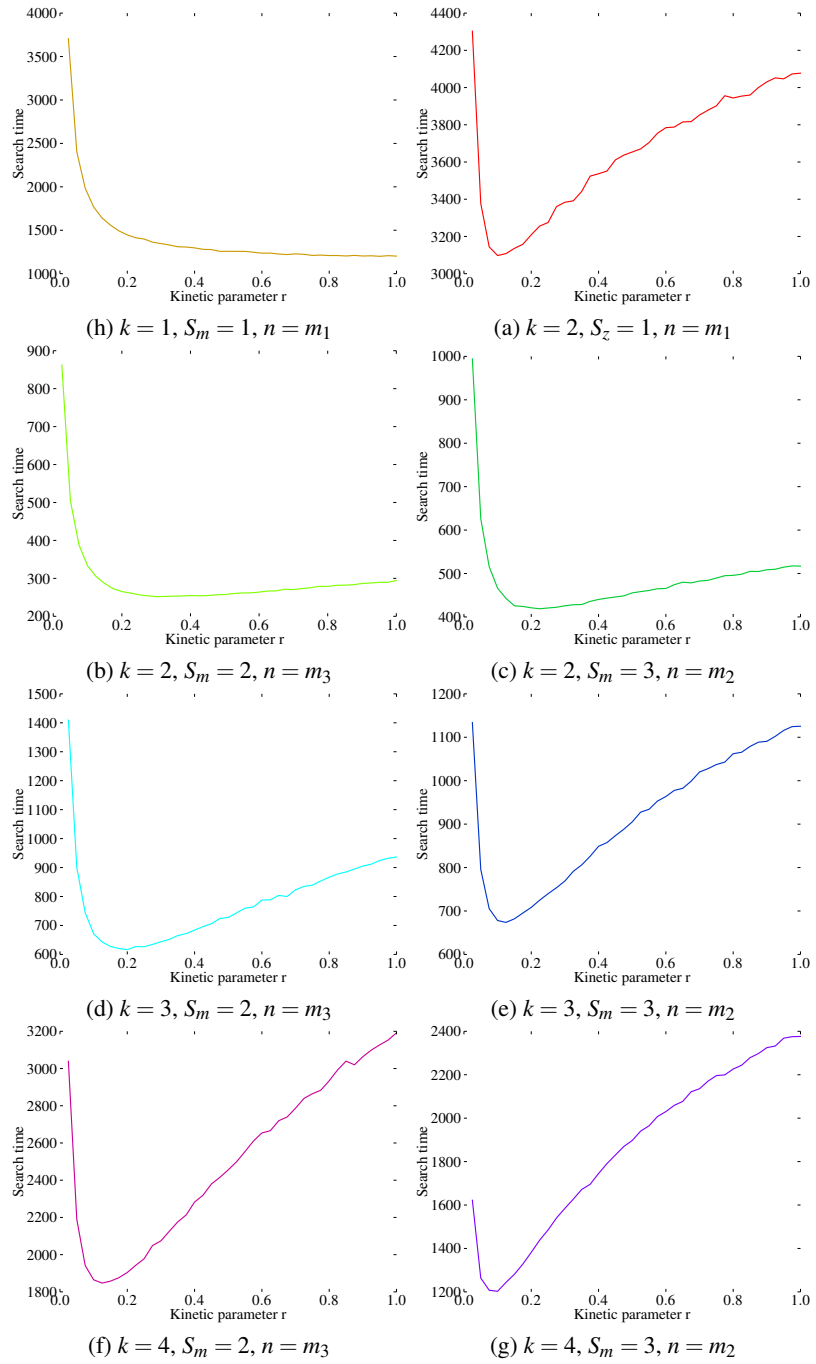


Fig. 4: Enlarged plots from Figure 3, in which the existence of an optimum r value for the multi-legged spiders can be discerned better.

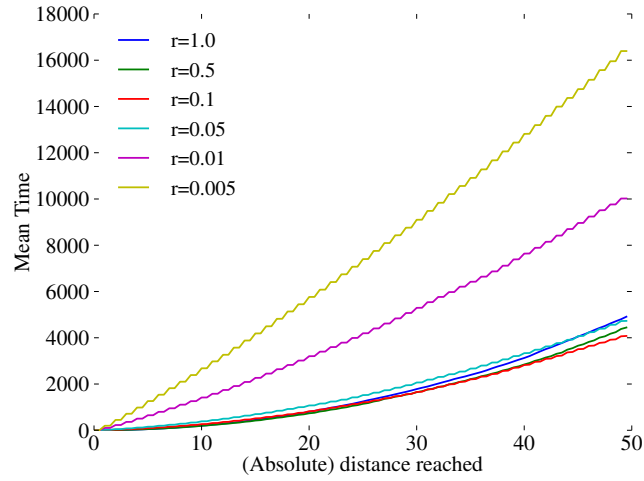


Fig. 5: Mean first passage time of a single two-legged spider moving over an infinite one-dimensional track initially covered with substrates.

$n = m_3$ and (c) $k = 2$, $S_m = 3$, $n = m_2$ get the least improvement. To understand this dependence, in future work we shall study gaits more systematically, i.e., have a scenario to vary parameters k , S , and n .

Search performance can also be affected by the exclusion between spiders and by the initial position of the spiders. For example, the longer search time of larger, four-legged spiders than of two-legged ones could be due in part to the relatively small simulated surface. Larger spiders may obstruct each other's progress more often, especially early on. To account for these effects, in future work we shall vary the placement of the targets and the spider starting positions, as well as the number of targets and spiders.

A number of assumptions had to be made to reduce the complex interactions of physical molecular spiders on DNA origami tiles, which have not yet been fully experimentally characterized and understood, to a tractable mathematical model amenable to efficient computer simulation. The highly abstract model presented here may sacrifice too much physical realism; at the opposite end of the abstraction spectrum are molecular dynamics approaches, but those are infeasible at the space and time scales of interest. Laboratory experiments can reveal the ground truth, but are too expensive for a full exploration of the parameter space. Instead, we plan to use mesoscale models [8] to refine these initial results.

Our system can be viewed as a hierarchical multi-agent system: the system consists of multiple spiders, and each spider consists of multiple legs. The legs and the spiders interact through exclusion, and also stigmergically as they modify the surface; and the legs of one spider interact through kinematic constraints. But we are not free to design this multi-agent system as we please to achieve some system-level behavior; instead, we are severely restricted in the design of the agents: they are just molecules of a particular kind, and molecules are dumb. Thus, recalling the principle of complexity theory

that simple local rules, iterated, may give rise to complex global behaviors, we are asking whether this also happens in a very primitive setting. We do not expect to be able to mimic the complexity of behaviors of even a single ant, which after all is billions times more structurally complex. However, we hope that our results will aid in the development of nanoscale molecular walkers. In future, we plan to continue the study of searching behaviors and to initiate a study of additional biochemically plausible modes of spider–spider interaction.

Acknowledgments. We thank Paul L. Krapivsky and Mark J. Olah for many discussions of spider behaviors. We thank the conference reviewers for their incisive and detailed comments. This material is based upon work supported by the National Science Foundation under grant 0829896/0829744.

References

1. 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* **128**(39) (2006) 12693–12699
2. 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** (May 2010) 206–210
3. Rothmund, P.W.K.: Folding DNA to create nanoscale shapes and patterns. *Nature* **440** (2006) 297–302
4. Pinheiro, A.V., Han, D., Shih, W.M., Yan, H.: Challenges and opportunities for structural DNA nanotechnology. *Nature Nanotechnology* **6** (2011) 763–772
5. Antal, T., Krapivsky, P.L., Mallick, K.: Molecular spiders in one dimension. *Journal of Statistical Mechanics: Theory and Experiment* **2007**(08) (2007) P08027
6. Antal, T., Krapivsky, P.L.: Molecular spiders with memory. *Physical Review E* **76**(2) (2007) 021121
7. Semenov, O., Olah, M.J., Stefanovic, D.: Mechanism of diffusive transport in molecular spider models. *Physical Review E* **83**(2) (February 2011) 021117
8. Olah, M.J., Stefanovic, D.: Multivalent random walkers — a model for deoxyribozyme walkers. In: *DNA 17: Proceedings of The Seventeenth International Meeting on DNA Computing and Molecular Programming*. Volume 6397 of *Lecture Notes in Computer Science*., Springer (2011) 160–174
9. Kolomeisky, A.B., Fisher, M.E.: Molecular motors: A theorist’s perspective. *Annual Review of Physical Chemistry* **58** (2007) 675–95
10. Bier, M.: The energetics, chemistry, and mechanics of a processive motor protein. *BioSystems* **93** (2008) 23–28
11. Astumian, R.D.: Thermodynamics and kinetics of molecular motors. *Biophysical Journal* **98** (June 2010) 2401–2409
12. Jamison, D.K., Driver, J.W., Rogers, A.R., Constantinou, P.E., Diehl, M.R.: Two kinesins transport cargo primarily via the action of one motor: Implications for intracellular transport. *Biophysical Journal* **99** (November 2010) 2967–2977
13. Lipowsky, R., Beeg, J., Dimova, R., Klumpp, S., Müller, M.K.I.: Cooperative behavior of molecular motors: Cargo transport and traffic phenomena. *Physica E* **42** (2010) 649–661
14. Driver, J.W., Jamison, D.K., Uppulury, K., Rogers, A.R., Kolomeisky, A.B., Diehl, M.R.: Productive cooperation among processive motors depends inversely on their mechanochemical efficiency. *Biophysical Journal* **101** (July 2011) 386–395

15. Yurke, B., Turberfield, A.J., Mills, Jr., A.P., Simmel, F.C., Neumann, J.L.: A DNA-fuelled molecular machine made of DNA. *Nature* **406** (August 2000) 605–608
16. Shirai, Y., Osgood, A.J., Zhao, Y., Kelly, K.F., Tour, J.M.: Directional control in thermally driven single-molecule nanocars. *Nano Letters* **5**(11) (2005) 2330–2334
17. Tian, Y., He, Y., Chen, Y., Yin, P., Mao, C.: A DNzyme that walks processively and autonomously along a one-dimensional track. *Angew. Chem. Int. Ed.* **44** (2005) 4355–4358
18. Venkataraman, S., Dirks, R.M., Rothmund, P.W.K., Winfree, E., Pierce, N.A.: An autonomous polymerization motor powered by DNA hybridization. *Nature Nanotechnology* **2** (2007) 490–494
19. Green, S.J., Bath, J., Turberfield, A.J.: Coordinated chemomechanical cycles: A mechanism for autonomous molecular motion. *Physical Review Letters* **101** (2008) 238101–+
20. Omabegho, T., Sha, R., Seeman, N.C.: A bipedal DNA brownian motor with coordinated legs. *Science* **324** (April 2009) 67–71
21. Bath, J., Green, S.J., Allen, K.E., Turberfield, A.J.: Mechanism for a directional, processive, and reversible DNA motor. *Small* **5**(13) (2009) 1513–1516
22. Gu, H., Chao, J., Xiao, S.J., Seeman, N.C.: A proximity-based programmable DNA nanoscale assembly line. *Nature* **465** (May 2010) 202–206
23. Santoro, S.W., Joyce, G.F.: A general purpose RNA-cleaving DNA enzyme. *Proceedings of the National Academy of Sciences of the USA (PNAS)* **94** (April 1997) 4262–4266
24. Samii, L., Linke, H., Zuckermann, M.J., Forde, N.R.: Biased motion and molecular motor properties of bipedal spiders. *Physical Review E* **81**(2) (Feb 2010) 021106–+
25. Semenov, O., Olah, M.J., Stefanovic, D.: Multiple molecular spiders with a single localized source—the one-dimensional case. In: *DNA 17: Proceedings of The Seventeenth International Meeting on DNA Computing and Molecular Programming*. Volume 6397 of *Lecture Notes in Computer Science.*, Springer (2011) 204–216
26. Samii, L., Blab, G.A., Bromley, E.H.C., Linke, H., Curmi, P.M.G., Zuckermann, M.J., Forde, N.R.: Time-dependent motor properties of multipedal molecular spiders. *Phys. Rev. E* **84** (Sep 2011) 031111
27. Bortz, A.B., Kalos, M.H., Lebowitz, J.L.: A new algorithm for Monte Carlo simulation of Ising spin systems. *Journal of Computational Physics* **17**(1) (1975) 10–18