

**Totally asymmetric exclusion process with extended objects: A model for protein synthesis**Leah B. Shaw,<sup>1,2,\*</sup> R. K. P. Zia,<sup>3</sup> and Kelvin H. Lee<sup>2</sup><sup>1</sup>*Department of Physics, Cornell University, Ithaca, New York 14853-2501, USA*<sup>2</sup>*School of Chemical and Biomolecular Engineering, Cornell University, Ithaca, New York 14853-5201, USA*<sup>3</sup>*Center for Stochastic Processes in Science and Engineering, Physics Department, Virginia Polytechnic Institute and State University, Blacksburg, Virginia 24061-0435, USA*

(Received 4 February 2003; published 18 August 2003)

The process of protein synthesis in biological systems resembles a one dimensional driven lattice gas in which the particles have spatial extent, covering more than one lattice site. We expand the well studied totally asymmetric exclusion process, in which particles typically cover a single lattice site, to include cases with extended objects. Exact solutions can be determined for a uniform closed system. We analyze the uniform open system through two approaches. First, a continuum limit produces a modified diffusion equation for particle density profiles. Second, an extremal principle based on domain wall theory accurately predicts the phase diagram and currents in each phase. Finally, we briefly consider approximate approaches to a nonuniform open system with quenched disorder in the particle hopping rates and compare these approaches with Monte Carlo simulations.

DOI: 10.1103/PhysRevE.68.021910

PACS number(s): 87.10.+e, 64.60.Cn, 05.70.Ln, 02.50.Ga

**INTRODUCTION**

The process of protein synthesis, important in biological systems, has been the focus of intensive study over the last few decades. We concentrate on protein synthesis in prokaryotes, particularly *Escherichia coli*, because it is relatively simple and well studied. The mechanism consists of ribosomes “reading” the codons of messenger RNA (mRNA) as the ribosomes move along an mRNA chain, and the recruitment and assembly of amino acids (appropriate to the codons being read) to form a protein. (See, e.g., Ref. [1], for more details.) Known as “translation,” this process is often described as three steps: initiation, where ribosomes attach themselves, one at a time, at the “start” end of the mRNA; elongation, where the ribosomes move down the chain in a series of steps; and termination, where they detach at the “stop” codon. Since ribosomes cannot overlap, their dynamics is subject to the “excluded volume constraint.” Apart from being impeded by another ribosome (steric hindrance), a ribosome cannot move until the arrival of an appropriate transfer RNA, carrying the appropriate amino acid (a combination known as aminoacyl-tRNA, or aa-tRNA). Thus, the relative abundances of the approximately 60 types [2] of aa-tRNA have significant effects on the elongation rate. Assuming reactant availabilities in a cell are in their steady state, with a time-independent concentration of ribosomes and aa-tRNA, there would be an approximately steady (average) current of ribosomes moving along the mRNA, resulting in a specific production rate of this particular protein. Our goal is the prediction of the protein production rates for various mRNA’s, as a function of the concentration of ribosomes and aa-tRNA’s.

The process of translation is well suited to modeling using a driven lattice gas in one dimension. In most relevant studies of one dimensional driven lattice gases, particles are in-

jected at some rate on one end of a chain of discrete lattice sites, then hop down the chain one site at a time with another rate, and finally exit the chain at the other end with a third rate. These three rates correspond to the rates of initiation, elongation, and termination. The excluded volume constraint is implemented by ensuring that each site can accommodate at most one particle. Because the dynamics is stochastic, even this “simple” model, though solvable exactly [3,4], is already not trivial. However, we believe two essential aspects of translation are missing from this model. First, if we model a codon by a lattice site, the ribosome would cover, typically, a dozen sites [5,6]. Second, there is nonuniformity in the hopping (elongation) rates along the chain, because a ribosome has to “wait” for the appropriate aa-tRNA before continuing, and the relative abundance of the different aa-tRNA’s is far from unity. Remarkably, the first issue was explored as early as 1968 [7,8], though only at the deterministic, “mean field” level.

In this paper, we present studies that address both of these issues, extending the work on the “simple model” known as TASEP (namely, single-site coverage, totally asymmetric simple exclusion processes with open boundaries) [9]. Our methods involve both Monte Carlo simulations and modern analysis techniques, including domain wall theory [10]. We have confirmed all key results from earlier studies [7,8], and several new insights have emerged. The paper is organized as follows. The details of the model are delineated first, along with brief summaries of known results. Section II is devoted to a *closed* (i.e., periodic) system with particles of arbitrary size. Though not a direct model for translation, TASEP on a “ring” provides simple solutions as well as useful insights in the form of *exact* relations for relevant parameters, such as current-density relations. In Sec. III, we turn to the central topic: TASEP with extended objects and open boundaries. Section IV is devoted to nonuniform hopping (elongation) rates. We close with a brief summary and speculate on the relevance of our model as a mechanism for

\*Electronic address: lbs22@cornell.edu

the nonlinear relationship between mRNA and protein levels observed in biological experiments [11,12].

## I. THE DRIVEN LATTICE GAS AS A MODEL FOR PROTEIN SYNTHESIS

### A. Model specifications

We model an mRNA with  $N$  codons as a chain of sites, each of which is labeled by  $i$ . The first and last sites,  $i = 1, N$ , are associated with the start and stop codons, respectively. At any time, attached to the mRNA are  $M$  ribosomes (also referred to as “particles”), which we label by  $\alpha$  ( $\alpha = 1, \dots, M$ ). Being a large complex of molecules, each ribosome will cover  $\ell$  sites (codons), with  $\ell = 12$  typically [5,6]. By contrast, nearly all studies of the asymmetric simple exclusion processes (ASEP) are devoted to  $\ell = 1$ . Any site may be covered by a single ribosome or none. In case of the latter, we will refer to the site as “empty” or “occupied by a hole.” For convenience, we define  $\tilde{M}$  as the number of holes on the chain, so that

$$\tilde{M} + \ell M = N. \quad (1)$$

For open systems, a ribosome at the end can be attached without covering all  $\ell$  codons, so that this equality is only approximately true. To locate the ribosome, we arbitrarily choose the *lowest* site covered. For example, if the first  $\ell$  sites are empty, a ribosome can bind in an initiation step, and then it is said to be “on site  $i = 1$ .” Therefore, a complete specification of the configuration (state, or microstate) of the mRNA is the set of locations:  $\{i_\alpha\}$ . One disadvantage of this labeling is that, with each initiation event, the  $\alpha$  of every ribosome will change (increase by unity). Alternatively, we can use *site occupation* numbers

$$n_i = \begin{cases} 1 & \text{if site } i \text{ is covered by any part of a ribosome,} \\ 0 & \text{if site } i \text{ is empty.} \end{cases}$$

With these conventions, we define several density parameters: (1)  $\rho_r \equiv M/N$  is the ribosome (or particle) density; (2)  $\rho \equiv M\ell/N = \sum_i n_i/N$  is the coverage density; (3)  $\rho_h = 1 - \rho$  is the hole density; and (4)  $\rho_s \equiv \rho_r + \rho_h$  is defined for convenience.

All of these quantities are time dependent, because initiation and termination occur *independently*. For mathematical reasons, we will first consider a *closed* system (with periodic boundary conditions, i.e., the ends of the chain tied to form a ring), for which these densities are fixed and Eq. (1) holds strictly. As will be clear later, it is also convenient to label configurations by specifying the number of holes between successive ribosomes:  $\{h_\alpha\}$ , where  $h_\alpha$  is the number of holes *in front of* the  $\alpha$ th ribosome. In the terminology of traffic models,  $h_\alpha$  is also known as the “headway” of this particle. Though not absolutely necessary, we could define  $h_0$  as the number of holes behind the first ribosome.

Next, we specify the dynamics of our model. An attached ribosome located at site  $i$  will move to the next site ( $i + 1$ ) with a rate  $k_i$ , *provided* site  $i + \ell$  is empty. For Monte Carlo simulations, it is convenient to update configurations in discrete time units. Then, it is better to use probabilities  $p_i$  ( $0$

$\leq p_i \leq 1$ ), so that a ribosome on site  $i$  will be moved or not with probability  $p_i$  or  $1 - p_i$ , respectively. We purposefully associate these hopping probabilities with a site because a site is associated with a particular codon. Thus, the jump rate from that site may depend on the relative abundance of the appropriate aa-tRNA. Apart from these probabilities, another aspect of our stochastics is random sequential updating: i.e., during each Monte Carlo step (MCS),  $M + 1$  particles are chosen at random, in sequence, to attempt moves. They are selected from a pool that includes the  $M$  particles on the lattice plus another unbound particle that can initiate if there are  $\ell$  holes at the beginning of the chain. Let us illustrate with a few examples. First,  $p_0$  is associated with the start codon and, *if* the first  $\ell$  sites are empty, a particle will be placed on the  $i = 1$  site with this probability. Next, a random particle (say at site  $i$ ) is chosen and, provided it has a hole in front ( $n_{i+\ell} = 0$ ), will be moved with probability  $p_i$ . Naturally, it will not be necessary to check for headway for the “last” ribosome ( $\alpha = M$ ). Finally, the stop codon will be associated with  $p_N$ . For simulations of the closed system, there will be no “beginning” or “end,” so that there are no special steps for initiation or termination.

In our computational studies, 100 identical systems of  $N$  sites are simulated in parallel to obtain good statistics. Simulations of closed systems begin with particles evenly distributed around the ring and run for 3600 MCS to ensure that steady state is reached. Open systems begin empty and are run for 12,000 MCS (for  $N < 500$ ) or  $100N$  MCS (for  $N \geq 500$ ) to reach steady state. After steady state is attained, data including the current and density distribution can be collected. Density data are typically collected every 100 MCS. We often use continuous time Monte Carlo [13] because it runs far more quickly than and provides the same results as standard Monte Carlo.

### B. Brief survey of known results

Extensive investigations of the simple totally asymmetric exclusion process (TASEP, defined as point particles hopping with unit rate along a line) with open boundaries can be found in the literature. Simulations have been performed [14], and exact analytic results for the steady state exist [3,4]. Depending on the initiation (or injection) and termination (or depletion) rates, the system will settle into one of three phases. Introduced above as  $p_0$  and  $p_N$ , respectively, the initiation and termination probabilities are mostly referred to as simply  $\alpha$  and  $\beta$  in the literature. From their dominant characteristics, the three phases are known as low density, high density, and maximal current. A phase diagram in this  $\alpha$ - $\beta$  plane has been determined, showing second order transitions between the maximal current phase and the others, as well as a first order transition between the high- and low-density regions. Subtle correlations further divide the last two into subregions. When disorder is introduced, i.e., not all the  $p_i$ 's are equal, then methods for exact analytic approaches fail (except in the extremely dilute limit, where only the motion of a single particle is of concern [15]). Indeed, even a single slow rate in a *closed* system poses serious difficulties [16–18]. However, Kolomeisky [19] obtained ap-

proximate steady state solutions and phase diagrams for an *open* system with a single nonuniform rate in the bulk by splitting the system into two smaller systems connected by the nonuniform rate. Tripathy and Barma [20] considered a closed system, but with a finite fraction of identical slow sites. Based on a combination of Monte Carlo simulations and numerical solutions of mean field equations, they found current-density relations. Though not a model for elongation, a related problem is “particlewise disorder,” in which the unequal hopping rates are associated with the particles rather than the sites [21]. Further references on TASEP with disorder may be found in a recent review [22]. Also indirectly related to our one dimensional models are driven lattice gases with quenched disorder in *higher* dimensions [23,24]. Finally, we mention that there are many studies on the ASEP in which a particle has a finite probability of stepping backwards [25]. Back steps are not generally believed to occur in elongation, and we will not consider such processes. All of these studies are restricted to  $\ell = 1$ .

Systems with extended objects ( $\ell > 1$ ) have been rarely investigated, despite their introduction over three decades ago as a model for biopolymerization [7]. Using a mean field approach, MacDonald *et al.* set up mean field equations for the average site occupation  $\langle n_i \rangle$  and considered both closed [7] and open [8] systems. In the former case, exact solutions were found, leading them to a current versus density relation. For the latter, the authors resorted to numerical solutions to find the phase diagram for a variety of initiation and termination rates. A phase diagram similar to the simple TASEP, as well as nontrivial density profiles and the associated currents, was obtained. More recently, there is renewed interest in this problem. Naming this system “ $\ell$ -TASEP,” Sasamoto and Wadati [26] focused on the time dependence of  $M$  particles in an infinite lattice and, using the Bethe ansatz, found exact results for the conditional probability that the particles are found at certain sites given an initial configuration. The main conclusion is that the dynamics of  $\ell$ -TASEP lies in the same universality class as ordinary TASEP. This line of inquiry has been further generalized to a system containing a *distribution* of particle sizes [27,28]. Though these investigations produced interesting results, they are not applicable to our situation, namely, finite systems with open boundaries. In particular, for finite lattices, these studies are restricted to *closed* systems, for which the stationary states are trivial, as we will recapitulate in Sec. II. Finally, a recent work by Lakatos and Chou [29] considered uniform open systems with extended objects. Using a discrete Tonks gas partition function, they derived the current versus density relation first presented by MacDonald *et al.* [7]. Via a refined mean field theory, they extended this result to predict currents and bulk densities for the open system, which they confirmed by Monte Carlo simulations. The phase diagram and its properties are consistent with those initially obtained by MacDonald and Gibbs [8]. We are not aware of published results on open systems with both extended objects and quenched disorder.

## II. TASEP OF EXTENDED OBJECTS ON A RING

For simplicity and mathematical reasons, it is convenient to discuss a uniform closed system with periodic boundary

conditions. Here,  $M$  identical particles of size  $\ell$  move on a ring of  $N$  sites. Due to the excluded volume constraint, there are  $\tilde{M} = N - \ell M$  uncovered sites (holes). The system evolves by discrete time steps, with random sequential updates. Specifically, during each MCS,  $M$  particles are chosen at random in sequence, and each is moved forward one site provided there is headway. Since we have a stochastic dynamics, the complete description of this system is  $P(C, t)$ , the probability that it is found in configuration  $C$  after  $t$  steps (starting from some initial  $C_0$ ). To label a configuration, we choose an arbitrary particle to be the first ( $\alpha = 1$ ) and supply  $\{h_\alpha\}$ , the set of number of holes in front of the  $\alpha$ th particle (with  $\alpha = 1, \dots, M$ ). Clearly, we may also think of a configuration as a series of  $M$  “gaps” (between the particles) with  $h_\alpha$  being the number of holes in the  $\alpha$ th gap. So,  $P(h_1, h_2, \dots, h_M; t)$  is an explicit form for  $P(C, t)$ . Note that, since the system is closed, there is a constraint on  $h_\alpha$ , i.e.,

$$\sum_{\alpha} h_{\alpha} = \tilde{M}$$

is a constant in time.

Random sequential updating can be translated into an equation that governs the time evolution of  $P(C, t)$ , namely, a master equation. Starting with the initial  $P(C, 0) = \delta(C, C_0)$  (where  $\delta$  is the Kronecker delta),  $P(C, t)$  is expected to settle into a unique, *time-independent* distribution,  $P^*(C)$ , which we will refer to as the “steady state.” If this system were evolving towards thermal equilibrium, the dynamics would satisfy detailed balance [30], and  $P^*(C)$  would be given by the well-known Boltzmann factor. However, the dynamics of our system definitely violates detailed balance, so that, associated with a *nonequilibrium* steady state,  $P^*(C)$  is not known in general. Fortunately, this closed system belongs to a class for which a simple solution is known [31]; namely, every configuration occurs with equal probability [32]. Thus,  $P^*(C)$  is precisely the reciprocal of the total number of configurations consistent with the given parameters  $(N, M, \ell)$ . With such a simple distribution, we can compute many quantities of interest, such as the probability distribution of the current (and hence the average current) and the headway.

Apart from overall factors, the total number of configurations is just  $Z(\tilde{M}, M)$ , the total number of lists  $\{h_\alpha\}$  subject to Eq. (1). Because this is a well-known combinatorial problem (appearing in, e.g., the Bose gas), we simply quote the result,

$$Z(\tilde{M}, M) = \frac{(\tilde{M} + M - 1)!}{\tilde{M}!(M - 1)!} \equiv \binom{\tilde{M} + M - 1}{\tilde{M}}.$$

The actual number of distinct configurations is  $(N/M)Z(\tilde{M}, M)$ , because there are  $N$  lattice sites on which the first particle can be placed but the  $M$  particles are identical. Thus,

$$P^*(C) = \frac{\tilde{M}!M!}{N(\tilde{M}+M-1)!}$$

independent of  $C$ .

To compute the probability distribution of the current and its average, we need more detailed information on the above ‘‘partition.’’ Defining the current  $J$  as the number of particles that moved in one step (normalized by the system size  $N$ ), we see that we will need  $H$ , the number of gaps with one or more holes. The reason is that, for each such gap, the ribosome behind it can move and contributes one ‘‘unit’’ to the current, so that

$$J = \frac{H}{N}.$$

Since all configurations are equally probable, the statistical weight associated with this  $J$  is just the total number of configurations with a given  $H$ . This quantity, denoted by  $Z(H; \tilde{M}, M)$  in analog to  $Z(\tilde{M}, M)$ , may be found from its definition:

$$Z(H; \tilde{M}, M) = \sum_{\{h_\alpha\}} \left[ \delta\left(\tilde{M}, \sum_\alpha h_\alpha\right) \delta\left(H, \sum_\alpha [1 - \delta(h_\alpha, 0)]\right) \right],$$

where the sum is over all possible lists  $\{h_\alpha\}$  and the Kronecker  $\delta$ 's select only those that satisfy Eq. (1) and have  $H$  gaps with  $h_\alpha > 0$ . Once  $Z(H; \tilde{M}, M)$  is known, the full distribution for the current is

$$p(J; \tilde{M}, M) = \frac{Z(H; \tilde{M}, M)}{Z(\tilde{M}, M)}.$$

Now, the explicit form of  $Z(H; \tilde{M}, M)$  can be obtained either through the generating function

$$W_M(\zeta, \eta) \equiv \sum_{\tilde{M}, H} Z(H; \tilde{M}, M) \zeta^{\tilde{M}} \eta^H = \left[ 1 + \frac{\zeta \eta}{1 - \zeta} \right]^M$$

or by standard combinatorial techniques,

$$Z(H; \tilde{M}, M) = \binom{M}{H} \binom{\tilde{M} - 1}{H - 1}.$$

Thus, the explicit current distribution is

$$p(J) = \binom{M}{H} \binom{\tilde{M} - 1}{H - 1} \bigg/ \binom{\tilde{M} + M - 1}{\tilde{M}}, \quad (2)$$

where  $H$  on the right stands for  $JN$ . An alternate form, showing the dependence of this distribution on the control parameters ( $\tilde{M}, M, N = \tilde{M} + \ell M$ ) is

$$p(J | \tilde{M}, M, N) = \frac{1}{M \tilde{M} (\tilde{M} + M - 1)!} \frac{JN}{(\tilde{M} - JN)! (M - JN)!} \left[ \frac{M! \tilde{M}!}{(JN)!} \right]^2.$$

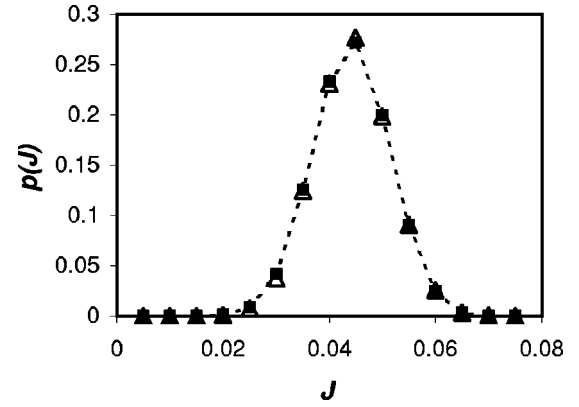


FIG. 1. Distribution of currents  $p(J)$  for  $N=200$ ,  $M=15$ , and  $\ell=12$ . Squares (connected by dashed lines) are values predicted by Eq. (2), and triangles are values observed in Monte Carlo simulations. A single lattice was simulated to steady state, and instantaneous current ( $H/N$ ) was determined every 100 MCS thereafter for  $1.2 \times 10^6$  MCS.

To illustrate this distribution, we show in Fig. 1 both this prediction and simulation data for the case of  $N=200$ ,  $M=15$ , and  $\ell=12$ . Clearly, there is excellent agreement between theory and simulation.

Containing less information, but easier to grasp, is the average current  $\bar{J} \equiv \sum J p(J)$ . Its computation is somewhat easier, since it is  $\sum_{H=1}^M H Z(H; \tilde{M}, M) / N Z(\tilde{M}, M)$ , with the numerator easily gleaned from  $\partial_\eta W_M(\zeta, \eta) |_{\eta=1}$ . The result is

$$\bar{J} = \frac{M}{N} \frac{\tilde{M}}{\tilde{M} + M - 1}. \quad (3)$$

As we will see, the dependence of this average current on the density of particles plays a central role. Expressing this quantity in terms of  $\rho_r \equiv M/N$ , we have  $\bar{J} = \rho_r (1 - \ell \rho_r) / [1 - (\ell - 1) \rho_r - 1/N]$ . An appealing form, which displays both the intensive nature of  $J$  and its underlying particle-hole symmetry, is

$$\bar{J} = \frac{\rho_r \rho_h}{\rho_s - 1/N}. \quad (4)$$

A third form, frequently referred to in the literature as the ‘‘current-density relationship,’’ is writing  $\bar{J}$  as a function of  $\rho$ , the coverage density ( $\rho \in [0, 1]$ ),

$$\bar{J}(\rho) = \frac{\rho}{\ell} \frac{1 - \rho}{1 - \rho + \rho/\ell - 1/N}. \quad (5)$$

In the limit  $N \rightarrow \infty$ , this result was first presented by MacDonald *et al.* [7]. A generalization of the well-known expression for  $\ell = 1$  [i.e.,  $\bar{J} = \rho(1 - \rho)$ ], this  $\bar{J}(\rho)$  is no longer symmetric about  $\rho = 1/2$ . Instead, the optimal density increases from  $1/2$  to  $\sqrt{\ell}/(1 + \sqrt{\ell})$ , while the maximum current is lowered from  $1/4$  to  $(1 + \sqrt{\ell})^{-2}$ . As these quantities will appear frequently, we will denote them by

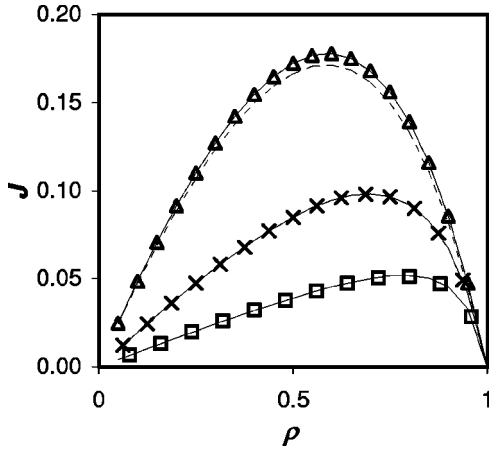


FIG. 2. Current  $J$  vs coverage density  $\rho = M\ell/N$  for closed systems. Symbols are Monte Carlo results, solid curves are predicted values from Eq. (3), and the broken curve is the prediction from Eq. (7). Triangles are for  $\ell=2$ ,  $N=40$ ,  $\times$ 's for  $\ell=5$ ,  $N=80$ , and squares for  $\ell=12$ ,  $N=150$ .  $J$  was determined by averaging over  $1.2 \times 10^5$  MCS and 100 identical systems after steady state was reached.

$$\hat{\rho} \equiv \frac{\sqrt{\ell}}{1 + \sqrt{\ell}} \quad \text{and} \quad \hat{J} \equiv (1 + \sqrt{\ell})^{-2}. \quad (6)$$

To appreciate this shift graphically, we present, in Fig. 2, both analytic and simulation results. In addition, to show the effects of the finite size corrections (due to the  $1/N$  term), we include a curve of the limiting form [7,29]

$$\bar{J} \rightarrow \frac{\rho_r \rho_h}{\rho_s} \quad (7)$$

for the  $N=40$  case.

In connection with these expressions, a conclusion for *conditional* probabilities can be drawn. Since  $\bar{J}$  is precisely the joint probability of finding a ‘‘covered’’-hole pair, we see that  $\rho_r/\rho_s$  is the probability that site  $i$  is covered, given that site  $i+1$  is empty and, similarly,  $\rho_h/\rho_s$  is the probability that site  $i+1$  is empty, given that site  $i$  is covered. These conditional probabilities will play a role in our understanding of the behavior of open systems.

In addition to the average current, we can also compute its fluctuations exactly,

$$\Delta J^2 \equiv \sum J^2 p(J) - \bar{J}^2 = \frac{\bar{J}^2}{\rho_s N}.$$

Typical of noncritical thermodynamic systems, in which the (fractional) deviations are  $O(N^{-1/2})$ , the full distribution  $p(J)$  approaches the standard Gaussian form:  $\exp[-(N/2)\rho_s(J/\bar{J}-1)^2]$ . Indeed, this is the form we see in the example shown in Fig. 1.

Finally, we turn to another quantity of interest: the statistics of ‘‘headway.’’ Since there are  $M$  gaps in each configuration and there are  $Z(\tilde{M}, M)$  configurations, we have a total of  $MZ(\tilde{M}, M)$  gaps. Out of these, we wish to compute the

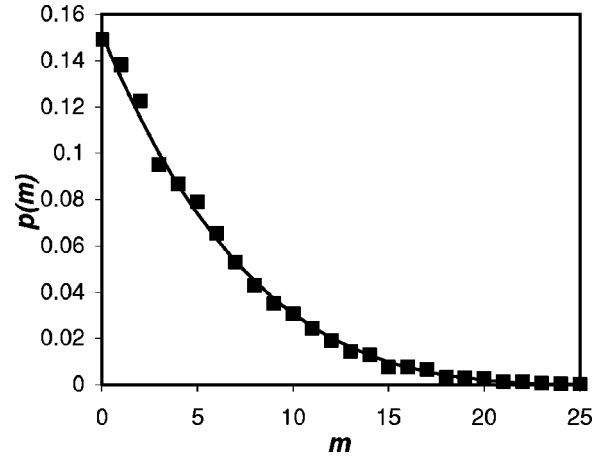


FIG. 3. Distribution of head spacings  $p(m)$  for  $N=100$ ,  $M=6$ , and  $\ell=12$ . The curve is the prediction from Eq. (8), and squares are simulation values. 100 lattices were simulated to steady state, and head spacing  $m$  for any particles on the first lattice sites was determined every 100 MCS thereafter for  $1.2 \times 10^5$  MCS.

number of gaps that contain precisely  $m$  holes, which we denote by  $Z(m; \tilde{M}, M)$ . From its definition

$$Z(m; \tilde{M}, M) \equiv \sum_{\{h_\alpha\}} \delta\left(\tilde{M}, \sum_\alpha h_\alpha\right) \left( \sum_\alpha \delta(m, h_\alpha) \right),$$

we find the associated generating function

$$W(\zeta, \eta) \equiv \sum_{\tilde{M}, m} Z(m; \tilde{M}, M) \zeta^{\tilde{M}} \eta^m = M \left( \frac{1}{1 - \zeta \eta} \right) \left( \frac{1}{1 - \zeta} \right)^{M-1},$$

leading to the distribution of head spacings

$$p(m; \tilde{M}, M) \equiv \frac{Z(m; \tilde{M}, M)}{MZ(\tilde{M}, M)} = \binom{\tilde{M} + M - m - 2}{\tilde{M} - m} \binom{\tilde{M} + M - 1}{\tilde{M}}. \quad (8)$$

In the limit of large  $N$ , this expression simplifies to

$$p(m) \rightarrow \frac{\rho_r}{\rho_s} \left( \frac{\rho_h}{\rho_s} \right)^m.$$

This distribution is reproduced faithfully in Monte Carlo simulations, as shown by an example in Fig. 3. It is easy to understand this result intuitively if we regard  $\rho_h/\rho_s$  as the probability of having a single hole in the headway. With independent hole statistics, we have  $p(m) \propto (\rho_h/\rho_s)^m$ . The average number of holes in the headway and the associated standard deviation can be computed easily:  $\rho_h/\rho_r$  and  $\sqrt{(\rho_h/\rho_r)^2 + \rho_h/\rho_r}$ , respectively.

Though the system considered in this section, particles traveling on a ring, bears little resemblance to the translation process, it is sufficiently simple for us to derive a number of exact results. Apart from their own interest, these results provide crucial insights, such as the current-density relationship,

for postulating appropriate equations in a coarse grained, mean field approach to the physical problem at hand.

### III. EXTENDED OBJECTS IN OPEN SYSTEMS

In this section, we attempt the next step towards a realistic model for protein synthesis by considering systems with open boundaries. The section is organized as follows. We describe the problem and introduce some terminology. Next, we study the effects of the open boundaries alone, by analyzing the continuum limit of a *symmetric* exclusion process. Finally, we consider the asymmetric exclusion process with open boundaries. We present the phase diagram found from simulations and an extremal principle analysis, and we show the ability of the continuum limit to predict steady state density profiles.

In the open system, the first site ( $i=1$ ) is no longer “in front of” the  $i=N$  site. Instead, a particle (of extent  $\ell$ ) will be placed at  $i=1$  with probability  $\alpha$  (previously labeled by  $p_0$ ), *provided* all the first  $\ell$  sites are empty. This models the initiation process. As for elongation, we continue to restrict ourselves to uniform rates here. In the language of TASEP, every (randomly) chosen particle *will move* by one site (increasing  $i$  by unity) if it has some headway. For particles on the last  $\ell$  sites, there is no hinderance, so they will always move if chosen. Finally, to simulate termination, a particle on the  $N^{\text{th}}$  site will be removed from the system with rate  $\beta$  (previously labeled by  $p_N$ ). To repeat, in a Monte Carlo step (MCS),  $M+1$  particles are chosen randomly (in sequence) to attempt a move. They are selected from a pool including the  $M$  particles on the lattice plus an unbound particle that may initiate.

Since this system is no longer closed,  $M$  and  $\tilde{M}$  are fluctuating quantities. Of course, their average values will be controlled by the rates  $\alpha$  and  $\beta$ . Our goal is to find the average densities and the average current of such a model, as functions of  $(\alpha, \beta)$ . In the  $\ell=1$  case, it is known that the system exhibits three different “phases” as  $\alpha$  and  $\beta$  are varied, only one of which resembles the closed system above (in the sense that the current approaches  $\hat{J}$  for large  $N$ ). Beyond overall averages, we seek the density profile, which will not only be nontrivial, but which also displays drastically different properties as we move about in the  $\alpha$ - $\beta$  plane. The main goal of this paper is to study the effects of  $\ell > 1$  on both the phase diagram and these density profiles. Unfortunately,  $P^*(C)$  for a system with open boundaries is not known in general. Even for the  $\ell=1$  case, only a limited set of quantities may be computed exactly. To make progress, we resort here to a more phenomenological approach, in the spirit of hydrodynamics or Lanudau-Ginzburg free-energy functionals. Considering coarse-grained densities and the continuum limit, we postulate equations of motion, based on some of the properties of the closed system. In principle, such equations can be “derived” from the master equation for  $P(C, t)$ , using the mean-field approximation (i.e., ignoring all correlations). In a sense, this is also the approach of MacDonald *et al.* [7], except that they focused on  $n_i^{(L)}(t)$  and  $n_i^{(0)}(t)$ , the probabilities for site  $i$  to be occupied by a ribosome and a

hole, respectively, at time  $t$ . By keeping the spatial coordinate discrete, they faced difference equations and succeeded in finding solutions only numerically. In contrast, by using a continuous spatial coordinate, we have ordinary differential equations instead, giving us *analytic* solutions. However, lost in the notion of “coarse-graining” are the period  $\ell$  structures that feature prominently behind “blockages.” Nevertheless, our approach appears to capture the essential (gross) features of these systems.

In our heuristic approach, we imagine  $N$  to be large enough to justify taking the continuum limit, i.e., replacing the discrete site label  $i$  by a continuous coordinate:  $x$ . For simplicity, define

$$x \equiv i/N$$

so that  $x$  lies within the unit interval. (If physical units of length are desired, we may introduce  $a$  as the lattice spacing, corresponding to the length of a codon, i.e., three bases of the mRNA. Then  $L_{mRNA} \equiv Na$  would be the length of the mRNA in question.) Similarly, continuous *local* densities will take the place of the discrete occupation variables. For example, the coverage density  $\rho(x)$  will be used instead of  $n_i$ . Since the maximum occupancy is unity, the hole density is just

$$\rho_h(x) = 1 - \rho(x).$$

Following our considerations above, we also define the ribosome density by

$$\rho_r(x) = \rho(x)/\ell.$$

Note that, despite the continuum limit, these equations display the meaning of  $\ell$ , which serves as a measure of the “size” (or “extent”) of ribosome. Of course, we are also interested in their time dependence, so that we must consider  $\rho(x, t)$  in general. Now, ribosomes rarely detach from the mRNA during elongation, so that we are justified in regarding these densities as *conserved* fields. Thus, the appropriate equation of motion is the continuity equation, i.e.,

$$\frac{\partial}{\partial t} \rho_r(x, t) = -\vec{\nabla} \cdot \vec{J}_r = -\frac{\partial}{\partial x} J_r(x, t), \quad (9)$$

where  $J_r(x, t)$  is the local (ribosome) current. Our first task is to find how this current depends on the (local) density  $\rho_r(x, t)$ , i.e., to find the functional form for  $J_r[\rho_r]$ . Then we will arrive at an acceptable equation for  $\rho_r$ . To find the steady state profile  $\rho_r^*(x)$ , which satisfies  $\partial_t \rho_r^* = 0$ , we see that this state is associated with a constant ( $x, t$ -independent) current  $J_r^*$ . Therefore, our problem consists of finding the solution to

$$J_r[\rho_r^*] = \text{const},$$

subject to the appropriate boundary conditions. To keep the notation simple, from here on we will drop the subscript  $r$  and write  $J$  or  $J(x, t)$  for the local current (as well as  $J^*$  for the constant current in the stationary state).

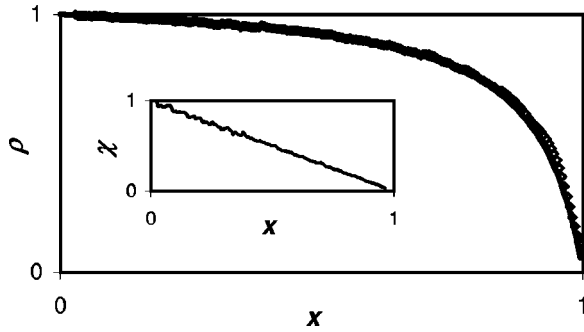


FIG. 4. Density profile for free diffusion in a system with  $N=200$  and  $\ell=12$  from  $\rho=1$  to  $\rho=0$ . Symbols are Monte Carlo data. 100 identical systems were simulated in parallel for  $10^5$  MCS to reach steady state, and then density profiles were collected every 100 MCS for an additional  $1.2 \times 10^4$  MCS. The curve is the predicted density profile. The inset shows the effective particle density profile  $\chi$ , constructed from the simulation data. The current observed in simulations was  $J^*=0.00252$ , which is close with the expected value of  $0.0025=D[\chi(0)-\chi(1)]$  for  $D=1/2N$ .

#### A. The case of free diffusion and an effective density variable for extended objects

To show how we build an appropriate equation for the TASEP, let us begin with a study of the effects of open boundaries *alone*. In other words, let us consider a *symmetric* exclusion process for extended objects, i.e., a system of large particles diffusing freely on a line, subjected to the excluded volume constraint and the controls  $\alpha, \beta$ . For simulations, a randomly chosen particle is moved one site forward or backward with equal probability (0.5), provided it does not run into its neighbor. The only exception is the first particle, which is prohibited from jumping backwards into the “source.”

For the  $\ell=1$  case ( $\rho=\rho_r$ ), the steady state profile is trivially linear,

$$\rho^*(x) = (1-x)\rho^*(0) + (x)\rho^*(1)$$

with current

$$J^* = \frac{1}{N}[\rho^*(0) - \rho^*(1)].$$

(Note that, since the current is controlled by the gradient of the local density only, it vanishes necessarily in the  $N \rightarrow \infty$  limit. So, we must keep the  $N$  explicitly here.) Meanwhile, the boundary densities are fixed by matching the injection/depletion rates to the internal current:

$$\alpha\rho_h^*(0) = \alpha[1 - \rho^*(0)] = J^* = \beta\rho^*(1),$$

and our problem is completely solved. These well known results can be traced to the fact that Eq. (9) assumes the form of the simple diffusion equation  $\partial_t \rho \propto \partial_x^2 \rho$ , since  $J[\rho] \propto -\nabla \rho$ . Though there are many ways to arrive at this result, it is less clear how to generalize it to the case of extended objects. In particular, as displayed in Fig. 4, the profile from simulations (for  $N=200, \ell=12, \alpha=\beta=1$ ) is far from linear.

Here, we will show that there is a natural generalization for the (particle) density in the case of extended objects and that its profile is again linear.

We proceed by returning to the basics, starting with the current  $J_r$  being proportional to both the conductivity (mobility) and the drive. For the former, we take the result from the preceding section. There, the drive is constant, so that the right hand side of Eq. (7),  $\rho_r \rho_h / \rho_s$ , can be interpreted as the conductivity or mobility (neglecting the finite-size effect term from the closed system). Meanwhile, the drive for free diffusion should be the gradient of a pressure  $\mathcal{P}$ , so that

$$J_r[\rho] = D \left[ \frac{\rho_r \rho_h}{\rho_s} \right] [-\nabla \mathcal{P}], \quad (10)$$

where  $D$  is a constant, to be fitted to data. Since we have scaled the system size to unity, we should keep in mind that  $D$  is a quantity of  $O(1/N)$ , to be consistent with the continuum limit. For the pressure  $\mathcal{P}$ , we follow the standard route of statistical mechanics and write

$$\mathcal{P} = \frac{\delta \mathcal{H}}{\delta \rho_r},$$

where  $\mathcal{H}$  is a free energy functional (e.g., the Landau-Ginzburg “Hamiltonian” in case of the ordinary lattice gas). Here, we have a noninteracting system, so that a reasonable form for  $\mathcal{H}$  is just the entropy [35]:

$$\mathcal{H} = \int dx [\rho_r \ln \rho_r + \rho_h \ln \rho_h - \rho_s \ln \rho_s].$$

Using

$$\rho_h = 1 - \ell \rho_r, \quad \rho_s = 1 - (\ell - 1) \rho_r$$

and carrying out the steps, we arrive at a *modified* diffusion equation:

$$\partial_t \rho_r = D \partial_x [\rho_s^{-2} \partial_x \rho_r]. \quad (11)$$

Note that, for  $\ell > 1$ , the effective “diffusion constant,”  $D/\rho_s^2$ , is *density dependent*. Thus, the steady state profile will not be linear in  $x$ . Instead, it satisfies

$$\frac{D}{(\rho_s^*)^2} \frac{\partial \rho_r^*}{\partial x} = -J^*, \quad (12)$$

which will definitely lead to nonlinear profiles.

Although we have derived this diffusion equation via phenomenological techniques, we speculate that it could be obtained through a more rigorous derivation. Specifically, if a matrix product description of the  $\ell > 1$  system becomes available, the methods of Derrida *et al.* [36] may extend to calculate a large deviation functional, which would play the role of a nonequilibrium free energy for this system. Should such a functional exist, we expect that our diffusion equation would result from a variation of this functional.

To find  $\rho_r^*$  explicitly, notice that, due to  $\rho_s = 1 - (\ell - 1)\rho_r$ , the left hand side is just a simple derivative of

$1/\rho_s^*$ . Thus, the profile  $1/\rho_s^*(x)$  will be *linear*. However,  $\rho_s$  does not reduce to a sensible variable as  $\ell \rightarrow 1$ , leading us to define

$$\chi \equiv \frac{1}{\ell-1} \left[ \frac{1}{\rho_s^*} - 1 \right] = \frac{\rho^*}{\ell - (\ell-1)\rho^*}.$$

Not only does  $\chi$  lie in the unit interval; it reduces to the usual particle density  $\rho^*$  in the limit  $\ell \rightarrow 1$ . Most importantly, it is uniquely related to  $\rho^*$  for all  $\ell$ . First introduced in Ref. [28],  $\chi$  is the ‘‘natural’’ generalization of the particle density for extended objects. We will refer to it as the effective particle density (EPD). Meanwhile, a hole remains a single-site entity, so its density needs no modification. In terms of the EPD, the various densities are

$$\rho_s^* = \frac{1}{1 + (\ell-1)\chi},$$

$$\rho_r^* = \frac{\chi}{1 + (\ell-1)\chi},$$

$$\rho_h^* = \frac{1-\chi}{1 + (\ell-1)\chi},$$

so that  $\chi$  is just

$$\chi \equiv \rho_r^* / \rho_s^*. \quad (13)$$

The current-density relationship [Eq. (7)] for the steady state in the *ring* is again a simple product,

$$\bar{J}_{ring} = \chi \rho_h. \quad (14)$$

Note that this product will serve as the *mobility* in our mean field approach here and can be expressed in terms of  $\chi$  alone,

$$\chi \rho_h = \frac{\chi(1-\chi)}{1 + (\ell-1)\chi}.$$

The main advantage of using the EPD is the reemergence of the familiar combination  $\chi(1-\chi)$  in the numerator. Finally, corresponding to the optimal density [Eq. (6)], we have

$$\hat{\chi} = \frac{1}{1 + \sqrt{\ell}},$$

a quantity that plays a significant role in systems with open boundaries. Note that the corresponding hole density ( $1 - \hat{\rho}$ ) also assumes the same value, so that  $\hat{J}$  is just  $\hat{\chi}^2$ . The underlying ‘‘particle-hole’’ symmetry generalizes to one under exchange of  $\chi \leftrightarrow \rho_h$ , or

$$\chi \leftrightarrow \frac{(1-\chi)}{1 + (\ell-1)\chi},$$

or, in terms of the more physical  $\rho$ ,

$$\rho \leftrightarrow \frac{1-\rho}{1-\rho+\rho/\ell}.$$

We now find that Eq. (12) simplifies to the familiar equation from ordinary diffusion,

$$D \frac{\partial \chi}{\partial x} = -J^*.$$

As a result, the profile in terms of  $\chi$  is again linear in  $x$  [37]. For completeness, we write the solution

$$\chi(x) = (1-x)\chi(0) + (x)\chi(1)$$

with

$$\chi(0) - \chi(1) = J^*/D.$$

The explicit solution ensues once the constraints of injection/depletion rates are imposed. Note that, in general, we would have  $J^* \propto D$ , so that the current inherits the  $O(1/N)$  from  $D$ , in contrast to the  $O(1)$  behavior for systems with nonzero drive.

To close, we illustrate how well the theory agrees with data by showing the analytic, nonlinear profile of  $\rho^*(x)$  with boundary conditions  $\rho(0) = 1$ ,  $\rho(1) = 0$ :

$$\rho^*(x) = \frac{1-x}{1 - (1-1/\ell)x},$$

as a curve in Fig. 4.

## B. TASEP in open systems

We now turn to the other extreme case, where backward jumps are completely excluded (TASEP), generalizing the model of Ref. [3] to extended objects. With open boundaries, this model incorporates another essential feature of translation. Our interest is again the average current and density profile, as a function of  $\alpha$  and  $\beta$ , the rates of initiation and termination.

### 1. Phase diagram from extremal principle and simulations

Although we are unable to generalize the methods of Refs. [3,4] to arrive at exact solutions for an open system with particles covering  $\ell > 1$  sites, we find that the phase diagram determined by the mean field techniques of Refs. [8,29] can be understood also by extending the extremal principle, an hypothesis first proposed by Popkov and Schütz [38] and later exploited successfully by others (e.g., Refs. [39,40]). In this approach, the open boundaries are regarded as connections to reservoirs with appropriate densities, so that by keeping the same jump rates as in the bulk,  $\alpha$  and  $\beta$  are realized. Defining  $\rho_-$  and  $\rho_+$  as the reservoir densities at the initiation and termination boundaries, respectively, the extremal principle relates the current in the open system to the  $J[\rho]$  for a closed, periodic system with the same bulk dynamics [41]:



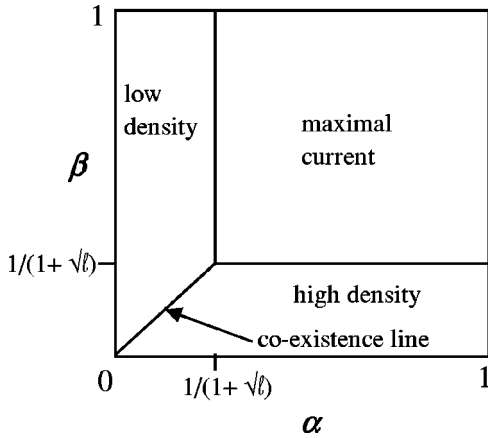


FIG. 5. Phase diagram predicted by the extremal principle.

$$J = \begin{cases} \max J[\rho] & \text{for } \rho_+ < \rho < \rho_- \\ \min J[\rho] & \text{for } \rho_- < \rho < \rho_+ . \end{cases}$$

Unfortunately, there is no prescription for finding  $\rho_-, \rho_+$  from  $\alpha, \beta$  in general. Exploiting results from the exactly solvable  $\ell = 1$  case (where  $\rho_- = \alpha$  and  $\rho_+ = 1 - \beta$ ) and from previous studies of the  $\ell > 1$  case [8,29], we argue in favor of

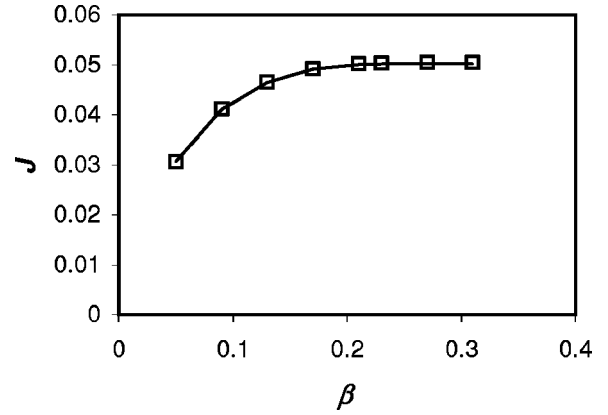
$$\rho_-(\alpha) = \frac{\ell \alpha}{1 + \alpha(\ell - 1)} \quad \text{and} \quad \rho_+(\beta) = 1 - \beta. \quad (15)$$

These reservoir densities can be understood as follows. Recall that, for a closed system, the probability for site  $i - 1$  to be filled, *given* that site  $i$  is empty, is  $\rho_r/\rho_s = (\rho/\ell)/(1 - \rho + \rho/\ell)$ . We now argue that when the first site of an open system is empty and will be filled with probability  $\alpha$ , it can be thought of as being coupled to a reservoir of the appropriate density, i.e.,  $\rho_-$  such that  $\alpha = (\rho_-/\ell)/(1 - \rho_- + \rho_-/\ell)$ . Solving for  $\rho_-$  leads to the expression above for  $\rho_-(\alpha)$ . The expression for  $\rho_+(\beta)$  is not readily explained by similar arguments, so discussion of its origin will be deferred until the following section. An important feature associated with this choice is that the current is symmetric under  $\alpha \leftrightarrow \beta$ , as observed in simulations. Most importantly, these choices lead to phase diagrams in good agreement with Monte Carlo data.

By combining these choices with the extremal principle, we find that, although the result is qualitatively similar to the  $\ell = 1$  system, there are quantitative changes for the  $\ell > 1$  case. First, there is a shift in the location of transition lines, from  $1/2$  to  $\hat{\chi} = 1/(1 + \sqrt{\ell})$  (as shown in Fig. 5). Then, the current and bulk densities are also modified:

$$\bar{J}(\alpha, \beta) = \begin{cases} \frac{\alpha(1 - \alpha)}{1 + \alpha(\ell - 1)} & \text{for } \alpha < \beta < \hat{\chi} \quad (\text{low density}), \\ \frac{\beta(1 - \beta)}{1 + \beta(\ell - 1)} & \text{for } \beta \leq \alpha < \hat{\chi} \quad (\text{high density}), \\ \hat{J} & \text{for } \alpha, \beta \geq \hat{\chi} \quad (\text{max current}) \end{cases} \quad (16)$$

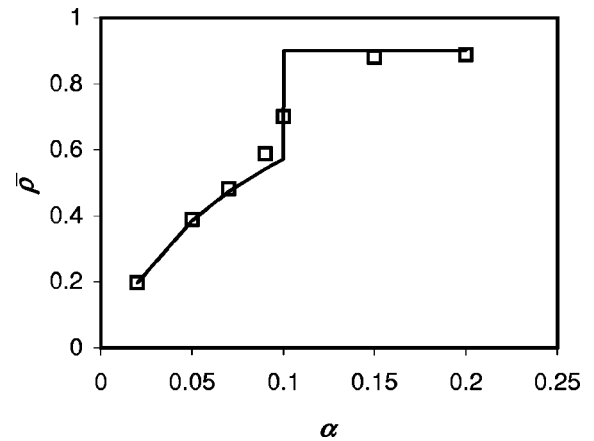
and


 FIG. 6. Dependence of current on  $\beta$  for  $\alpha = 1$ ,  $\ell = 12$ , and  $N = 1000$ . Symbols are simulation results (determined from 100 systems simulated in parallel for  $1.2 \times 10^4$  MCS after steady state was reached) and the curve is the prediction from Eq. (16).

$$\bar{\rho}(\alpha, \beta) = \begin{cases} \rho_- & \text{for } \alpha < \beta < \hat{\chi} \quad (\text{low density}), \\ \rho_+ & \text{for } \beta < \alpha < \hat{\chi} \quad (\text{high density}), \\ \hat{\rho} & \text{for } \alpha, \beta \geq \hat{\chi} \quad (\text{max current}). \end{cases} \quad (17)$$

As examples of how well these predictions fit simulation data, we show plots of  $\bar{J}(1, \beta)$  (Fig. 6) and  $\bar{\rho}(\alpha, 0.1)$  (Fig. 7). As expected, when  $\alpha$  was rate limiting (low density phase), a bulk coverage density of  $\rho_-(\alpha)$  was induced, and when  $\beta$  was rate limiting, a bulk coverage density of  $\rho_+(\beta)$  was induced. Given the success of the extremal principle analysis, it would be appropriate to search for more fundamental theories from which the extremal principle could be derived.

The transition between low and high density phases is clearly first order, displayed as a jump in the bulk density plot (Fig. 7). As in the  $\ell = 1$  case [10], domain wall theory can be used to understand our results. In particular, for  $\alpha = \beta < \hat{\chi}$ , we have observed a shock front (between the low


 FIG. 7. Dependence of average coverage density  $\bar{\rho}$  on  $\alpha$  for  $\beta = 0.1$ ,  $\ell = 12$ , and  $N = 1000$ . Symbols are simulation results (determined from 100 systems simulated in parallel and sampled every 100 MCS for  $1.2 \times 10^4$  MCS after steady state was reached) and curve is the prediction from  $\rho_-(\alpha)$  and  $\rho_+(\beta)$  in Eq. (17).

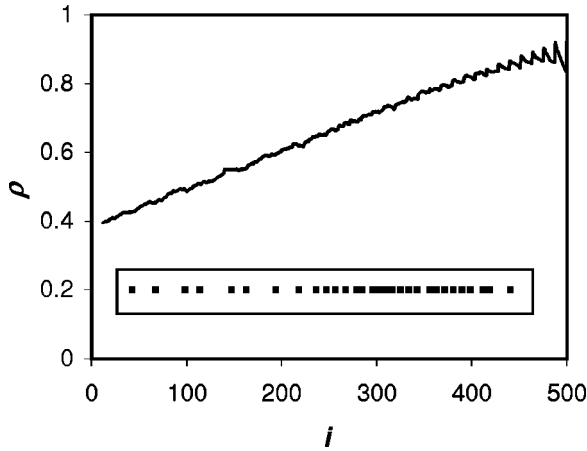


FIG. 8. Steady state density profile for  $\alpha=0.05$ ,  $\beta=0.05$ ,  $N=500$ ,  $\ell=12$ . The profile was obtained from simulations of 100 systems run to steady state and then sampled every 100 MCS for  $1.2 \times 10^4$  MCS. (Nonlinearities near the termination boundary result from ribosomes tending to “pile up” at  $i=N, N-\ell, N-2\ell$ , etc., due to  $\beta < 1$  but with decreasing correlations as  $i$  decreases.) The inset shows a typical configuration of this system, with the shock front near the center. (Each ribosome is represented by a dot.)

and high density regions) diffusing freely along the lattice. As a result, the *average* profile is linear, as shown in Fig. 8. To appreciate the shock, we display in the inset a typical configuration (showing a shock) where each ribosome is represented by a dot. In addition, we exhibit another aspect of this coexistence by sampling the average (coverage) density in the central 10% of a large ( $N=1000$ ) lattice. Compiling a histogram (closed diamonds in Fig. 9), we find a bimodal distribution typical of systems at a first order transition. Peaks are expected to correspond to the densities  $\rho_-(\alpha)$  and  $\rho_+(\beta)$ . To connect with local fluctuations in the “pure” systems, we compile similar histograms for a closed system (i.e., 10% of a ring with 1000 sites) with overall density set at  $\rho_-(\alpha)$  and  $\rho_+(\beta)$ . For comparison, we show a simple linear combination of such distributions in Fig. 9 (open squares). We expect correspondence between densities in the window in the open system, which alternate between approximately  $\rho_-(\alpha)$  and  $\rho_+(\beta)$ , and an appropriately weighted average of window densities in the two closed systems. The deviations can be understood as the contribution of configurations with the shock in the sampling window. Roughly, this may occur about 10% of the time, which is also the order of magnitude of the deviation from the simple average of pure systems.

While most of the features presented here were known to MacDonald *et al.* [7,8] and have also been obtained by Lakatos and Chou [29], our efforts are to go beyond mean field approaches, showing the different perspective offered by domain wall theory and the extremal principle. Shock effects on the coexistence line are particularly well described by this approach.

## 2. Steady state profiles from the continuum approach

Following an understanding of the phase diagram, we turn to more details of the system, namely, steady state density

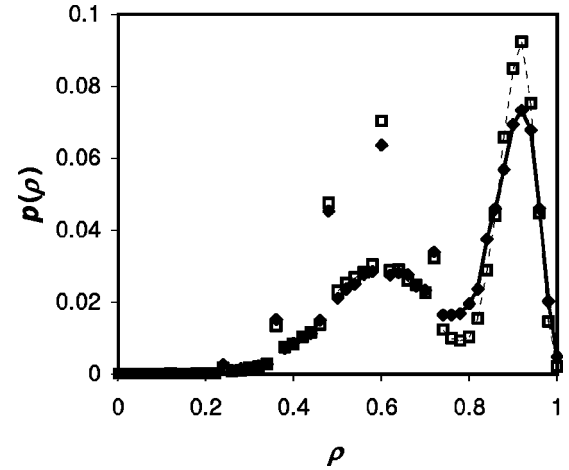


FIG. 9. Probability distribution of densities in the central 10% of a system with  $\alpha=0.1$ ,  $\beta=0.1$ ,  $N=1000$ , and  $\ell=12$ . The actual probability distribution (solid diamonds with solid curve) was obtained by simulating 100 systems to steady state, sampling the central densities every 100 MCS for  $5 \times 10^4$  MCS, and compiling a frequency histogram. Peaks correspond to the densities  $\rho_-(\alpha)$  and  $\rho_+(\beta)$ . For comparison, a composite density distribution (open squares with dashed curve) was assembled from similar frequency histograms for two closed systems with average densities  $\rho_-(0.1)$  and  $\rho_+(0.1)$ . (Again,  $N=1000$ ,  $\ell=12$ , and 10% of the system was sampled.) The composite distribution required a linear combination of 53% low density and 47% high density to minimize the absolute difference from the actual probability distribution. See text for details.

profiles. Our continuum approach again leads us to differential equations. As considerable insight can be gained by studying simple differential equations, we believe it is worthwhile to devote a section to the continuum approach.

Starting with the case of free diffusion, we extend Eq. (11) to the driven system by simply adding an “Ohmic” term to the current:  $J = \sigma E$ , where the drive is associated with strength  $E$  and the conductivity will be the density dependent factor found before [first used in Eq. (10)]. Thus, our starting equation is

$$\frac{\partial \rho_r}{\partial t} = - \frac{\partial}{\partial x} J(x) = \frac{\partial}{\partial x} \left[ \frac{D}{\rho_s^2} \frac{\partial \rho_r}{\partial x} - \frac{\rho_r \rho_h}{\rho_s} E \right]. \quad (18)$$

As before, we suggest that this driven diffusion equation might be obtained from a variation of an appropriate large deviation functional, the  $\ell=1$  version of which has been derived previously [42,43].

As a reminder, our  $x$  is actually the *fractional* length along the mRNA, so that it is dimensionless. Similarly, the densities are also unitless (e.g.,  $\rho_h \in [0,1]$ ), so that  $D/E$  carries the information of the real length of the chain. An estimate of this ratio can be obtained by taking the naive continuum limit of a discrete hopping model, with the result  $1/2N$ . In the Appendix, we will show how this arises from taking such a limit of the MacDonald *et al.* [7] current equation. For here, we may regard  $D/E$  as a phenomenological parameter. Further, since we are focusing only on *totally* asymmetric

processes, we could just as well set  $E$  to unity [as in Eq. (7)], and measure  $J$  and  $D$  “in units of  $E$ .”

Taking these considerations into account, we seek the steady state profile  $\rho_r^*(x)$ , as in Eq. (12), by setting the square bracketed terms in Eq. (18) to the steady state current,

$$D(\rho_s^*)^{-2} \frac{\partial \rho_r^*}{\partial x} - \rho_r^* \rho_h^* (\rho_s^*)^{-1} = -J^*.$$

In terms of our “natural” variable  $\chi$  [Eq. (13)], this equation reduces to

$$\chi' \equiv \frac{\partial \chi}{\partial x} = \frac{-1}{D[1+(\ell-1)\chi]} [\tilde{J}\chi + J^* - \chi(1-\chi)], \quad (19)$$

where  $\tilde{J} \equiv J^*(\ell-1)$ . Thus, we see that this  $\ell > 1$  generalization is quite similar to the  $\ell=1$  equation:  $\rho' = [\rho(1-\rho) - J^*]/D$ .

Now, the zeros of  $\chi'$  will play an important role, occurring at

$$\chi_{>} \equiv \frac{1-\tilde{J}+R}{2} \quad \text{and} \quad \chi_{<} \equiv \frac{1-\tilde{J}-R}{2}, \quad (20)$$

where

$$R \equiv \sqrt{(1-\tilde{J})^2 - 4J^*}.$$

It is not surprising that the maximal current  $\hat{J}$  (i.e.,  $(1 + \sqrt{\ell})^{-2}$  in our system of units) is a key player, so that  $R$  can be written in the following form:

$$R = (\ell-1) \sqrt{(J^* - \hat{J})(J^* - \check{J})}.$$

Here,  $\check{J} \equiv (1 - \sqrt{\ell})^{-2}$  appears as a natural counterpart to  $\hat{J}$ . Another advantage of this form is its relationship to the extremal current principle. As we will see, for finite systems,  $J^*$  will be slightly [ $O(1/N^2)$ ] larger than  $\hat{J}$  for the maximal current phase, giving us complex roots and profiles with inflections. On the other hand, for the other two phases,  $J^* < \hat{J}$ , leading to fixed points in the  $N \rightarrow \infty$  limit.

Equation (19) can be integrated to find an implicit function for the density profile in each phase. For explicit solutions, we must specify the boundary conditions,

$$J^* = \alpha[1 - \rho(0)],$$

$$J^* = \frac{\beta}{1 + \beta(\ell-1)} \rho(N), \quad (21)$$

so that

$$\chi(0) = \frac{\rho(0)/\ell}{1 - \rho(0) + \rho(0)/\ell} = \frac{\alpha - J^*}{\alpha + (\ell-1)J^*} \quad (22)$$

and

$$\chi(1) = \frac{\rho(N)/\ell}{1 - \rho(N) + \rho(N)/\ell} = \frac{[1 + \beta(\ell-1)]J^*}{\ell\beta - (\ell-1)[1 + \beta(\ell-1)]J^*}. \quad (23)$$

The boundary conditions can be understood as follows. From  $\rho_-(\alpha)$  in Eq. (15), we can show that the effective reservoir at the initiation boundary has an EPD of  $\chi_-(\alpha) = \alpha$ . Equation (14) then implies the initiation boundary condition in Eq. (21). Particles on the final  $\ell$  lattice sites experience no steric hindrance, so the current through the final  $\ell$  sites is just

$$J^* = \rho_r(i) \quad \text{for } i = N - \ell + 1, \dots, N - 1,$$

$$J^* = \beta \rho_r(N).$$

These relations can be used to express  $\rho(N) = \sum_{i=N-\ell+1}^N \rho_r(i)$  in terms of  $J^*$ , leading to the termination boundary condition in Eq. (21). The boundary conditions for  $\chi$  will serve to fix the constant of integration as well as  $J^*$ , which is still an unknown in Eq. (19). [The current thus determined is expected to match closely the current predicted by Eq. (16.)] We discuss the various phases separately.

For smaller values of  $J^*$  (i.e.,  $J^* < \hat{J}$ ),  $\chi(x)$  has fixed points at  $\chi_{>}$  and  $\chi_{<}$ . Boundary conditions determine whether the steady state profile approaches  $\chi_{<}$  (low density) or  $\chi_{>}$  (high density). Each solution corresponds to part of the phase diagram shown in Fig. 5. There is a “kink” solution when  $\chi(0), \chi(1) \in (\chi_{<}, \chi_{>})$  with  $\chi(0) \geq \chi_{<}$  and  $\chi(1) \leq \chi_{>}$  that corresponds to the first order transition line.

In the high density phase ( $\beta < \alpha$ ) where  $\chi_{>} > \chi_{<}$ , the density profile is given implicitly by

$$\begin{aligned} [1 + (\ell-1)\chi_{>}] \ln\left(\frac{\chi - \chi_{>}}{\chi_0 - \chi_{>}}\right) - [1 + (\ell-1)\chi_{<}] \ln\left(\frac{\chi - \chi_{<}}{\chi_0 - \chi_{<}}\right) \\ = -\frac{R\chi}{D} \end{aligned}$$

[where  $\chi_0 \equiv \chi(0)$  is the EPD at initiation]. It may be noted that as  $N \rightarrow \infty$ , the bulk density and termination boundary density both approach  $\chi_{>}$ . Setting  $\chi(1) = \chi_{>}$  and eliminating  $J^*$  from the termination boundary condition [Eq. (23)] and the definition of  $\chi_{>}$  [Eq. (20)] yields a bulk  $\chi$  of  $(1 - \beta)/[1 + \beta(\ell-1)]$ . We expect the bulk density in the high density phase to match the termination reservoir density. Indeed, the bulk  $\chi$  value calculated here is consistent with the reservoir density  $\rho_+ = 1 - \beta$  given above [Eq. 15].

Though it is impossible to write a closed form for  $\chi(x)$  in general, a convenient expression is

$$\chi = \chi_{>} + (\chi_0 - \chi_{>}) \exp[-\mu_+ x] \left(\frac{\chi - \chi_{<}}{\chi_0 - \chi_{<}}\right)^\gamma, \quad (24)$$

where

$$\mu_+ \equiv \frac{R}{D[1 + (\ell-1)\chi_{>}]}$$

and

$$\gamma = \frac{1 + (\ell-1)\chi_{<}}{1 + (\ell-1)\chi_{>}}.$$

Note that  $R$ ,  $\chi_{<}$ , and  $\chi_{>}$  can be conveniently estimated using the  $J^*$  value given by Eq. (16). The advantage here is

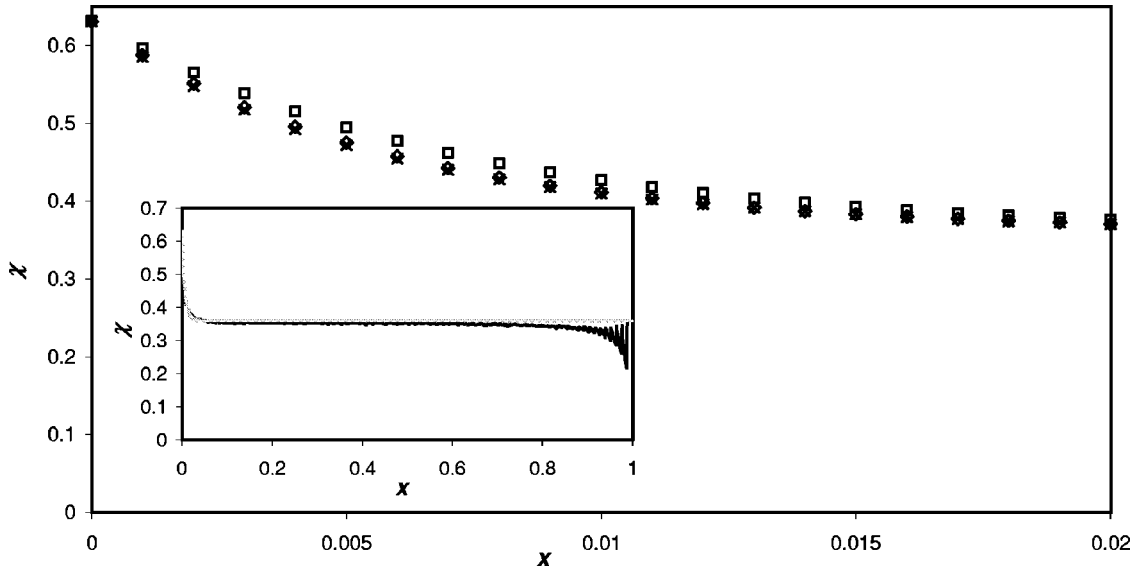


FIG. 10. Rapid convergence of the iterative method to  $\chi(x)$  for small  $x$ . (Larger  $x$  values are omitted because all iterations of  $\chi$  are nearly identical.)  $\chi^{(1)}$ ,  $\chi^{(2)}$ , and  $\chi^{(3)}$  are represented by squares, triangles, and  $\times$ 's, respectively. The system considered here is  $\alpha=1$ ,  $\beta=0.13$ ,  $N=1000$ ,  $\ell=12$ .  $D$  was set to  $1/3N$  to obtain a good fit. The inset compares the actual steady state  $\chi$  from simulations (bold curve) with the predicted  $\chi^{(3)}$  (lighter curve). In the profile from simulations, particle depletion near the termination end can be seen.

that the unknown factor  $(\chi - \chi_{<}/\chi_0 - \chi_{<})^\gamma$  displays only limited variation, since  $\chi$  lies outside the interval  $[\chi_{<}, \chi_{>}]$ . Thus, we expect to find better and better approximations by exploiting an *iterative* scheme. Starting with the first iteration

$$\chi^{(1)} = \chi_{>} + (\chi_0 - \chi_{>}) \exp[-\mu_+ x],$$

this procedure involves substituting repeatedly for the  $\chi$  in the unknown factor of Eq. (24), namely,

$$\chi^{(k)} = \chi_{>} + (\chi_0 - \chi_{>}) \exp[-\mu_+ x] \left( \frac{\chi^{(k-1)} - \chi_{<}}{\chi_0 - \chi_{<}} \right)^\gamma.$$

In Fig. 10, we show an example of three such iterations, converging rapidly to the real  $\chi(x)$ , and in the inset a comparison with the  $\chi(x)$  observed in simulations. Near the termination end of the system, the EPD from simulations shows a depletion below the bulk density. This reduction is characteristic of the high density phase and was originally observed in the numerical results of Ref. [8]. The continuum limit does not capture this feature.

A similar analysis may be carried out for the low density phase when  $\chi < \chi_{<}$ . In this case, the boundary layer occurs at the termination end, so it is convenient to apply the  $\chi(1)$  boundary condition explicitly and use the  $\chi(0)$  boundary condition to determine  $J^*$ . This method also works for high and low density phase profiles when  $\chi(0), \chi(1) \in (\chi_{<}, \chi_{>})$ , as long as  $\alpha$  and  $\beta$  are not too similar. The analysis fails close to the first order transition line, as we would expect since Eq. (19) predicts a “kink” rather than a linear density profile on the transition line.

For the maximal current phase, we know that  $J^* > \hat{J}$ , so that  $R$  is purely imaginary. Since  $\chi'$  is negative definite, we have a downward sloping profile, which is the generalization of the  $\tan(-x/\xi)$  [or  $\cot(x/\xi)$ ] profiles in the ordinary driven

lattice gas [44]. In the large  $N$  limit, a macroscopically large region of the profile will be almost flat (corresponding to  $\chi' \approx 0$ ), where the density assumes the optimal value  $\hat{\rho}$  (or  $\chi \approx \hat{\chi}$ ). Meanwhile,  $J^*$  will approach the maximal current  $\hat{J}$  from above with  $O(1/N^2)$  terms. The details are somewhat involved, so that we will show the solution and discuss its properties only for large  $N$ .

Defining the real quantity  $\tilde{R} = -iR$ , Eq. (19) can be integrated to show that the steady state density profile is given by

$$\ln \left[ \frac{\chi^2 - (1 - \tilde{J})\chi + J^*}{\chi_0^2 - (1 - \tilde{J})\chi_0 + J^*} \right] + \frac{4 + 2(\ell - 1)(1 - \tilde{J})}{\tilde{R}(\ell - 1)} (\theta - \theta_0) = -\frac{2x}{D(\ell - 1)}, \quad (25)$$

where

$$\theta(x) \equiv \arctan \frac{\tilde{R}}{1 - \tilde{J} - 2\chi}$$

and  $\chi_0 = \chi(0)$  is given by Eq. (22). Evaluating Eq. (25) at the termination end [ $x=1, \chi=\chi(1)$ ] and using Eq. (23), we arrive at an equation for determining  $J^*$  in terms of the control parameters  $\alpha, \beta$ . Needless to say, this equation is too complex to solve analytically. Nevertheless, we can gain some insight by considering the large  $N$  limit. Defining  $\epsilon \equiv 1/N$  for convenience, recall that  $D$  is  $O(\epsilon)$ . Next, let us assume that  $J^* = \hat{J} + O(\epsilon^2)$  and show that it is justified later. This leads us to  $\tilde{R}$  being  $O(\epsilon)$  and  $(1 - \tilde{J})/2 = \hat{\chi} + O(\epsilon^2)$ , so that  $\chi_{>}, \chi_{<} = \hat{\chi} + O(\epsilon^2) \pm iO(\epsilon)$ . Thus, the quadratic form in the argument of the  $\ln$  in Eq. (25) never becomes smaller than  $\tilde{R}^2/4 = O(\epsilon^2)$ , so that this term never exceeds  $O(\ln N)$ .

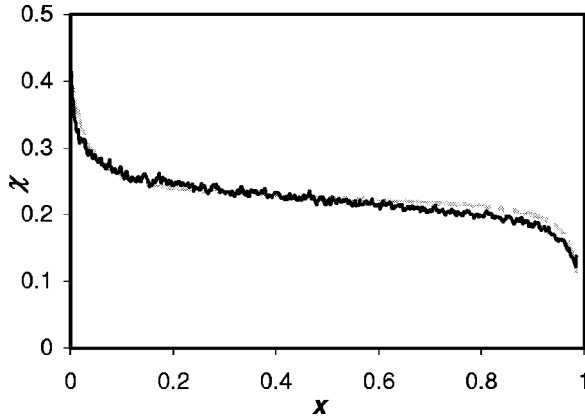


FIG. 11. Actual and predicted steady state density profiles in the maximal current phase. The system considered here is  $\alpha=0.5$ ,  $\beta=0.5$ ,  $N=800$ ,  $\ell=12$ .  $D$  was set to  $4/5N$  to obtain a good fit.  $J^*$  was set to  $\hat{J}+3.24\times 10^{-5}$  to satisfy the termination boundary condition. Profiles are the actual steady state  $\chi$  from simulations (bold curve) and the predicted  $\chi^{(7)}$  (lighter curve).

Since the other terms are generally  $O(N)$  (from  $D$  and  $\tilde{R}$ ), we can write an approximate equation by neglecting the  $\ln$  term, leaving an expression for  $\theta$ , and taking the cotangent of both sides:

$$\chi - \hat{\chi} = \frac{1}{2} \tilde{R} \cot[x/\xi + \text{const.}] + O(\epsilon^2), \quad (26)$$

where  $\xi \cong 2D\sqrt{\ell/\tilde{R}}$  is of  $O(1)$ . Note that the constant here can be written as  $\text{arccot}[2(\chi_0 - \hat{\chi})/\tilde{R}]$ , which vanishes as  $\tilde{R}/(\chi_0 - \hat{\chi})$  in the limit  $\epsilon \rightarrow 0$ . Thus, for typical values of  $x$ , we see that  $\chi = \hat{\chi} + O(\epsilon)$ . For the termination end,  $J^*$  must be carefully fixed so that the argument of the co-tangent in Eq. (26) approaches  $\pi$  in an appropriate way for the right hand side to equal  $\chi(1) - \hat{\chi}$ . This requirement is consistent with the original assumption that  $J^* = \hat{J} + O(\epsilon^2)$ .

Returning to the finite  $N$  case, although  $\chi(x)$  cannot be determined explicitly, we find that an iterative method is again helpful and converges rapidly. This time, successive iterations of  $\chi$  are substituted into the term  $\ln\{\chi^2 - (1 - \tilde{J})\chi + J^*\} / \{\chi_0^2 - (1 - \tilde{J})\chi_0 + J^*\}$  of Eq. (25). As Fig. 11 shows, there is good agreement between the steady state profiles from simulation data and this mean field theory.

#### IV. DISORDERED HOPPING RATES IN THE OPEN SYSTEM

We finally turn to the difficult problem of the TASEP with extended objects and quenched disorder. At each codon of the mRNA, a ribosome translating the mRNA must wait for an appropriate aa-tRNA to decode that codon. aa-tRNA availability ranges over about an order of magnitude [45], and it is thought that the elongation (hopping) rates may also range over an order of magnitude, with the slowest ones comparable to the initiation and termination rates [46].

Since previous studies of the TASEP with quenched disorder involve  $\ell=1$ , their relevance to the process of trans-

lation may be questioned. Indeed, we believe that the system will display essential differences when these point particles are replaced by extended objects. First, as we have seen in the uniform elongation studies above, generalizing to the  $\ell > 1$  case led to significant changes. Now, with nonuniform rates, steric hindrance should play an even larger role, since extended particles can block multiple sites with potentially different rates. Thus, we will devote this section to a limited study of the effects of  $\ell > 1$  on open systems with nonuniform elongation rates: (a) systems with a single slow site, (b) bounds on the current in general open systems, and (c) as an illustration of systems with full disorder, simulation results for a real mRNA sequence.

#### A. Simple model of a single internal blockage

We begin with the simplest form of quenched disorder, namely, having a single *internal* site  $i$  with a reduced elongation rate:  $r < 1$ . If this blockage were moved to the termination site, then  $r$  would carry the label  $\beta$  to conform with the notation above. In most of the simulation studies for this section, we focus on  $\alpha = \beta = 1$  with the blockage at the center of the mRNA (i.e.,  $i = 100$  with  $N = 200$ ). Though the naive expectation is that the blockage should limit the current in the same way, regardless of its location, we observe that the current is noticeably reduced if the slow site is internal. This reduction can be understood as follows. Once a particle moves past the slow site, the particle behind it is not necessarily free to move to the slow site. Instead, because of the random sequential updating, the particle behind has a nonzero chance of being blocked by the particle ahead. With just two particles in the system, this effect can be accounted for exactly. Beyond the scope of this paper, the details will be published elsewhere. Here, let us present some simple heuristic arguments that lead us to results that agree well with simulation data.

Extrapolating from the current in the uniform system [Eq. (16)], we speculate that

$$J = \rho_{source} \frac{P_{free}}{T}, \quad (27)$$

where  $\rho_{source}$  is the average coverage density behind the slow site and  $P_{free}$  is the approximate probability for a particle just beyond the slow site to advance. Lastly,  $T$  is the average time to travel through the slow site and the  $\ell - 1$  sites that precede it. For example, in the termination-limited case,  $\rho_{source} = 1 - \beta$ ,  $T = \ell - 1 + 1/\beta$ , and  $P_{free} = 1$  (no steric hindrance beyond the termination site), giving  $J = \beta(1 - \beta)/(1 + \beta(\ell - 1))$ . We estimate  $\rho_{source}$  from the bulk density induced if the lattice were truncated immediately after the slow site.  $P_{free}$  can be estimated from the density-dependent head spacing in a closed system [Eq. (8)], using the bulk density that would be induced if the lattice began with the slow site.

To determine  $T$  approximately, we consider the behavior of two particles in an infinite system with a single slow site (located at the origin). Denoting the position of the leading and following particles by  $\zeta$  and  $\eta$ , respectively, the evolution of this system can be regarded as a random walk in the

TABLE I. Actual and predicted reductions in current due to an internal slow site with rate  $r$ . Currents with the slow site at one end are listed for comparison. All currents are multiplied by 100. Simulations were performed with  $N=200$ ,  $\alpha=\beta=1$  with the blockage at the center. These results are statistically the same if the blockage is placed at site 20 or 180.

$r$	$T_{\max}$	$J$ for slow site at end	Actual $J$	Predicted $J$
0.01	13.5	0.89	0.86	0.87
0.1	12.9	4.29	3.87	3.90
0.2	12.3	5.00	4.68	4.57

$\zeta$ - $\eta$  plane, confined to  $\zeta - \eta \geq \ell$ . During each time step, the walker moves, with equal probability, either “upwards” ( $\eta \rightarrow \eta + 1$ ) or “to the right” ( $\zeta \rightarrow \zeta + 1$ ). When it arrives at the  $\zeta = \eta + \ell$  line, the walker remains stationary with probability  $\frac{1}{2}$  (and moves “to the right” otherwise). Assuming that there are  $m + 1$  holes between the particles immediately after the leading one leaves the slow site, so that the walker is “initially” located at  $(\zeta, \eta) = (1, -\ell - m)$ , we computed numerically  $\tau_m$ , the average time for the walker to arrive at the  $\eta = 0$  line (i.e., for the second particle to reach the slow site). Note that these  $\tau$ 's take into account the steric hindrance due to the leading particle, which we assume is always free to move. To extract  $T$ , we make the assumption that the left particle advances its first  $m + 1$  steps in time  $m + 1$ , leaving  $\ell - 1$  more steps to reach the slow site. Now, the probability  $p(m)$  for finding a gap of  $m$  holes before the leading particle passes the slow site can be estimated by inserting the bulk density before the slow site ( $\rho_{\text{source}}$ ) in Eq. (8). Taking the average over these initial starting positions and accounting for the time to move over the slow site ( $1/r$ ), we obtain  $T = \sum_m [\tau_m - (m + 1)] p(m) + 1/r$ . Predictions for the current from Eq. (27), compared with the actual current from Monte Carlo simulations, are shown in Table I for several values of the slow rate  $r$ . Despite the approximations involved in this approach, the agreement is surprisingly good.

### B. Bounds for the current

To model a real mRNA, we must allow for arbitrary translation rates associated with each codon. Let us denote the rate at codon  $i$  by  $k_i$ . Due to the excluded volume constraint, however, it is meaningful to consider also  $K_{\ell,i}$ , the maximum rate for a ribosome to translate a stretch of  $\ell$  sites beginning with site  $i$ :

$$K_{\ell,i} \equiv \left( \sum_{q=i}^{i+\ell-1} \frac{1}{k_q} \right)^{-1}.$$

Now, consider a “window” of any stretch of  $\ell$  consecutive sites in the lattice. If one particle is moving within this window, the following particle must wait until the first one passes through entirely before it can begin translating these  $\ell$  sites. Thus the characteristic time associated with the current

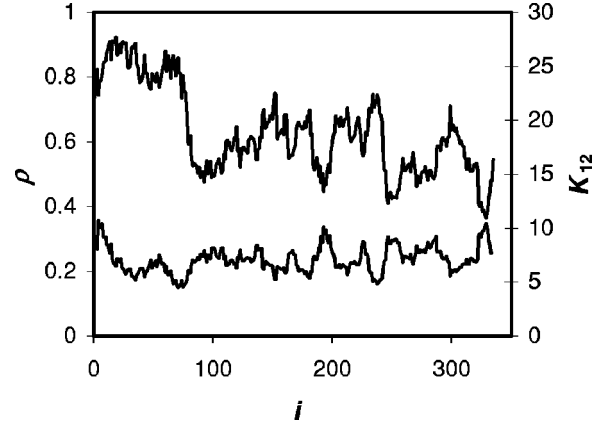


FIG. 12. Steady state coverage density  $\rho$  (upper curve) and maximum translation rate  $K_{12}$  in each window of  $\ell = 12$  codons (lower curve) for the *ompA* gene of *E. coli* when the elongation rates are limiting. Elongation rates at each codon were assumed proportional to availabilities of corresponding tRNA.

must be at least the time required to translate the slowest stretch of  $\ell$  codons. In this way, we find an upper bound to the current, i.e.,

$$J \leq \min_{i \in \{1, \dots, N-\ell+1\}} K_{\ell,i}. \quad (28)$$

One might imagine the slowest segment of  $\ell$  codons acting as a “gatekeeper” and preventing the current from exceeding the value in Eq. (28).

To arrive at a lower bound, we need only replace the elongation rates at each site by the slowest elongation rate. From Eq. (16) above, we have the current of a system with uniform rate unity. Thus, the minimum current for the disordered system is simply

$$J \geq \left( \min_{i \in \{1, \dots, N-1\}} k_i \right) \bar{J}(\alpha, \beta), \quad (29)$$

where  $\alpha$  is the ratio of the initiation rate to the slowest elongation rate, and similarly for  $\beta$ . Though the gap between these bounds for a real system may be too large to be of significant predictive value, they can provide some guide to our understanding of the current.

### C. Simulation of a real gene sequence

To illustrate the full problem of disorder, we have simulated translation of several real mRNA sequences from *Escherichia coli* strain MG1655, obtained from Ref. [47]. Elongation rates at each codon were estimated using commonly accepted values for the availability of tRNA in *E. coli* [45]. The rate at each codon was assumed proportional (with an arbitrary proportionality constant) to the availability of the tRNA decoding that codon, as in Ref. [48]. Corresponding data were not available for estimating initiation and termination rates, so a range of initiation and termination rates was studied. We assumed that ribosomes cover  $\ell = 12$  codons [5,6].

Figure 12 shows the steady state coverage density profile for the reasonably well studied gene *ompA* when the elonga-

tion rates are limiting. The lower curve in the figure shows the maximum translation rate  $K_{12}$  for each window of  $\ell = 12$  codons, calculated from

$$K_{12,i} = \left( \sum_{q=i}^{i+11} \frac{1}{k_q} \right)^{-1}$$

for each site  $i$ . The minimum of the  $K_{12}$  values is 4.51 in arbitrary units, thus giving an upper bound for the current [Eq. (28)]. This value is significantly higher than the actual current of 3.52. [For comparison, Eq. (29) gives a lower bound of 0.99 for the current in this system.] It should be noted that the minimum  $K_{12}$  occurs at codon 71, which is approximately the location behind which the ribosome density is very high, due to ribosomes “piled up” behind the slow region. In general, lower values for the rate  $K_{12}$  correspond to higher ribosome densities, and higher  $K_{12}$  to lower ribosome densities, leading to an approximate symmetry between  $K_{12}$  and  $\rho$ . Thus the  $K_{12}$  values are useful in understanding the ribosome density profiles observed.

### CONCLUSIONS

This work generalizes the well-studied  $\ell=1$  TASEP model to particles with extended sizes. Since there is a difference between particle density  $\rho_r$  and density of occupied sites  $\rho$ , the familiar particle-hole symmetry ( $\rho \leftrightarrow 1 - \rho$ ) takes the form  $\rho_r \leftrightarrow 1 - \rho$  here. Exact results for TASEP on a uniform ring, including probability distributions for the current and for particle headway, were found. Particularly useful in the latter part of this study is the new current-density relationship, which is interpreted as a density dependent mobility factor (or “diffusion constant”).

An extremal principle [38] based on domain wall theory allowed the closed system current-density relation to predict currents and bulk densities in the uniform open system as functions of the initiation and termination rates  $\alpha$  and  $\beta$ . The phase diagram for the open system was thus determined using the extremal principle. Domain wall theory also provided an explanation for the linear density profiles and other unique characteristics observed at the first order transition between the high and low density phases. Given the ability of this theory to describe particles with length  $\ell > 1$ , it might be exploited further to determine fluctuations in number of bound particles in the steady state and behavior in the pre-steady state regime, as has been done for the  $\ell=1$  system [49].

Based on the new mobility factor, a simple continuum limit led to a differential equation for the density profile in open systems. Though nonlinear in general, this equation can be transformed, in the case of a *symmetric* exclusion process, to the familiar *linear* diffusion equation for an effective particle density  $\chi$ . As a result, the stationary density profile, expressed in terms of  $\chi$ , is again linear. For the *asymmetric* exclusion process, the differential equation is more complex, due to an extra term to account for the drive. Analytic expressions, albeit implicit, for the stationary profiles were obtained and solved numerically through a rapidly converging, iterative procedure. The results matched simulation data

closely. However, the predictive power of this approach is limited, since the “coarse-grained” parameter  $D$  cannot be derived from the microscopic jump rates. Naive estimates for it are of the right order of magnitude but quantitatively inaccurate, reflecting the importance of particle correlations. Regarding  $D$  as a phenomenological quantity, we simply fit it to data. At a more detailed level, open questions about density profiles remain. In particular, our continuum limit cannot explain the particle depletion and period- $\ell$  structure near the termination end of high density systems. Profiles of the particle density  $\rho_r$  are also of interest, though beyond the scope of this simple continuum theory. Absent from the  $\ell=1$  TASEP, peaks with spacing  $\ell$  extend far behind a blockage, with a decay length well beyond typical microscopic scales (data not shown). Evidently, when extended objects are included, even a simple uniform TASEP tantalizes us with a rich variety of behavior.

Finally, systems with quenched disorder in the particle hopping rates were briefly considered. Effects of a single internal slow site on the current were estimated with fair accuracy by considering the average delay a particle near the slow site experiences due to a particle ahead of it. Also, bounds on the current were determined for general disordered systems. A real gene sequence was simulated, leading to a complicated density profile. The parameter  $K_{\ell,i}$ , the maximum rate to translate a stretch of  $\ell$  sites beginning with site  $i$ , proved helpful in understanding the shape of the density profile, but the disordered system remains far from solved.

We close with some speculation about the relevance of this work to an understanding of translation. The advent of functional genomics technologies to measure simultaneously mRNA and protein expression profiles from many thousands of genes provides special opportunities to begin to understand gene expression regulation. Our results for currents (i.e., protein production rates) and ribosome densities for uniform systems cannot be directly compared with such experimental data from typical bacterial cells because translation is not approximated well enough by a uniform system. However, it is possible to use the uniform system results to interpret data from an mRNA artificially constructed to be uniform. Although there are no reports in the literature of systems that are approximately uniform, the use of an *in vitro* translation system [50] provides an opportunity to make such experimental observations.

It is known that the relationship between mRNA and protein levels in typical cells is nonlinear [11,12]. Specifically, when cells are grown under two different sets of conditions, the amount of protein corresponding to a particular gene may be down-regulated while its corresponding mRNA is up-regulated, and the opposite may be true for other genes measured from the same samples. We expect that in biological systems, the initiation rate  $\alpha$  should be an increasing function of the availability of ribosomes within the cell. The protein production rate [current  $J$ , Eq. (16)] is, in turn, a nondecreasing function of  $\alpha$ . This analysis thus suggests that the observed nonlinear relationship can arise from changes in the availability of ribosomes given the nonlinear relationship between  $J$  and  $\alpha$ . However, this situation would cause all pro-

tein production rates to change in the same direction, that in which the ribosome availability changes. It would not permit some proteins to be up-regulated while others are down-regulated—which is the situation observed experimentally. Further nonlinearity would arise if mRNA's were to compete for available aa-tRNA as well as for ribosomes. Detailed modeling of this effect will require a better understanding of systems with quenched disorder, in which the elongation rates result from aa-tRNA availability.

### ACKNOWLEDGMENTS

We thank Johannes Hager, James Sethna, Beate Schmittmann, Vassily Hatzimanikatis, Amit Mehra, Tom Chou, and Anatoly Kolomeisky for their helpful suggestions. R.K.P.Z. acknowledges support from the National Science Foundation through Grant Nos. DMR-9727574 and 0088451. K.H.L. acknowledges support from the NSF through Grant Nos. BES-0120315 and 9874938. L.B.S. thanks the NSF and the Corning Foundation for financial support. This research was conducted using the resources of the Cornell Theory Center, which receives funding from Cornell University, New York State, federal agencies, foundations, and corporate partners.

### APPENDIX: DERIVATION OF MODIFIED DIFFUSION EQUATION FROM DISCRETE VERSION

A modified diffusion equation qualitatively equivalent to Eq. (19) can be obtained by taking the naive continuum limit of the discrete mean field equations of MacDonald *et al.* We begin with Eq. (7) of [7] for the current from lattice site  $j$  to  $j+1$ :

$$q_j = \frac{n_j^{(L)} n_{j+1}^{(0)}}{n_{j+1}^{(0)} + n_{j+L}^{(L)}}$$

and make the following correspondences with our continuum notation:

$$n_j^{(0)} \rightarrow \rho_h(x),$$

$$n_j^{(L)} \rightarrow \rho_r(x),$$

$$L \rightarrow \ell,$$

$$q \rightarrow J,$$

where  $x$  lies in  $[0, N]$  here. Thus we have

$$J(j \rightarrow j+1) = \frac{\rho_r(x) \rho_h(x+1)}{\rho_h(x+1) + \rho_r(x+\ell)}. \quad (\text{A1})$$

Performing a series expansion of Eq. (A1) and keeping terms up to second order, we find that

$$\begin{aligned} J(j \rightarrow j+1) &= \frac{\rho_r(x) \left[ \rho_h(x) + (1) \frac{\partial \rho_h(x)}{\partial x} + \frac{1}{2!} \frac{\partial^2 \rho_h(x)}{\partial x^2} + \dots \right]}{\left[ \rho_h(x) + (1) \frac{\partial \rho_h(x)}{\partial x} + \frac{1}{2!} \frac{\partial^2 \rho_h(x)}{\partial x^2} + \dots \right] + \left[ \rho_r(x) + \ell \frac{\partial \rho_r(x)}{\partial x} + \frac{\ell^2}{2!} \frac{\partial^2 \rho_r(x)}{\partial x^2} + \dots \right]} \\ &= \frac{\rho_r \rho_h}{\rho_s} + \frac{\rho_r \rho_h'}{\rho_s} + \frac{1}{2} \frac{\rho_r}{\rho_s^2} [\rho_r \rho_h'' - \ell^2 \rho_h \rho_r''] \end{aligned}$$

by using  $\rho_h(x) = 1 - \ell \rho_r(x)$ . Similarly,

$$\begin{aligned} J(j-1 \rightarrow j) &= \frac{\rho_r \rho_h}{\rho_s} - \frac{\rho_r' \rho_h}{\rho_s} + \frac{\rho_r \rho_h}{\rho_s^2} \rho_s' + \frac{\rho_r \rho_h}{2 \rho_s} \left[ 2 \left( \frac{\rho_s'}{\rho_s} \right)^2 \right. \\ &\quad \left. - 2 \frac{\rho_r'}{\rho_r} \frac{\rho_s'}{\rho_s} + \frac{\rho_r''}{\rho_r} - (\ell-1)^2 \frac{\rho_r''}{\rho_s} \right]. \end{aligned}$$

It can then be shown that

$$\begin{aligned} \frac{\partial J}{\partial x} &= J(j \rightarrow j+1) - J(j-1 \rightarrow j) \\ &= \frac{\partial}{\partial x} \left[ \frac{-1}{2 \rho_s^2} [\ell(1-\ell) \rho_r^2 + 1] \frac{\partial \rho_r}{\partial x} \right] + \frac{\partial}{\partial x} \left( \frac{\rho_r \rho_h}{\rho_s} \right). \end{aligned}$$

This leads to the modified diffusion equation

$$J = D \left( \frac{1 + \ell(1-\ell) \rho_r^2}{\rho_s^2} \right) \left( - \frac{\partial \rho_r}{\partial x} \right) + E \frac{\rho_r \rho_h}{\rho_s}, \quad (\text{A2})$$

where we have defined  $D = 1/2$  and  $E = 1$ .

Finally, we rewrite Eq. (A2) in terms of the effective particle density  $\chi$ . Thus we find that the mean field equations of Ref. [7] predict

$$\frac{\partial \chi}{\partial x} = - \frac{1 + (\ell-1) \chi}{1 + (\ell-1) \chi - (\ell-1) \chi^2} [\tilde{J} \chi + J - \chi(1-\chi)]. \quad (\text{A3})$$

We expect Eq. (A3) to be comparable to Eq. (19), the steady state density profile equation derived previously from simple arguments. Indeed, both equations give the same fixed points for  $\chi$ , thus producing qualitatively identical density profiles. Further, their quantitative differences have little effect on the shape of the density profile (data not shown).



- [1] L. Stryer, *Biochemistry*, 4th ed. (W.H. Freeman, New York, 1995).
- [2] F. Neidhardt and H. Umbarger, in *Escherichia coli and Salmonella*, 2nd ed. edited by F. C. Neidhardt (ASM Press, Washington, D.C., 1996).
- [3] B. Derrida, E. Domany, and D. Mukamel, *J. Stat. Phys.* **69**, 667 (1992).
- [4] G. Schütz and E. Domany, *J. Stat. Phys.* **72**, 277 (1993).
- [5] R. Heinrich and T. Rapoport, *J. Theor. Biol.* **86**, 279 (1980).
- [6] C. Kang and C. Cantor, *J. Mol. Struct.* **181**, 241 (1985).
- [7] C. MacDonald, J. Gibbs, and A. Pipkin, *Biopolymers* **6**, 1 (1968).
- [8] C. MacDonald and J. Gibbs, *Biopolymers* **7**, 707 (1969).
- [9] G. Schütz, in *Phase Transitions and Critical Phenomena*, edited by C. Domb and J. L. Lebowitz (Academic Press, San Diego, 2000).
- [10] A. Kolomeisky, G. Schütz, E. Kolomeisky, and J. Straley, *J. Phys. A* **31**, 6911 (1998).
- [11] S. Gygi, Y. Rochon, B. Franza, and R. Aebersold, *Mol. Cell. Biol.* **19**, 1720 (1999).
- [12] T. Ideker, V. Thorsson, J. Ranish, R. Christmas, J. Buhler, J. Eng, R. Bumgarner, D. Goodlett, R. Aebersold, and L. Hood, *Science* **292**, 929 (2001).
- [13] A. Bortz, M. Kalos, and J. Lebowitz, *J. Comput. Phys.* **17**, 10 (1975).
- [14] J. Krug, *Phys. Rev. Lett.* **67**, 1882 (1991).
- [15] B. Derrida, *J. Stat. Phys.* **31**, 433 (1983). For generalizations and more recent studies, see, e.g., A. Kolomeisky and M. Fisher, *J. Chem. Phys.* **113**, 10867 (2000), and references therein.
- [16] S. Janowsky and J. Lebowitz, *Phys. Rev. A* **45**, 618 (1992).
- [17] S. Janowsky and J. Lebowitz, *J. Stat. Phys.* **77**, 35 (1994).
- [18] T. Seppäläinen, *J. Stat. Phys.* **102**, 69 (2001).
- [19] A. Kolomeisky, *J. Phys. A* **31**, 1153 (1998).
- [20] G. Tripathy and M. Barma, *Phys. Rev. E* **58**, 1911 (1998).
- [21] M. Bengrine, A. Benyoussef, H. Ez-Zahraouy, J. Krug, M. Loulidi, and F. Mhirech, *J. Phys. A* **32**, 2527 (1999).
- [22] J. Krug, *Braz. J. Phys.* **30**, 97 (2000); e-print cond-mat/9912411.
- [23] V. Becker and H. Janssen, *J. Stat. Phys.* **96**, 817 (1999).
- [24] K. Lauritsen and H. Fogedby, *Phys. Rev. E* **47**, 1563 (1993).
- [25] See, e.g., B. Derrida, *Phys. Rep.* **301**, 65 (1998), and references therein.
- [26] T. Sasamoto and M. Wadati, *J. Phys. A* **31**, 6057 (1998).
- [27] F. Alcaraz and R. Bariev, *Phys. Rev. E* **60**, 79 (1999).
- [28] A. Ferreira and F. Alcaraz, *Phys. Rev. E* **65**, 052102 (2002).
- [29] G. Lakatos and T. Chou, *J. Phys. A* **36**, 2027 (2003).
- [30] D. Mukamel, in *Soft and Fragile Matter: Nonequilibrium Dynamics, Metastability and Flow*, edited by M. Evans and M. Cates (Scottish Universities Summer School in Physics, St. Andrews, 2000); e-print cond-mat/0003424.
- [31] F. Spitzer, *Adv. Math.* **5**, 246 (1970).
- [32] Recently, authors rediscovering this result named it the “pairwise balance” condition [33,34]. In our case, for every configuration  $C$  that makes a transition to  $C'$ , the inverse is forbidden, i.e.,  $R_{C \rightarrow C'} = 1; R_{C' \rightarrow C} = 0$ . Nevertheless, another  $C''$  can be identified with  $C'$  such that  $R_{C'' \rightarrow C} = 1$  and be uniquely “paired” to account for the cancellation of  $P^*(C)R_{C \rightarrow C'} - P^*(C'')R_{C'' \rightarrow C}$  in the master equation. Specifically, every  $C$  consists of (one or more) solid “trains” of ribosomes. Let  $C'$  be the result of the “leading” ribosome of *one* of these trains hopping forwards. Then  $C''$  differs from  $C$  by the “trailing” ribosome being a single site behind the end of *this* train.
- [33] P.-M. Binder, M. Paczuski, and M. Barma, *Phys. Rev. E* **49**, 1174 (1994).
- [34] G.M. Schutz, R. Ramaswamy, and M. Barma, *J. Phys. A* **29**, 837 (1996).
- [35] If a “coarse grained cell” of  $n$  sites contains  $n_r$  ribosomes and  $n_h$  holes, then the total number of possible arrangements is  $(n_r + n_h)! / n_r! n_h!$ . Writing  $n_r = n \rho_r$ , etc., and taking the logarithm leads to the expression here.
- [36] B. Derrida, J. Lebowitz, and E. Speer, *J. Stat. Phys.* **107**, 599 (2002); See also L. Bertini, A. De Sole, D. Gabrielli, G. Jona-Lasinio, and C. Landim, *Phys. Rev. Lett.* **87**, 040601 (2001); L. Bertini, A. De Sole, D. Gabrielli, G. Jona-Lasinio, and C. Landim, *J. Stat. Phys.* **107**, 635 (2002).
- [37] Note that it is always possible to achieve a linear profile (in these one-dimensional systems) by an appropriate transformation of the density variable, as long as the effective “diffusion constant” is a single valued function of  $\rho^*$ .
- [38] V. Popkov and G. Schütz, *Europhys. Lett.* **48**, 257 (1999).
- [39] J. Hager, J. Krug, V. Popkov, and G. Schütz, *Phys. Rev. E* **63**, 056110 (2001).
- [40] J. Hager, *Phys. Rev. E* **63**, 067103 (2001).
- [41] In this section, we will use only the  $N \rightarrow \infty$  limit of Eq. (5).
- [42] B. Derrida, J. Lebowitz, and E. Speer, e-print cond-mat/0205353.
- [43] B. Derrida, J. Lebowitz, and E. Speer, *Phys. Rev. Lett.* **89**, 030601 (2002).
- [44] D. Boal, B. Schmittmann, and R. Zia, *Phys. Rev. A* **43**, 5214 (1991).
- [45] J. Solomovici, T. Lesnik, and C. Reiss, *J. Theor. Biol.* **185**, 511 (1997).
- [46] J. Bergmann and H. Lodish, *J. Biol. Chem.* **254**, 11927 (1979).
- [47] *E. coli* Genome Project at the University of Wisconsin–Madison, <http://www.genome.wisc.edu/sequencing/k12.htm>
- [48] T. Lesnik, J. Solomovici, A. Deana, R. Erhlich, and C. Reiss, *J. Theor. Biol.* **202**, 175 (2000).
- [49] L. Santen and C. Appert, *J. Stat. Phys.* **106**, 187 (2002).
- [50] Y. Cenatiempo, *Biochimie* **68**, 505 (1986).