Macroscopic model of self-propelled bacteria swarming with regular reversals

Richard Gejji,[1,2] Pavel M. Lushnikov,[3] and Mark Alber[1]

[1]*Department of Applied and Computational Mathematics and Statistics, University of Notre Dame, Notre Dame, Indiana 46656, USA*

[2]*Mathematical Biosciences Institute, Ohio State University, 1735 Neil Avenue, Columbus, Ohio 43210, USA*

[3]*Department of Mathematics and Statistics, University of New Mexico, Albuquerque, New Mexico 87131, USA*

(Received 6 January 2011; revised manuscript received 30 December 2011; published 6 February 2012)

Periodic reversals in the direction of motion in systems of self-propelled rod-shaped bacteria enable them to effectively resolve traffic jams formed during swarming and maximize the swarming rate of the colony. In this paper, a connection is established between a microscopic one-dimensional cell-based stochastic model of reversing nonoverlapping bacteria and a macroscopic nonlinear diffusion equation describing the dynamics of cellular density. Boltzmann-Matano analysis is used to determine the nonlinear diffusion equation corresponding to the specific reversal frequency. Stochastic dynamics averaged over an ensemble is shown to be in very good agreement with the numerical solutions of this nonlinear diffusion equation. Critical density p_0 is obtained such that nonlinear diffusion is dominated by the collisions between cells for the densities $p > p_0$. An analytical approximation of the pairwise collision time and semianalytical fit for the total jam time per reversal period are also obtained. It is shown that cell populations with high reversal frequencies are able to spread out effectively at high densities. If the cells rarely reverse, then they are able to spread out at lower densities but are less efficient at spreading out at higher densities.

DOI: 10.1103/PhysRevE.85.021903 PACS number(s): 87.18.Ed, 05.40.−a, 05.65.+b, 87.18.Hf

I. INTRODUCTION

Many bacteria, including species found in diverse soil and water environments, are able to spread rapidly over surfaces by the process of swarming, which is the collective motion of many thousands of cells. Bacteria capable of swarming range from innocuous carbon-cycle organisms to harmful pathogens. Swarming involves directional movement due to pulling with type IV pili or either propulsion due to rotating flagella or slime jets [1]. In certain cases, these mechanisms work together and allow cells to swarm at a rate faster than each individual type of motility [2,3].

For example, *Myxococcus xanthus*, ubiquitous bacteria found in soil, are very efficient swarmers. These bacteria have elongated rod-type shapes (about 7 μm in length and 0.5 μm in width) and they move by gliding over a substrate in the direction of their longer axis [1,2,4,5]. They align and travel together in the same direction [see Fig. 1(a)] as well as reverse direction of their motion about every 8 min [1,5,6]. Mutant species of *M. xanthus* that are unable to reverse are also unable to swarm [5,7].

After *M. xanthus* are inoculated in the center of agar plate, they start growing and moving, and the swarm expands. Ninety percent of the expansion is caused by cell movement and only 10% by growth [3]. It has been shown that a reversal period of 8.8 min maximizes the expansion rate for a given average cell velocity of 4 μm/min [5]. Such motion is limited by new cells moving out from the center. Therefore, a cell, in many cases, cannot move all 8 min in the direction toward the center. When encountering a cell moving in the opposite direction, the cell stops and waits until it is time to start moving again away from the center. The swarm expands symmetrically in all directions [see Fig. 1(b)]. This symmetry dictates that there is a net movement only in radial directions. In a swarm *M. xanthus* bacteria try to escape from the central region of high cell density to the low density region at the swarm edge and then to an unoccupied area where nutrient and oxygen are abundant [5]. Here we study the role of reversals in this process in relatively high-density domains close to the central region.

Reversals of *M. xanthus* cells require an inversion of cell polarity and coordination of the A and S motility systems (slime production and pili IV motility) achieved by a set of proteins encoded by the frz operon [9–11]. Reversals are needed for cells to reorient themselves as part of a biased random walk resulting in movement of groups of cells during aggregation and fruiting body formation under starvation (see Ref. [11] for a review). It was suggested in Ref. [5] that in the presence of nutrient during swarming of *M. xanthus* the oscillatory cell motion is used for the net migration of cells from the center of the swarm to the swarm edge where nutrient and oxygen are abundant and there is little contest for either.

The two-dimensional (2D) off-lattice microscopic stochastic model (MSM) described in Ref. [2] has been able to predict optimal reversal rates for specific choices of bacterial velocities and aspect ratios leading to the maximal swarming rates of the colony, which were confirmed in experiments [5]. It has been also shown in Ref. [5] that such a choice of the optimal reversal rate allows cells to align better and resolve traffic jams, resulting in the maximal order of alignment. The model takes into account cell shape and direction of motion of each *M. xanthus* in the colony determined by the two motility mechanisms: pili IV and slime production. Recently, the subcellular elements model (SCE) was developed and calibrated to study the role of bending and slime capsule adhesion of cells swarming in rafts [12].

Myxococcus. xanthus behaves differently in the presence of nutrient and under starvation. Under starvation, *M. xanthus* is observed to aggregate and form standing waves of cell density (ripples) close to the edge of the colony. Individual cells are aligned parallel to each other in ripples and move back and forth in one direction, regularly reversing the direction of their movement. Continuous models [13–15] and a lattice gas

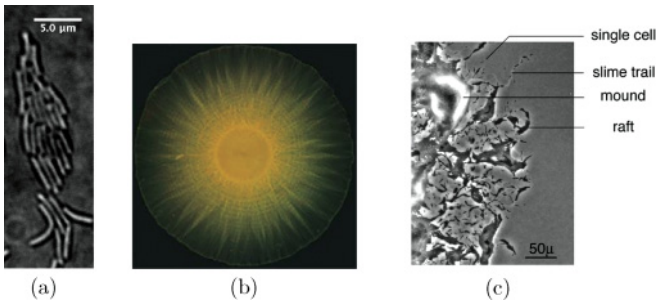

FIG. 1. (Color online) (a) *Myxococcus xanthus* cells aligned together into a cluster (raft) during swarming. (Image was obtained in Dr. Shrout's laboratory by Dr. Alber's student Cameron Harvey.) (b) Swarm of *M. xanthus*, picture taken by Lars Jelsbak. The edge of the swarm displays a single layer of cells that are spreading outward away from the cell center [1]. (c) Distribution of cells at the swarm edge [5,8]. The multicellular structures, slime trails, mounds, and rafts are labeled. The swarm is expanding in the radial direction, which is to the right of the image.

cellular automation model [16] were developed based on the assumption that reversal periods of individual cells depend on C-signaling between cells which are close to each other, and on the existence of a refractory period (period of time after reversal when cells do not react to C-signaling). These models have been used for studying the formation of ripples and the aggregation process. A lattice based models was also used [17,18] for studying cell streaming during the aggregation of *M. xanthus* under starvation. Cells in streams also move mostly in one dimension.

The biological justification for the one-dimensional (1D) discrete and continuous models described in this paper is based on observation of the movement of cells in rafts during the swarming phase [19] [see Figs. 1(a) and 2]. Individual cells also reverse regularly in rafts but they do not exchange C-signal with each other. In this paper, we focus on studying the self-propelled motion of rod-shaped bacteria without specifying motility engines. We do not incorporate cell division, the directional effects of slime, or the social motility governed by pili into the model. Instead we use a basic model for studying bacterial swarming caused by the cell-cell collisions (jams) and regular cell reversals, in rafts propagating near the edge of the swarm [see Fig. 1(c)]. We assume that cells cannot

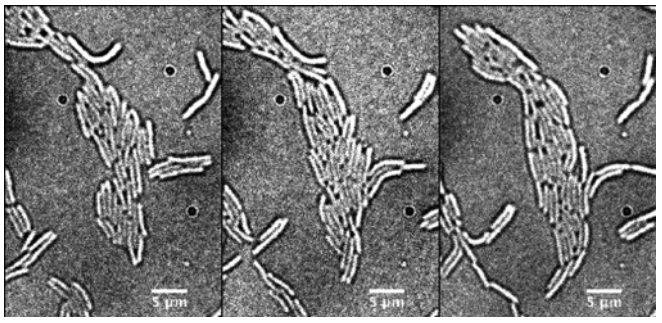

FIG. 2. Sequence of images of a moving raft in a swarm of *M. xanthus* consisting of approximately 100 cells. Cells are aligned with each other. (Images were obtained in Dr. Shrout's laboratory by Dr. Alber's student Cameron Harvey.)

climb onto each other and that no more than one cell could be positioned at any moment in time at specific location in space (by using excluded volume constraint). Cells are modeled as self-propelled rods that glide on slime on the substrate with the same constant velocity. They reverse the direction of their motion periodically in time, which serves as a mechanism for diffusion of spatial positions of cells.

The main result of the paper is an establishment of a connection between one-dimensional (1D) microscopic and macroscopic models describing the swarming of bacteria reversing at different frequencies. The microscopic 1D discrete stochastic model, which is a 1D simplified version of the full 2D model [2], describes the motion of cells in a raft at the edge of a swarm [see Figs. 1(c) and 2]. The 1D excluded volume constraint generally prevents cells from forming rafts. That effect might facilitate transport in the full 2D model [2] but we neglect it in our 1D model. The 1D macroscopic model is a nonlinear diffusion equation describing the cellular density dynamics for self-propelled nonoverlapping rods moving with regular reversals. The combination of a stochastic discrete model and its continuous limit, in the form of a nonlinear diffusion equation, constitutes a multiscale modeling environment. This environment allows one to zoom in and study individual bacteria and then zoom out and investigate the emerging behavior of a large number of bacteria in a swarming colony.

Although only a few continuous models based on biological cell behavior exist that take the volume of cells into account and prevent cells from overlapping, such models are more biologically relevant and can provide novel insights. Recently, continuous limit models describing the dynamics of cellular density were derived from the microscopic motion of randomly moving cells exhibiting volume exclusion in the presence of chemotaxis [20–24]. In particular, Ref. [24] introduces a nonlinear diffusion equation model with a chemotactic term for describing amoeba aggregation and without finite-time collapse of the solution [25]. This is in contrast with the standard but biologically less realistic Keller-Segel equation (sometimes also called Patlak-Keller-Segel equation) with constant diffusion coefficient [26,27] that neglects the size of bacteria resulting in the solution (bacterial density) having a collapse in finite time [28–30].

Another 1D continuous limit equation was recently derived from a model of cells that interact using Hooke's Law [31]. This equation also displays nonlinear fast diffusion and looks similar to the porous medium equation but with a negative exponent. This model agrees well with the discrete system from which it is derived and it is capable of effectively making biological predictions for cells that can be modeled as stiff springs.

The paper is organized as follows. In Sec. II an MSM of cellular dynamics is introduced that describes the 1D motion of self-propelled rods with periodic reversals in the direction of their motion. In Sec. III general settings for MSM simulations are presented and results of multiple 1D dynamics MSM simulations with initially localized distributions of bacterial colonies are described. In Sec. IV elementary laws of collisions (jams) between cells are derived and equilibrium motion of cells is determined in the limit of zero noise of the reversal period. Without interactions (in a vanishing cell density limit),

each cell experiences almost periodic motion in space and time. Without noise in the reversal time that motion would be strictly periodic. However, the experimentally observed [32] small noise in the reversal time results in the random walk of the average position (averaged over time period $2T$) of the center of mass of each cell at time scales above $2T$. Here T is the average reversal time for each cell. For finite cellular densities we introduce different types of collisions (jams) between cells, including a pairwise jam and a cluster jam. We also find the critical density p_0 below which cellular diffusion is dominated by diffusion (random walk) of individual cells while above p_0 diffusion is dominated by the collisions between cells. In Sec. VI multiple collisions and cell clustering for large cellular densities are studied. In Sec. VII a nonlinear diffusion equation of the general form

$$\partial_t p = \partial_x [D(p)\partial_x p] \tag{1}$$

is introduced, where $p(x)$ is a local cell density (measured in units of volume fraction, i.e., the ratio of volume occupied by cells to the total volume of space), x is the spatial coordinate, and $D(p)$ is the nonlinear diffusion coefficient determined by using Boltzmann-Matano (BM) analysis [33] of ensemble averaged MSM simulations of cells moving with different reversal frequencies. Equation (1) gives microscopically averaged dynamics of cellular density versus the microscopic description of the MSM model. We compare the dynamics of cellular density from MSM simulations with the numerical solutions of Eq. (1) for different reversal frequencies and find a very good agreement between these two types of simulations for $p > p_0$. This confirms that the dynamics of cellular density is indeed of a nonlinear diffusion type described by Eq. (1). In Sec. VIII an analytical approximation for pairwise collision time and semianalytical fit for the total jam time per reversal period are described. In Sec. IX, the main results of the paper and future directions are discussed. In Appendix A, BM analysis is reviewed. Appendix B provides additional testing of the accuracy of BM approach for the cellular distributions of finite size.

II. MICROSCOPIC STOCHASTIC MODEL OF BACTERIAL MOTION

In this section we introduce a computational discrete MSM of cellular dynamics describing the motion of self-propelled rods on a 1D lattice with periodic reversals of the direction of their motion.

We simulate a fixed number of cells of length L that move back (left) and forth (right) with velocity v in a domain on the x axis with periodic boundary conditions. We assume that each cell reverses the direction of its motion on average every T min. The reversal period fluctuates with the variance ΔT_0^2 and is sharply peaked near T, i.e., $\Delta T_0/T \ll 1$ in accordance with the experimental data [32]. The positive integer n denotes the nth reversal of the given cell. The Poisson probability distribution is used to define stochastic realizations of the reversal time T_n for the nth reversal

$$f(k) = \frac{\lambda^k e^{-\lambda}}{k!}, \quad k = 0,1,2,\ldots \tag{2}$$

as follows:

$$T_n = k\Delta T_1, \tag{3}$$

where $\lambda = T^2/\Delta T_0^2$ and $\Delta T_1 = \Delta T_0^2/T$. Because the statistical averages for (2) are $\langle f(k)\rangle = \lambda$ and $\langle (f(k) - \langle f(k)\rangle)^2\rangle = \lambda$ we obtain that $\langle T_n\rangle = T$ and $\langle (T_n - \langle T_n\rangle)^2\rangle = \Delta T_0^2$. With that definition the stochastic realizations of T_n can take only discrete values $0, \Delta T_1, 2\Delta T_1, \ldots$. Since $\Delta T_0/T \ll 1$ we conclude that this provides a good approximation of the *M. xanthus* reversals with a continuous set of values for the reversal time.

To quantify the difference between reversal times of neighboring bacteria, a fluctuating reversal phase is assigned to each bacteria as follows. Inside the time intervals $(0,2T)$, $(2T,4T)$, $(4T,6T)$, . . . , each bacteria has an assigned reversal phase, ϕ, between 0 and $2T$ determined by the time when cell reverses from moving to the right to moving to the left (i.e., cell reverses from right-directed motion to the left-directed motion at times $t = \phi, 2T + \phi, 4T + \phi, \ldots$). The phase ϕ of each cell fluctuates at every reversal period because of the fluctuations of T_n, i.e., while the period T is fixed, the value of ϕ constantly fluctuates. These fluctuations follow the Poisson distribution of Eqs. (2) and (3). Cells reverse from the left-directed motion to the right-directed motion near average times $t = \phi, T + \phi, 3T + \phi, \ldots$ [with the reversal time fluctuating around these average times according to Eqs. (2) and (3)]. The initial phase of each cell is chosen randomly. The condition $\Delta T_0/T \ll 1$ ensures that the change of ϕ over T is small. At much larger time scales, $t \gg T$, the phase ϕ undergoes random walk. These random walks of N cells are independent of each other.

Notice that, in dimensionless units, $L = 1$ and $v = 1$. Unless otherwise specified, we choose $T = 8$. That choice is consistent with the typical experimental value of T in dimensionless units $L = v = 1$. Each cell is represented by a cluster of a finite number of lattice sites on a 1D grid. In a typical simulation, each cell includes 10 lattice sites, i.e., distance between neighboring lattice sites is $\triangle x = 0.1$ [see Fig. 3(a)]. We assume time step to be $1/\triangle x$ in dimensionless units to keep the velocity $v = 1$. However, we also ran multiple simulations with smaller values of $\triangle x$ to make sure that increasing the number of lattice sites per cell (but keeping $L = 1$ and decreasing time step to keep $v = 1$) does not significantly change our results. In other words, we choose $\triangle x$ small enough to be in agreement with the continuous limit $\triangle x \to 0$.

Three dimensionless parameters completely determine the continuous limit $\triangle x \to 0$ description of cell dynamics. The first parameter is vT/L, which is the ratio of the average distance traveled by cells between reversals and the cell length. (The value of this parameter is $vT/L = 8$ for the typical value of time period $T = 8$.) The second dimensionless parameter is the local cellular density $p(x)$ measured in units of volume fraction p, i.e., the ratio of volume occupied by cells and the total volume of available space (in 1D, volume is simply length). The third parameter is $\Delta T_0/T$, i.e., relative size of the reversal time fluctuations.

For example, we can choose the velocity, the reversal period, the fluctuation of the reversal period, and the cellular length as $v_{\rm dim} = 10\,\mu$m/min, $T_{\rm dim} = 8$ min, $\Delta T_{0,\rm dim} = 0.9$ min, and $L_{\rm dim} = 10\,\mu$m, respectively, in dimensional units.

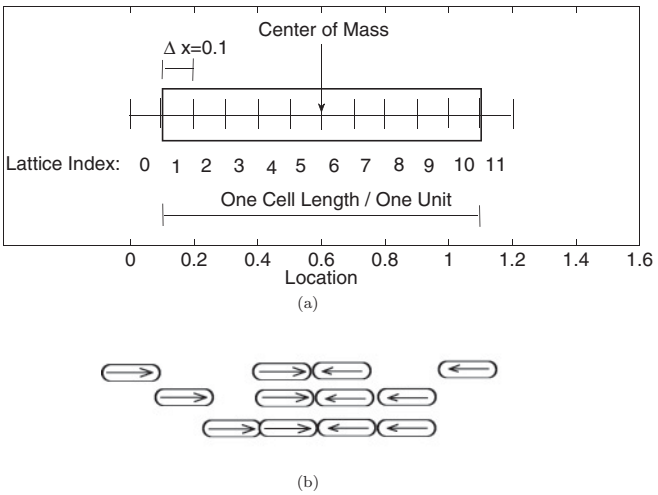

FIG. 3. (a) A single cell is represented in MSM as a cluster of lattice sites (pixels). (b) A sketch of a typical formation of a cluster shown as a sequence at three moments in time (time progress from top to bottom). First, two neighboring cells moving in opposite directions are jammed (upper line). The third cell then approaches to join a cluster (center line) and, subsequently, the fourth cell joins the cluster (bottom line).

This yields $v_{\rm dim}T_{\rm dim}/L_{\rm dim} = 8$ corresponding to the dimensionless values described above. This choice is consistent with the cell length and reversal period used in previous computational models [2,5] and observed in experiments [7]. Unless otherwise specified, we choose below $\Delta T_1 = 0.1$. For $T = 8$ this results in $\Delta T_0 \simeq 0.9$. Experiments with *M. xanthus* typically show only small fluctuations of the reversal period T so the probability distribution function is sharply peaked near the average reversal period T [32]. Our typical choice $\Delta T_0 \simeq 0.9$ reflects these small fluctuations. We also show in Sec. VII B that MSM simulations fail to match solutions of the nonlinear diffusion equation (1) for any form of $D(p)$ in the absence of noise in the reversal period (i.e., for $\Delta T_0 = 0$). In contrast, for a small but finite value of noise ($\Delta T_0 \neq 0$), we observe a very good agreement between MSM simulations and the solutions of Eq. (1) with $D(p)$ determined by BM analysis. This suggests that noise (although small) in the reversal period of bacteria contributes critically to their macroscopic behavior by allowing them to diffuse.

The MSM is a stochastic model. During each time step, a sequence of randomly chosen N cells attempt to move one at a time. MSM determines the movement of each cell based on the occupancy of the next lattice site in the direction of cell motion determined by ϕ at the given time t. If this location is free, the cell is moved one lattice site in that direction (keeping constant length $L = 1$). If the location is not free, the cell does not move, i.e., it is jammed [see Fig. 3(b) for a typical example of jamming]. It is possible that the same cell may move more than once during a time step, and, as a result, some other cells may not move at all. Also, note that the random selection of cells may create small gaps between cells that are following each other side by side.

The creation of such gaps results in extra diffusion that each cell experiences in addition to the directed motion with the speed v. This diffusion is a pure artifact of the finite width Δx of each lattice site and it vanishes as $\Delta x \to 0$. We checked in simulations that the reduction of Δx from 0.1 to 0.001 results in only small changes in the cellular density dynamics (see Sec. VII B for more discussion on this topic). However, the collision time between cells is more sensitive to the value of Δx. Therefore, we used a reduced grid $\Delta x = 0.001$ for the simulations in Sec. VIII. Generally, for all quantities plotted in all figures in the paper we used $\Delta x = 0.1$ unless we explicitly specify a different value.

Unless otherwise specified, the simulations were run in a one-dimensional lattice domain of length 4000 centered at $x = 0$. Initial distribution of cells of width 1000 centered at $x = 0$ of a top-hat shape was used (i.e., the density of cells is approximately constant $p \equiv p_{\rm max}$ for $-500 < x < 500$ and zero everywhere else). [See curve at $t = 0$ in Fig. 4(a) for an example of a top-hat boundary condition.] Because the domain is symmetric between x and $-x$, it replicates a no-flux boundary condition at $x = 0$ after averaging over the statistical ensemble of simulations. In general, we choose lattice domain of a large enough length to avoid influence of the periodic boundary conditions (i.e., to maintain zero cellular densities at both right and left boundaries).

III. MSM SIMULATIONS

Cells near the edge of a bacterial swarm similar to the ones shown in Fig. 1(a), move mostly in radial directions. Therefore, their collective motion can be analyzed by averaging over angles determining their radial motion. In what follows, we also assume in our model that motion along the radius is dominant while rotation is only a correction, which we neglect.

Multiple MSM simulations of 1D dynamics of initially localized distributions of bacterial density have been performed followed by ensemble averaging over initial conditions chosen to represent each desired spatial distribution of the average initial density of cells $p(x,t)|_{t=0}$. The result of such ensemble averaging is the time- and spatial-dependent density $p(x,t)$. (The ensemble serves to approximate averaging over angles of the full 2D problem.) We typically used the "top-hat" initial distribution (constant density around the center of the domain and zero density to the left and to the right of the center) provided by a dense initial packing of bacteria in the domain of width 1000 around $x = 0$. A typical number of stochastic realizations in the statistical ensemble was chosen to be 20 000. The cellular density (measured in units of volume fraction) was determined by calculating the average number of times a given location was occupied [see Fig. 4(a)]. Qualitatively, the cell densities spread out symmetrically away from the center of the top hat.

The cells' movements frequently cause them to collide with each other. When two cells are trying to move into each other's space, they stall (jam) until at least one reverses. This stalling, on average, shifts the mean location of their oscillatory movement away from the location at which they stall. If no other cells are nearby, the cells may collide again or separate further away due to fluctuations in the mean location of their oscillatory movement. If other cells are nearby, these outer cells have their mean location shifted outward while the original cells' mean locations are shifted closer together. As

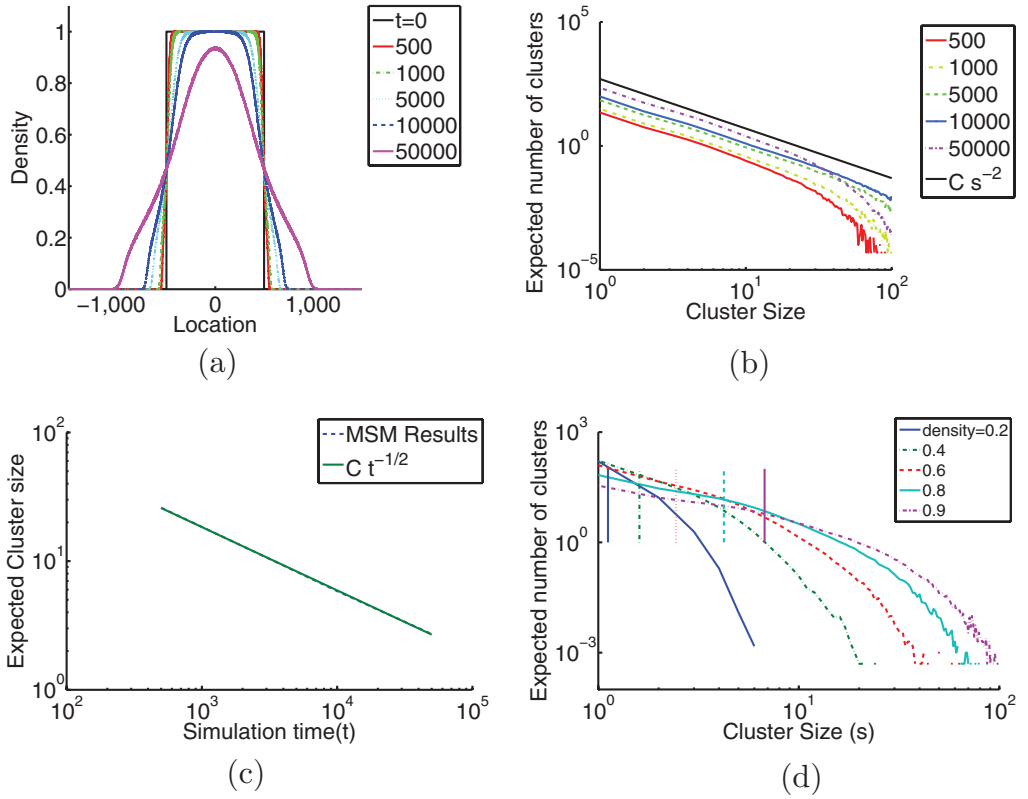

FIG. 4. (Color online) Results of MSM simulations of rods reversing every $T = 8$ performed 20 000 times. [(a)–(c)] Simulations with the top-hat initial conditions and (d) simulations with initial constant average density. (a) Average cell density at different times. (b) Expected cluster frequencies for different times (dashed line), i.e., the average number of clusters of a given size obtained in MSM simulations. The fit to Cs^s law is shown, where C is the fitting parameter and s is the cluster size. (c) Average cluster size over time on a log-log scale (dashed line). The solid line is the fit to the $Ct^{-1/2}$ law, where C is the fitting parameter. These two curves are almost indistinguishable. (d) Expected cluster frequencies at the time $t = 50\,000$ as a function of cluster size s for simulations with constant initial density and the statistical ensemble size 2000. Vertical bars represent the average cluster size for each corresponding curve.

a result of these shifts, cells steadily spread away from each other.

Cells cluster together in highly packed domains. Two cells can jam together, forming a two-cell cluster. If another cell is moving in a direction of a two-cell cluster, then it may join the cluster, forming three-cell cluster, and so on [see Fig. 3(b) for a qualitative picture of such clustering]. To measure the level of clustering, we calculated the frequencies of cluster sizes at different moments of time, which is shown in log-log plots in Fig. 4(b) of MSM simulations with the top-hat initial conditions. The frequency of cluster sizes follows a -2 power law and the average cluster size decreases over time as initially densely packed cells expand. Figure 4(c) shows that this decrease follows a power law decay over time with calculated exponent -0.4965, which is close to the $-1/2$ power law expected from diffusion. This suggests that, early in time, many cells are found in large clusters where they cannot move. At later times, the large clusters rapidly break up into smaller ones, allowing cells to move. Moreover, individual cells near the boundaries spend a significantly greater percentage of their time moving.

Figure 4(a) shows a temporal sequence of the cellular density profiles of the ensemble-averaged MSM simulations demonstrating decrease of the spatial gradients as time progresses. This decrease is slow in comparison with the velocity $v = 1$ of individual cellular motion because during each period $2T$ an average shift of a position of a center of mass of each cell is typically small in comparison with $2Tv$. This suggests that the ensemble averaged distribution of cells at each moment of time and at each point in space is in statistical quasiequilibrium. Below we study dependence of the quasiequilibrium on local cellular density, cellular collision times, and cluster sizes. We also performed a second type of MSM simulations with periodic boundary conditions and uniform average densities to study statistical equilibrium of cellular motion. Figure 4(d) shows the result of such second type of simulations with different average densities p. The expected number of clusters versus the cluster size s (which represents the probability density distribution of different cluster sizes up to the normalization) shown in Fig. 4(d) indicates a fast increase in the number of large clusters with the increase of the average density.

IV. ELEMENTARY LAWS OF COLLISIONS (JAMS) BETWEEN CELLS AND EQUILIBRIUM MOTION OF CELLS IN THE LIMIT OF ZERO NOISE OF THE REVERSAL PERIOD

In the absence of noise in the reversal time T each cell experiences, in a vanishing cell density limit, periodic motion

in space and time. In that case the center of mass of each cell does not moving after averaging over time period $2T$.

However, experimentally observed small noise in the reversal time [32] results in the random walk of the average position of the center of mass of each cell. (By average position we mean here position averaged over period $2T$.) This random walk occurs because each cell moves to the right and to the left for unequal time period determined by the fluctuations of T. Thus, the random walk occurs at time scales above $2T$. The random walk results in collisions of cells for any finite cell density. As the density goes to zero these collisions become more and more rare because it takes more time for cells to span the average distance between them through random walk. Below, in this section, we consider the limiting case with the random walk neglected, i.e., $T = \text{const}$ to extract the major features of such motion. In the next section we reintroduce the finite noise in T to understand how it affects the system dynamics.

There is a finite probability for two neighboring cells to collide (jam) in case of a nonzero cell density. By jam we mean that one cell tries to move where another cell is located, but the excluded volume principle prevents it from moving. The term *jam* in this paper is similar to the term *collision*. The subtle difference is that by *collision* we mean that a cell jams with another cell with subsequent unjamming, i.e., the cell is free to move after a jam.

We distinguish two types of jams in this paper. The first type is a *pairwise jam*. It occurs when two neighboring cells are jammed directly because they try to move in opposite directions toward each other but that motion is prevented by the excluded volume principle. The second type of jam occurs when cell 1 tries to move in the same direction as a neighboring cell 2 but that cell 2 is jammed by another cell(s) (e.g., by cell 3). We refer to this type of a jam of cell 1 as *indirect jam*. Such a jam is an indirect one because there is no direct (pairwise) jam between cells 1 and 2. A typical example occurs when cell 1 moves toward neighboring cell 2 while cells 2 and 3 have a pairwise jam. After cell 1 touches cell 2 they together (cells 1, 2, and 3) form three-cell cluster with pairwise jam between cells 2 and 3 and indirect jam with cell 1. We also say that a given cell is in a "*cluster jam*" if it is either in a pairwise or an indirect jam. The pairwise collision time τ_{pair} is always smaller or equal to T because of reversal of the direction of cellular motion. In contrast, the cluster jam time τ_{cluster} can be arbitrary large if cells stay inside a large cluster. In Sec. VIII, moreover, we use total jam time τ per period T, i.e., the time during which a given cell remains jammed (either directly or indirectly) per period T. With such definition τ never exceeds T.

Assume that the density is small so mostly pairwise jams occur. In such a case the jamming of two cells lasts until one of the cells reverses. After that they move together in the same direction until the second cell reverses. After the second cell reversal, cells move in opposite directions away from each other. Assume that all other cells are still far away. After the first cell reverses for a second time, both cells then will move in the same direction, and after the second reversal of the second cell they will move toward each other. Exact calculation shows that these two cells will never jam again in the absence of other cells. Instead, exactly at the moment when these two cells touch each other, the first cell will reverse for a third time and they will move in the same direction again. This pattern of periodic motion without jamming of these two cells will continue for an arbitrarily long time (or until another, third cell, would approach them close enough to jam with one of these two cells). It means that any two isolated cells jam only once and after that both cells experience periodic motion without disturbing one another.

A similar interaction pattern occurs if we consider a system of three or more cells moving in an infinite spatial domain. After several collisions (jams) between these three or more cells, they also end up in the state where they no longer jam and all cells experience periodic motion without touching each other. The center of mass of each cell participating in a jam shifts relative to its average position at a distance $v\tau_{\text{pair}}$ to the left or to the right (depending on which side it has a jam), where τ_{pair} is the collision time. However, after all collisions end, the center of mass of each cell experiences periodic motion and no averaged over the period $2T$ motion is observed. We refer to such state as an equilibrium motion of *M. xanthus*. Note that equilibrium motion differs markedly from the equilibrium distribution (Gibbs distribution) in statistical mechanics [34] because *M. xanthus* are always self-propelled and are not subject to any type of thermal equilibrium. Starting with a finite number of initially densely packed *M. xanthus*, after finite number of collisions and provided *M. xanthus* divisions are neglected, the bacterial colony expands to such size that there will be no more collisions between cells. After that, the average size of the colony remains the same with bacteria moving periodically at equilibrium.

We now calculate the density of *M. xanthus* p_0 at which a transition occurs from motion with collisions to equilibrium motion. First, consider two neighboring cells and assume that they have phases ϕ_1 and ϕ_2, respectively. Generally $-2T \leqslant \phi_1 - \phi_2 \leqslant 2T$, but assuming periodicity over time $2T$, we can always add a multiple of $2T$ to each of the phases, $\tilde{\phi}_j \equiv \phi_j + n_j 2T$, $j = 1,2$ (n_j are integers), to keep the difference of modified phases inside a twice smaller interval: $-T \leqslant \tilde{\phi}_1 - \tilde{\phi}_2 \leqslant T$. For $p = p_0$ cells do not jam but during a part of the time interval $2T$ they move together (attaching to each other) in the same direction until one of them reverses. After that, they move in opposite directions from each other for the time interval $|\tilde{\phi}_1 - \tilde{\phi}_2|$. After that, the second cell reverses and both cells move in the same direction, and so on. The minimum separation between the centers of mass of these two cells is L and the maximum separation is $L + 2v|\tilde{\phi}_1 - \tilde{\phi}_2|$. The distance L_{dist} between the average positions of the centers of mass is equal to $L_{\text{dist}} = L + v|\tilde{\phi}_1 - \tilde{\phi}_2|$.

Now, to calculate the average density p_0 of many cells we average L_{dist} over phase differences $0 \leqslant |\tilde{\phi}_1 - \tilde{\phi}_2| \leqslant T$, resulting in the critical density

$$p_0 \equiv \frac{L}{\langle L_{\text{dist}} \rangle} = \left[T^{-1} \int_0^T (L + v\phi) d\phi \right]^{-1} = \frac{L}{L + vT/2}. \quad (4)$$

For the standard values $v = L = 1$, $T = 8$ it yields that $p_0 = 0.2$. If initially there is a localized distribution of cells with the average density $p > p_0$, then these cells would spread out with collisions until their density reaches $p = p_0$. If initially $p < p_0$, then some redistribution of cellular density

may occur when the average distance between centers of mass of two neighboring cells is $L_{\rm dist} < L + v|\tilde{\phi}_1 - \tilde{\phi}_2|$. Because average density is low, this would result only in a local redistribution of the positions of cells without much change in the macroscopic cellular density. After cells initially spread out no collisions or cellular density transport is observed. This conclusion is supported by observing a dynamic sequence of experimental images of cells moving in a raft, which is presented in Fig. 2.

We conclude that in order to observe transport of a system of self-propelled rods without noise in the reversal period T at long times one needs to incorporate in the model a source of the density gradient. In *M. xanthus* swarms such a source is present due to division of cells in the center of the bacterial colony. Thus, any transport of self-propelled rods without noise in T is a collective phenomena with the threshold density p_0 required for transport.

V. REINTRODUCTION OF THE NOISE OF THE REVERSAL PERIOD IN THE SMALL DENSITY LIMIT

We now add noise in the reversal period T to the analysis described in the previous section. With noise, collisions between cells occur, even for $p < p_0$, because the random walk of the average position of cells causes them to move at arbitrary large distances until they finally collide with other cells. When the density approaches zero, the frequency of collisions also goes to zero. But, if $p \to p_0$ from below, then cells collide typically at each period $2T$ with the collision time $\sim\Delta T_0$ (so for $\Delta T_0 \to 0$ collision time would vanish). Thus, p_0 separates two regimes of collisions. If $p < p_0$, then collisions are rare because of the noise in T while, if $p > p_0$, collisions are frequent. At the transition densities $p \sim p_0$ the contributions of both of these effects are comparable with each other.

Thus, a transport of *M. xanthus* is a mixture of two effects. The first effect is the diffusion of individual cells due to the noise in the reversal period T that dominates for small densities $p < p_0$. The second effect is due to the frequent collisions of cells during each period $2T$ making that effect essentially a collective one. Two regimes markedly distnguish *M. xanthus* from bacteria like *E. coli* or amoeba *Dictyostelium discoideum*, which diffuse as randomly moving Brownian-like particles [22–24] without any periodic motion.

VI. MULTIPLE COLLISIONS AND CELL CLUSTERING FOR LARGE CELLULAR DENSITIES

If the cellular density p is not small ($p > p_0$) so cells typically experience collisions during each period $2T$, then cell motion is more complicated than the one described in Sec. IV based on rare pairwise collisions. In addition, we assume below that there is nonzero noise in T as in Sec. V. Figures 5(a) and 5(d) show pairwise jam time (duration) versus number of collisions that occur between three adjacent cells for the average cellular density $p = 0.95$. In that case, cells occupy 95% of the total volume and each cell can cover up to $vT/L = 8$ cell volumes between two reversals, meaning that it could collide with multiple cells. Figure 5 demonstrates that distribution of pairwise jam time $\tau_{\rm pair}$ is random. This typically occurs closer to the bacterial colony center, where cell flux caused by cell divisions is large. It keeps the system far from the equilibrium motion state as described in Sec. IV. There are at least two situations where such a far from equilibrium state is possible. The first is the high-density gradient case caused by bacterial division (as mentioned above). The second case occurs if no-flux boundary conditions maintain a large density of *M. xanthus* in a domain with fixed volume. In both cases, the rate of bacterial jamming is high and the collision times are randomly distributed.

Another effect that occurs in the case of large densities is the high probability of the formation of clusters consisting of more than two bacteria. As the density of bacteria approaches 1, all bacteria jam in large clusters. Unjamming bacteria from a large cluster might take a lot of time because the leftmost or rightmost bacteria in the cluster needs to move away to provide space for the bacteria in the center of the cluster to move. As a result, many cells stay jammed in a cluster for a long time for large densities. Figure 6 shows that the averaged over ensemble of MSM simulations cluster collision time $\tau_{\rm cluster}$ diverges for $p \to 1$. Values of $\tau_{\rm cluster}$ for different Δx converge fast to the continuous limit $\Delta x \to 0$, e.g., curves for $\Delta x = 0.01$ and $\Delta x = 0.001$ are almost indistinguishable.

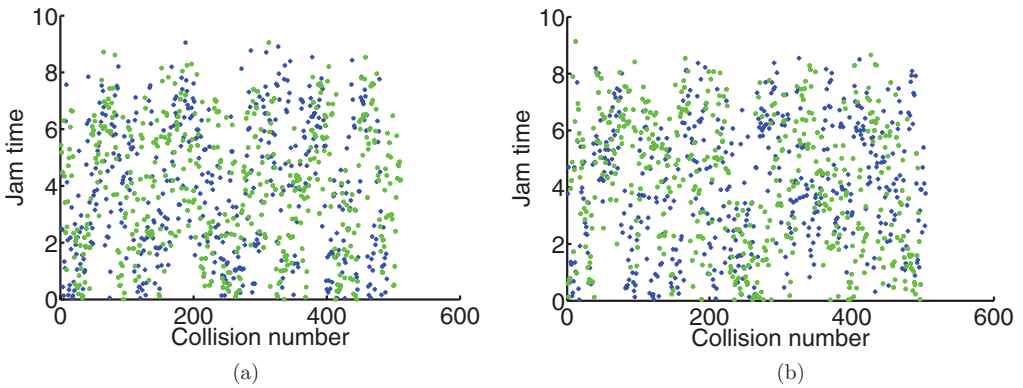

FIG. 5. (Color online) Two stochastic realizations of the pairwise jam time that a single cell experiences with its two adjacent neighbors. Collisions with the left and right neighbors are colored blue and green, respectively, for a total of 1000 different collisions at average cellular density 0.95 and $\Delta x = 0.001$.

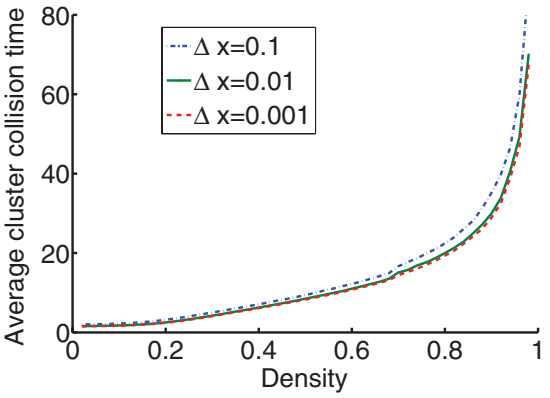

FIG. 6. (Color online) Average cluster collision time τ_{cluster} as a function of density p for different values of Δx. It is seen that $\tau_{\text{cluster}} \to \infty$ for $p \to 1$.

Figure 4(b) shows the distribution of cluster sizes for MSM simulations starting from the top-hat initial condition. Figure 4(d) provides cluster-size distribution for different densities p obtained in MSM simulations with uniform average density and periodic boundary conditions. (That second type of MSM simulations is described in Sec. III.) Figure 4(d) shows that the cluster-size probability density distribution becomes much wider for $p \gtrsim p_0$ indicating multiple collisions.

VII. MACROSCOPIC NONLINEAR DIFFUSION MODEL AND MSM SIMULATIONS

A. Nonlinear diffusion model and its limitations

If collisions are frequent and, additionally, the distribution of the reversal phases and initial position of cells are random, then we assume that collective dynamics of cells is diffusion-like and it is described by the equation of the general type Eq. (1). In this section, we obtain an approximation of the diffusion coefficient $D(p)$ in Eq. (1) to match the results of MSM simulations. This is achieved by running MSM simulations with a top-hat initial distribution and using BM analysis [33] applied to the ensemble-averaged MSM density profile on the right half of the spatial domain at time $t = t_D$. Here and below t_D describes the time at which we apply the BM analysis. (A description of the BM analysis is given in Appendix A). We then demonstrate that numerical solutions of Eq. (1) with $D(p)$ obtained using a BM analysis yield density dependence on space and time that is in very good agreement with the one obtained using MSM simulations. This justifies the initial assumption of collective dynamics of cells being of a diffusion-like type.

Results of MSM simulations unavoidably have noise due to the finite size of the stochastic ensemble used to determine cellular densities. Boltzmann-Matano analysis relies on calculating derivatives of density and we apply a Gaussian filter to the MSM density data to smooth out both p and all its derivatives [35]. We checked that the change of parameters of the Gaussian filter resulted in only small corrections without any systematic error.

Typically, to perform BM analysis we run MSM using an initial top-hat distribution of length 1000 in a domain of size 4000 (see the end of Sec. II for more details). For most of our simulations, the top hat is wide enough so during simulation time the density at the middle of the domain remains close to the initial density $p_{\max}$ (in most simulations $p_{\max} = 1$). This means that cells mostly move near the boundary of initial top-hat distribution while at the middle of the domain the cellular density is almost constant. This allows us to ignore the left half of the domain and treat the cell distribution as if it were stepwise shaped in an infinite domain. This is necessary in order to perform BM analysis that is exact for an infinite spatial interval with stepwise initial conditions only. In Appendix B we study the accuracy of BM analysis for a finite width of the top-hat initial conditions.

Another limitation of the BM analysis is that it requires calculating $[dp(x)/dx]^{-1}$. Due to the presence of the regions where the density is constant, singularities of $[dp(x)/dx]^{-1}$ may be generated during calculation of the non-linear diffusion coefficient near the end of the diffusion curves where p is close to 0 or $p_{\max}$. These artificial singularities result from a loss of numerical precision near singularity of $[dp(x)/dx]^{-1}$ which is clearly seen near $p = 0$ and $p = 1$ in all figures that include $D(p)$. To reduce such a loss of numerical precision, we perform only BM analysis in the neighborhood of the interface that encompasses the initial step of a top hat instead of the entire right half of the domain. It can be also mitigated by performing a cubic spline interpolation of $D(p)$ from the domain $0 < p < 1$ to values around $p = 0$ and $p = 1$. This, however, appears to be unnecessary because the loss of precision does not affect prediction of the density dynamics in Eq. (1) in any significant way.

Since the BM analysis approach for calculating the diffusion coefficient assumes that the nonlinear diffusion equation (1) is solved on an infinite domain, there will be errors in calculating the diffusion coefficient in cases of significantly high density near the boundaries of the finite computation domain. Therefore, if t_D were too large, the analysis would fail. Also, if t_D were too small, then not enough cells would reverse to generate diffusion. We found that any time between $t_D = 125$ and $t_D = 10\,000$ appears to be sufficient for generating reasonably universal diffusion curves for $T = 8$, as shown in Fig. 7(a). In other words, diffusion coefficient $D(p)$ curves generated at different times t_D are close to each other. This near independence of $D(p)$ from t_D justifies our assumption of the collective dynamics of cells being diffusion-like and being described by Eq. (1).

Unless otherwise specified below, we use $t_D = 500$ to generate the diffusion curves. To demonstrate that there is little difference in the numerical solutions of Eq. (1) with diffusion coefficients $D(p)$ chosen based on BM analysis with $t_D = 500$ versus $t_D = 10\,000$, we compared the resulting numerical solutions with the densities obtained using microscopic stochastic model simulations. Figure 7(c) shows that partial differential equation (PDE) density profiles $p(x)$ are almost indistinguishable for both values of t_D. Furthermore, the difference between the numerical solutions of the nonlinear diffusion equation and stochastic simulation results are negligible, except for the region $p < p_0$.

Diffusion curves for different reversal periods T were also calculated [see Fig. 7(b)]. Large reversal periods produce high diffusion at low densities and low diffusion at high densities. Small reversal periods result in low diffusion at low densities

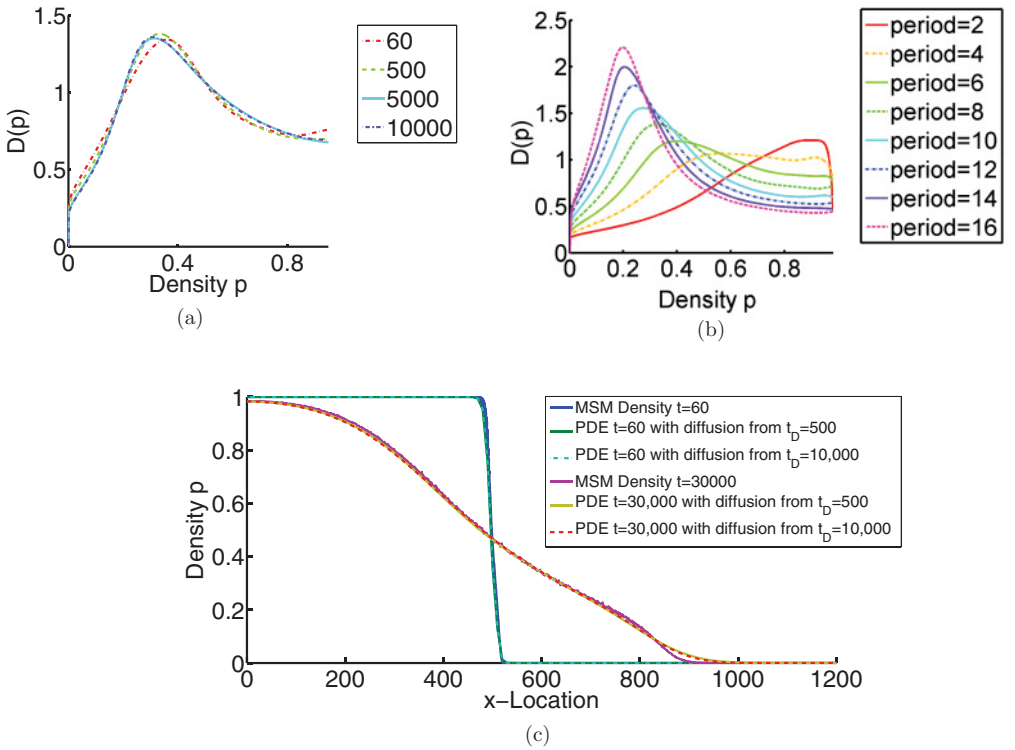

FIG. 7. (Color online) (a) Diffusion coefficient $D(p)$ generated by BM analysis at different times t_D for $T = 8$. (b) $D(p)$ generated by BM analysis for varying reversal periods T for diffusion curves calculated from BM analysis at $t_D = 500$. (c) Density profiles from MSM simulations vs. PDE density profiles [solutions of Eq. (1)] at $t = 60$ and $t = 30\,000$ obtained using diffusion coefficients from (a) with $t_D = 500$ and $t_D = 10\,000$. Only the right half of the spatial domain is shown. It is seen that the agreement between two types of density profiles is very good, except the region of small density, $p < p_0 = 0.2$.

and high diffusion at high densities. In the first case cells move left or right until they collide and they stay jammed for a long time. In the second case, the cells rapidly oscillate left and right. Once cells spread out, collisions become infrequent.

We now discuss the limitations of nonlinear diffusion model. Figure 7(c) demonstrates that there is a difference between MSM and BM prediction for small densities, namely diffusion appears to be suppressed in MSM simulations in comparison with the BM predictions. This difference qualitatively explains the small difference between diffusion curves $D(p)$ generated at different times t_D in Fig. 7(a). The discrepancy between MSM and BM occurs for $p \lesssim p_0$, where the critical density p_0 is defined in Eq. (4). For $p < p_0$ there are not many jams between cells (see Sec. IV) and, subsequently, the diffusion occurs mostly due to the fluctuations of T only. The diffusion coefficient D_0 of this collision-independent limit $p \to 0$ can be estimated from Eqs. (2) and (3) as

$$D_0 = \Delta T_0^2/(2T). \tag{5}$$

For the parameter values used in Fig. 7(a) it yields $D_0 = 0.05$, which is significantly smaller than $D(0)$ in Fig. 7(a). In addition, the finite values of Δx also contribute to D_0, modifying it as $D_0 = (\Delta T_0^2 + T\Delta x)/(2T)$, but this contribution is small for typical values of Δx we use in simulations.

This difference between D_0 and $D(0)$ can be attributed to two effects. The first effect is the loss of the numerical precision of BM analysis for $p \to 0$ as described above in this section. The second effect is due to the lack of the diffusion of MSM at small times, $t \sim T$. To understand the second effect we recall that the diffusion equation for the Brownian motion of particles can be derived from the Newton equation with random force if we neglect the mass, i.e., neglect the inertia [36]. As a result, the speed of propagation of the density in the diffusion equation can be infinite (which is apparent, e.g., if we look at the evolution of the fundamental solution of the linear diffusion equation). In contrast, MSM always describes finite propagation speed because, during period T, each cell can move for a distance no larger than vT, creating effective inertia with the maximum allowed propagation speed $v = 1$. That inertia effect is not important if the density gradients are small. But for the top-hat initial condition, required for BM analysis, the initial gradient is singular. Thus, the diffusion approximation is not applicable for initial evolution of MSM at times $t \sim T$. In addition, the finite size of cells can also contribute to a break of the nonlinear diffusion approximation at these small times. A profound effect of such finite propagation speed occurs for small densities which explains difference between MSM and BM for small densities. To describe these corrections from the initial large-density gradient, one can introduce the inertia-like effect, i.e., through either the introduction of the second time derivative into Eq. (1) (see, e.g., Ref. [37]) or the development of the version of 1D ballistic coalescence model [38]. We, however, do not pursue such corrected models here because they would create a significant effect only at small densities $p < p_0$, which is of limited interest in describing the growth

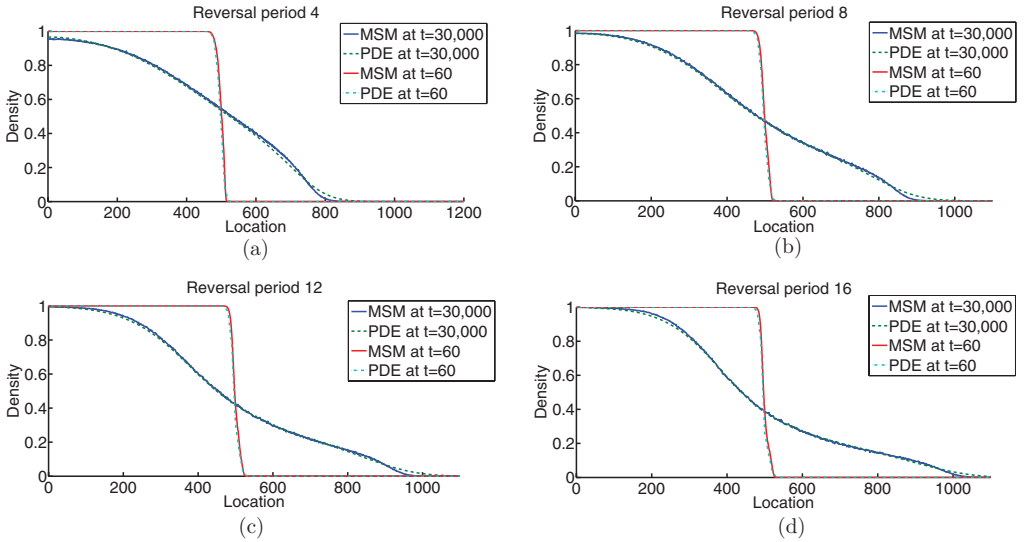

FIG. 8. (Color online) Numerical solution of the diffusion equation (1) [with $D(p)$ from the BM analysis] plotted against the ensemble-averaged MSM results at $t = 60$ and $t = 5000$ for different reversal periods T. PDE and MSM results are very close to each other except over the region of low density $p \lesssim p_0$, where p_0 is given by Eq. (4). Only the right half of the spatial domain is shown.

of a bacterial colony, as explained in the Introduction. For small gradients the diffusion is, of course, applicable for small densities, $p < p_0$, as demonstrated in Sec. VIII through the analysis of the collisions between cells.

B. Testing accuracy of the nonlinear diffusion model and macroscopic limit of MSM

To test that the diffusion curves in Fig. 7(b) actually predict the diffusion in the stochastic model, we compared numerical solutions of the diffusion equation (1) with $D(p)$ derived using BM analysis of MSM simulations with different reversal periods at early and at later times (see Fig. 8). A very good match is demonstrated for $p > p_0$, where the critical density p_0 is defined in Eq. (4).

To test whether the dynamics of the discrete stochastic system is consistently well approximated by the diffusion equation independently of initial density, we compared numerical solutions of the diffusion equation with the MSM simulations for the reversal period $T = 8$ and different initial conditions (different amplitudes of densities in the initial top-hat profile). We, first, generated random initial conditions with constant average density and periodic boundary conditions and allowed cells to move in the MSM simulations for $t = 50\,000$ to reach statistical equilibrium. After that we inserted the obtained equilibrium distribution as a top part of the top-hat initial condition and ran MSM simulations, starting with these spatially nonuniform initial conditions. Figure 9 shows a very good match between these simulations and numerical solutions of the nonlinear diffusion equation. Matching is not as good for smaller densities due to the qualitative change of diffusion and lack of collisions for $p < p_0$, as explained above.

Since cells move on a discrete grid at discrete time steps, we tested convergence of the system to a continuous description of cell movement by decreasing the grid spacing $\triangle x$ and by scaling the lengths and time steps appropriately. The spatial profiles of cell density obtained using ensemble-averaged MSM simulations are shown in Fig. 10. Figure 10 demonstrates that the reduction of $\triangle x$ from 0.1 to 0.001 results only in small changes in the cellular density dynamics with density curves for different $\triangle x$ being practically indistinguishable. This suggests that $\Delta x = 0.1$ already provides a good approximation for the cellular dynamics in a continuous limit.

The diffusion curves $D(p)$ obtained from BM analysis are more sensitive to the change of Δx in comparison with the sensitivity of $p(x)$ shown in Fig. 10. Figure 11 presents diffusion curves from BM analysis obtained at $t_D = 500$ from the same MSM simulations as in Fig. 10. $D(p)$ converges relatively well for $\Delta x \lesssim 0.01$. These changes in $D(p)$ do not undermine the efficiency of the BM analysis of predicting density dynamics. For example, it does not affect good agreement between the MSM and PDE simulations as shown in Fig. 7. For all values of $\Delta x \leqslant 0.1$, corresponding diffusion

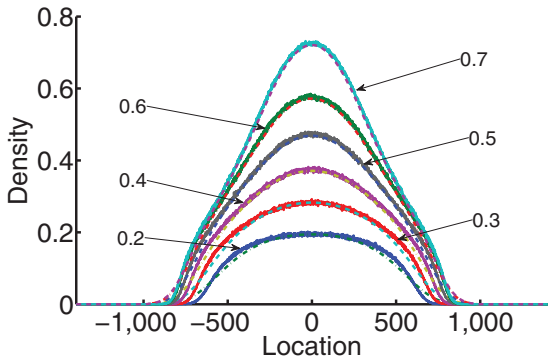

FIG. 9. (Color online) Comparison of ensemble-averaged MSM simulations with numerical solution of Eq. (1) [with $D(p)$ from BM analysis] using different initial densities. MSM densities are not smooth because of the finite size of the statistical ensemble. It is seen that PDE results are well approximations for MSM for $p > p_0$ if we average over fast fluctuations of MSM densities. Simulations results were taken at $t = 50\,000$.

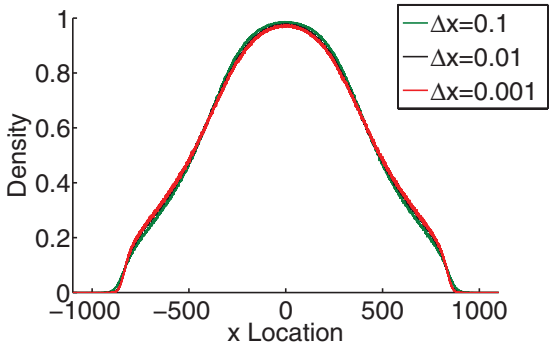

FIG. 10. (Color online) Spatial distribution of density at $t = 30\,000$ from the ensemble-averaged MSM simulations with different Δx. The initial condition had a top-hat distribution form of width 1000.

curves from BM analysis work well for describing density dynamics. The effect of the finite Δx results only in modifying diffusion through additional Δx-dependent fluctuations of the reversal time T. The cause of such modification is discussed in Sec. II.

We now address the effect of different values of ΔT_0. Experimental observations of fluctuations of the reversal period T of individual cells in a *M. xanthus* swarms have shown that they can be described by a probability density function for T sharply peaked near average reversal period T [32]. We study the role of addition of noise to the reversal period by changing the value of ΔT_0. For each ΔT_0 we determine the nonlinear diffusion coefficient $D(p)$ using BM analysis [33] of ensemble-averaged MSM simulations. We found that $D(p)$ changes with ΔT_0 as shown in Fig. 12. Then, we compared the density dynamics description from the MSM simulations with the solution of the nonlinear diffusion equation (1). For finite values of noise (typically for $\Delta T_0/T_0 \gtrsim 0.1$), the agreement between Eq. (1) and MSM simulations is very good [see Fig. 7(c)]. We found that Eq. (1) is not a good approximation of the density dynamics when no noise is added ($\Delta T_0 = 0$), i.e., MSM density no longer follows Eq. (1) with $D(p)$ obtained using BM analysis as shown in Fig. 13. Boltzmann-Matano analysis is exact for any nonlinear diffusion process with smooth $D(p)$. Our failure to fit MSM density for $\Delta T_0 = 0$ proves that in the zero-noise case the dynamics of the system cannot be described by any

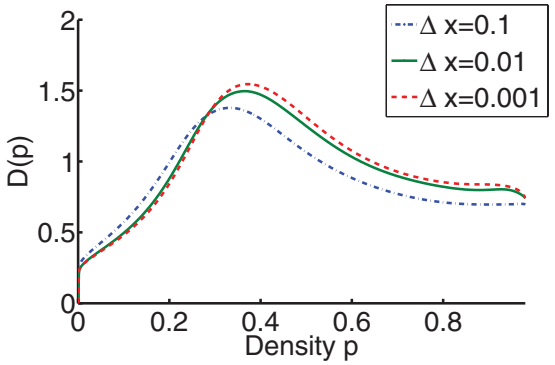

FIG. 11. (Color online) Diffusion coefficient $D(p)$ generated by BM analysis at $t_D = 500$ for the same simulation as in Fig. 10.

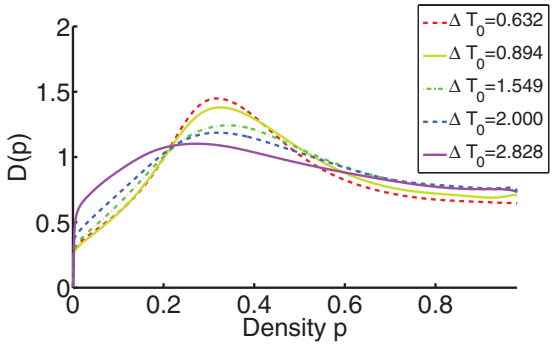

FIG. 12. (Color online) Diffusion curves $D(p)$ for different values of variance ΔT_0 of the reversal time T fluctuations [see Eq. (3) and after it for definitions] obtained from BM analysis of ensemble-averaged MSM simulations.

nonlinear diffusion process. We conclude that finite noise is necessary for the nonlinear diffusion approximation to work. Moreover, finite value of Δx creates effective noise in T as discussed above, meaning that effects of finite ΔT_0 and finite Δx generally add up. To distinguish these effects one can additionally reduce Δx.

Notice that macroscopic models of microscopic phenomena do not necessary follow the solutions of a nonlinear diffusion equation (1). In other words, the diffusion coefficient may depend on the density gradients but not on the density alone or it may be nonlocal in time as discussed in Sec. VII A. This

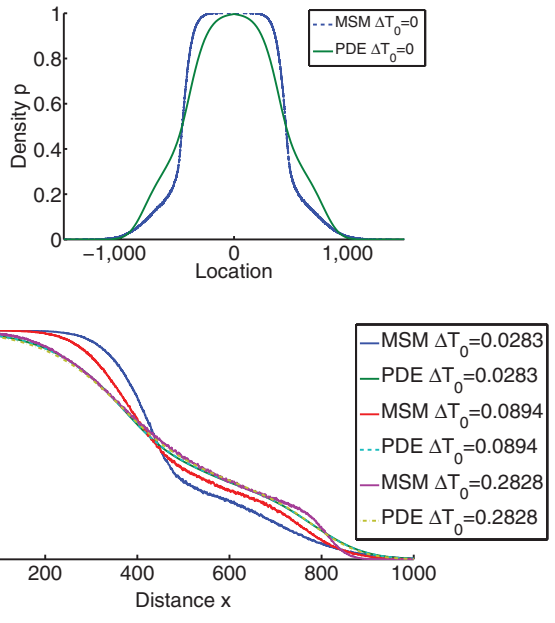

FIG. 13. (Color online) (a) Density curves at time $t = 30\,000$ for the MSM simulations with $\Delta T_0 = 0$ and PDE simulations (obtained using BM analysis with $t_D = 500$). (b) Density curves at $t = 30\,000$ for MSM simulations for a series of values of ΔT_0 and PDE simulations (obtained using BM analysis with $t_D = 500$). Only the right half of the spatial domain is shown. For larger values of $\Delta T_0 = 0.2828$, agreement between the MSM and PDE simulations is much better than that for smaller values of $\Delta T_0 = 0.0283$ and $\Delta T_0 = 0.0894$. Initial conditions in all cases have a form of a top-hat distribution of width 1000.

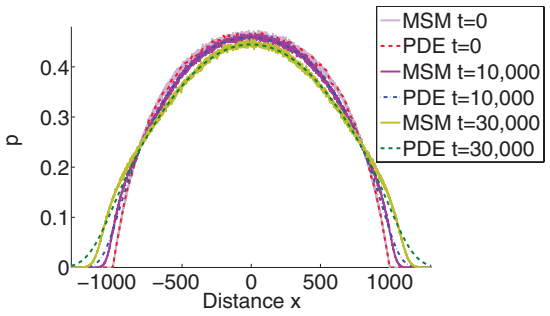

FIG. 14. (Color online) Comparison of ensemble-averaged MSM simulations with numerical solution of Eq. (1) [with $D(p)$ obtained using BM analysis and the top-hat distribution] obtained from the parabolic initial density. Seven hundred cells were used in the MSM simulations in the domain of size 4000 (only a central part of the spatial domain is shown). PDE results provide a very good approximation for the MSM simulation results for $p > p_0$. Curves are plotted for $t = 0$, 10 000, and 50 000. Other parameters are the same as in Fig. 10.

would mean the nonlocality of diffusion. However, extensive comparison of the MSM with BM-based $D(p)$ in this section suggests that it is not the case (assuming $p > p_0$ and finite ΔT_0 as discussed above). [If $D(p)$ were to be obtained using the BM analysis, then at different times t_D this analysis would produce different values of $D(p)$. Moreover, each $D(p)$ from the BM analysis would reproduce MSM density only for fixed $t = t_D$.] We show in this section that all of this is not true. Our results suggest that $D(p)$ obtained from the BM analysis is not a result of an uncontrolled fit but, instead, it reflects fundamental properties of the MSM which, at this time, we cannot derive from the first principles. To strengthen this argument, we also compared the MSM dynamics simulations with the initial distribution in the form of a parabola (not a top-hat initial distribution as above) with the density dynamics obtained using the BM analysis as shown in Fig. 14. Here $D(p)$ was obtained from the BM analysis for the top-hat distribution, i.e., completely independent from the form of the parabola. The results show, again, very good agreement between MSM and the nonlinear diffusion equation.

VIII. ANALYTICAL APPROXIMATIONS OF THE PAIRWISE AND TOTAL COLLISION TIMES

In this section an analytical approximation of the pairwise collision time and semianalytical fit to the total collision times are derived. We mostly focus on a limit of intermediate cellular density when $p > p_0$ but p is not very close to 1, so that a majority of collisions between cells are still pairwise and they do not result in larger clusters. Assume that a cell experiences on average jam(s) of total duration τ_1 from the left and τ_2 from the right during each reversal period $2T$. We include both pairwise jams and indirect jams into the definition of τ (see Sec. IV for detailed definitions of jams). In such a case a shift of the center of mass of a given cell per period $2T$ is $(\tau_1 - \tau_2)v$. The average collision time τ in a given direction (left or right) must be a slow function of x (i.e., $\tau_1 \simeq \tau_2 \equiv \tau$) to avoid large microscopic gradients. Typically τ can be viewed as the average (ensemble or time average) over many collisions (jamming events) for each given cell. It is necessary to stress that τ in this section is the (total) jam time (from both direct and indirect jams) per period T. This quantity differs from τ_{cluster} in Sec. IV because τ never exceeds T by its definition.

Although jam times can fluctuate strongly from collision to collision (as seen in Fig. 5), after averaging over several collisions, τ becomes a slow function of x and t. We also neglect for now the influence of the fluctuations of the reversal time, i.e., we assume that all reversal phases are constant. Below in this section we also separately discuss the effect of these fluctuations. Taking into account the finite value of τ we estimate the local cellular density as $p = L/\langle L_{\text{dist}}\rangle$, where the average distance between neighboring cells is $L_{\text{dist}} = L + v|\tilde{\phi}_1 - \tilde{\phi}_2| - v\tau$ and $\langle\cdots\rangle$ means statistical averaging over the uniformly distributed phases $\tilde{\phi}_1$ and $\tilde{\phi}_2$. This expression is, however, true only for $|\tilde{\phi}_1 - \tilde{\phi}_2| \geqslant \tau$ because distance between centers of mass of two neighboring cells is $\geqslant L$. For pairs of cells with smaller difference in phases $|\tilde{\phi}_1 - \tilde{\phi}_2| \leqslant \tau$ we have to take into account simultaneous collisions (jams) of three and more cells. During each triple collision two neighboring cells have pairwise jams and the third one has an indirect jam. If τ is small, $|\tilde{\phi}_1 - \tilde{\phi}_2| \leqslant \tau$, then cells 1 and 2 move parallel to each other most of the time. They either attach to each other or separate by a typical distance of $2v|\tilde{\phi}_1 - \tilde{\phi}_2|$. After reversing direction, cells always alternate between these two possibilities. A distance between average positions of the centers of mass of these two cells is $\sim v|\tilde{\phi}_1 - \tilde{\phi}_2|$. Almost all the time cells 1 and 2 move together, separated by that average small distance between them. After colliding with another (third) cell on the left (referred to as cell 0) or with a cell on the right (referred to as cell 3) they quickly form a three-cell cluster. Assume that the lifetime of each such cluster is about τ. The pairwise jam time for cells 0 (with cell 1) or cell 3 (with cell 2) then is $\simeq\tau$. For cells 1 and 2 each collision is either a pairwise jam with the jam time $\simeq\tau$ or a cluster jam with the jam time $\simeq\tau$. The average jam time is $\simeq\tau$ in both cases. The distance between average positions of cells 0 and 1 (or between cells 2 and 3) is $\simeq L + v(|\tilde{\phi}_1 - \tilde{\phi}_2| - \tau)$. Based on that we obtain the following approximate expression for the average distance between two neighboring cells combining contributions from $|\tilde{\phi}_1 - \tilde{\phi}_2| \geqslant \tau$ and $|\tilde{\phi}_1 - \tilde{\phi}_2| \leqslant \tau$:

$$\langle L_{\text{dist}}\rangle = T^{-1}\left[\int_\tau^T (L + v[\phi - \tau])d\phi + \int_0^\tau [L + vO(\phi)]d\phi\right] = L + v\left(\frac{T}{2} - \tau\right) + v\tau O\left(\frac{\tau}{T}\right), \quad (6)$$

where we included the contribution of the average distance $\sim v|\tilde{\phi}_1 - \tilde{\phi}_2|$ between cells 1 and 2 for $|\tilde{\phi}_1 - \tilde{\phi}_2| \leqslant \tau$ into the $O(\phi)$ term. Here and below $O(x)$ refers to $O(x) = c_1 x + c_2 x^2 + c_3 x^3 + \ldots$ with constants $c_1, c_2, \ldots$ generally ~ 1.

Terms $\propto v\tau^2/T$, $v\tau^3/T^2$, $v\tau^4/T^3$, etc., in Eq. (6) result from the three-cell, four-cell, five-cell, and so on, cluster contributions, respectively. To establish scaling associated with the number of cells in a cluster we note that the probability to have an n-cell cluster is roughly proportional to the probability P_{n-2} of $n-2$ neighboring cells simultaneously having small differences in phases $|\phi_i - \phi_{i+1}| \lesssim \tau$, $i = 1, 2\ldots, n-2$. Here $P_{n-2} \propto (\tau/T)^{n-2}$ because phases are statistically independent.

An n-cell cluster is formed by these $n-2$ cells together with two surrounding cells involved in pairwise jams with the average time τ. Similarly to the case of the three-cell cluster, the $n-2$ cells inside a cluster have an average jam time τ dominated by the indirect jams. The resulting contribution to the $\langle L_{\rm dist}\rangle$ is $\sim v\tau P_{n-2}$. Of course, for densely packed clusters such an approximation is an oversimplifcation but the general form of $O(x)$ remains the same. These qualitative arguments do not affect the quantitative calculations described below and yield qualitative understanding of the MSM dynamics.

Qualitatively, we can also interpret the formation of a large cluster as a loose analog of phase locking because cells with similar phases have a tendency to form clusters more easily, as explained earlier in this section (they push out other cells less in the process of jamming). We would like to stress that it is purely a kinematic effect because we assume here that fluctuations of the reversal phases of each cell are independent and follow the Poisson distribution Eqs. (2) and (3).

Equation (6) results in the following relation between cellular density and the collision time:

$$p_{\rm an}=\frac{L}{\langle L_{\rm dist}\rangle}=\frac{L}{L+vT/2-v\tau+v\tau O(\tau/T)}. \qquad (7)$$

After solving Eq. (7) for τ, we obtain the following analytical approximation for the average collision time:

$$\tau(p)=\left[\frac{T}{2}-\frac{L}{v\,p}+\frac{L}{v}+\tau O(\tau/T)\right]\Theta(p-p_0), \qquad (8)$$

where p_0 is given by Eq. (4), $\Theta(y)$ is a Heaviside step function [$\Theta(y)=1$ for $y>0$ and $\Theta(y)=0$ for $y<0$], and the factor $\Theta(p-p_0)$ is obtained from the condition that $\tau\geqslant 0$ (recall that it is shown in Sec. IV that jams are absent for $p<p_0$ fluctuations of the reversal time are neglected).

Neglecting $\tau O(\tau/T)$ in Eq. (8) means that we take into account only pairwise jams and neglect indirect jams resulting in the average pairwise collision time $\tau_{\rm pair}$:

$$\tau_{\rm pair}(p)=\left(\frac{T}{2}-\frac{L}{v\,p}+\frac{L}{v}\right)\Theta(p-p_0). \qquad (9)$$

Figure 15(a) compares $\tau_{\rm pair}(p)/T$ simulations that were obtained using MSM with simulations from Eq. (9). MSM simulations were performed with the periodic boundary conditions at the spatial interval of length 1000 and initial random placement of N cells (avoiding configurations forbidden by excluded volume principle). We used $\triangle x=0.05$ and $\triangle x=0.005$. N was chosen for each simulation to match given p (i.e., $N=1000\,p$). All types of collision times were calculated by running simulations through the final simulation time $t_{\rm final}=10^6$. We also assumed ergodicity and recorded collisions of all cells during each simulation. Convergence was tested by comparing the results for a subset of densities to results obtained with $t_{\rm final}=10^7$ and a good match was demonstrated. Ergodicity was also tested by comparing the collision time results from several different stochastic realizations with $t_{\rm final}=10^6$ and a very good match was shown for tested density values.

Figure 15(a) shows that MSM simulations and Eq. (9) are in a reasonably good agreement for $p>p_0$. For $p<p_0$ one would need to modify Eq. (9) to include fluctuations of the reversal time. Comprehensive analysis of such a modification is outside the scope of this paper. Here we consider the particular value of density $p=p_0$ as well as the limit of small densities, $p\ll p_0$.

For $p=p_0$, cells do not collide without fluctuations of T but they often get next to each other and move in parallel, as explained in Sec. IV. This means that with inclusion of fluctuations of T the typical pairwise collision time would be $\sim\Delta T_0$, i.e., $\tau_{\rm pair}(p_0)\sim\Delta T_0$ [in contrast to $\tau_{\rm pair}(p_0)=0$ in Eq. (9)], which is in good agreement with Fig. 15(a) [$\Delta T_0/T=0.09$ for the parameter values of MSM simulations in Fig. 15(a) while $\tau_{\rm pair}(p_0)/T\simeq 0.1$ for the dashed curve of Fig. 15(a)]. In the case of $p<p_0$, such a modification is smaller because, for smaller densities, cells collide only as a result of a random walk due to the fluctuations of T. The smaller the density the longer it takes for random walk to ensure collisions. This results in the decay of $\tau_{\rm pair}(p)$ to zero as $p\to 0$.

Consider now the limit $p\ll p_0$, where collisions between cells are rare (a typical time between collision is much above T). We average the cellular dynamics over period $2T$ and

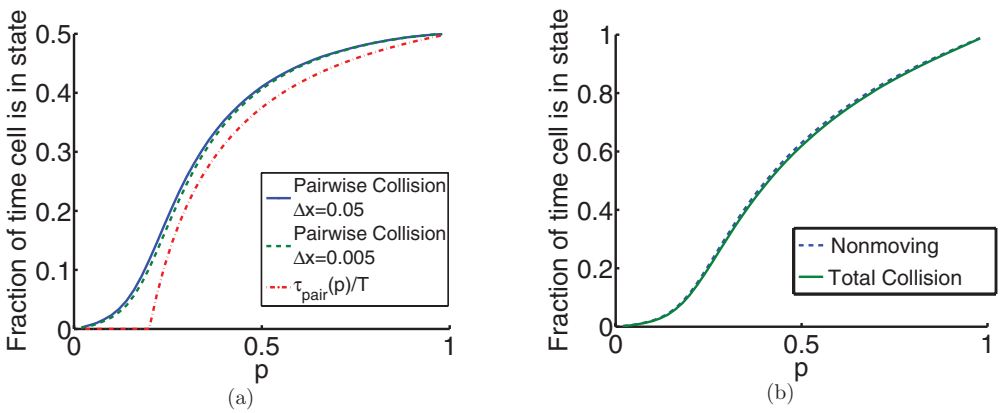

FIG. 15. (Color online) (a) Dependence of $\tau_{\rm pair}(p)/T$ from MSM simulations on the lattice size $\Delta x=0.05$ (dash-dotted line) and $\Delta x=0.005$ (dashed line) in comparison with the analytical expression Eq. (9) for $\tau_{\rm pair}/T$ (solid line). (b) Relative amount of time cells spend without moving per period T (dashed line) vs. relative total jam time τ/T (solid line) with $\Delta x=0.005$; τ includes both indirect and pairwise collisions. Decrease of Δx results in a better match between each pair of curves; $T=8$ for all simulations.

consider time scales larger than $2T$. After averaging, each cell experiences a random walk with jumps at the average distance $\pm\Delta T v$ at times $0, T, 2T, \ldots$; that is, we obtain a random walk at a discrete grid and in discrete time. To take into account collisions we introduce the effective size of cells from Eq. (6) in the limit of $\tau \to 0$:

$$L_{\text{eff}} = L + vT/2. \tag{10}$$

We now look at each cell as the effective object with the average size L_{eff} and we introduce the effective average distance Δx_{eff} between neighboring cells based on the normalization $(p/L)(L_{\text{eff}} + \Delta x_{\text{eff}}) = 1$ [see also Eq. (7) for the definition of p]. That normalization means that at the distance $L_{\text{eff}} + \Delta x_{\text{eff}}$ there is exactly one cell. Together [Eq. (10)] it gives

$$\Delta x_{\text{eff}} = \frac{L}{p} - L - \frac{vT}{2}. \tag{11}$$

Thus, in the limit $p \ll p_0$ we reduce, by averaging over period $2T$, the dynamics of the full MSM problem to 1D discrete random walk of cells with the effective length L_{eff} and the effective average intercellular distance Δx_{eff}. These effective cells are subject to excluded volume constraint. A problem of that type was studied in Ref. [24] in the continuous limit of infinitely small spatial jumps. In particular, the nonlinear diffusion equation of the general type Eq. (1) was derived. Using the diffusion coefficient D_0 (5) of the cells without collision (corresponds to the limit $p \to 0$), we obtain from Eq. (1) of Ref. [24] the following nonlinear diffusion coefficient:

$$D(p) = D_0 \left\{ P \frac{1 + (L_{\text{eff}}/L)^2 p^2}{[1 - (L_{\text{eff}}/L)p]^2} \right\}, \tag{12}$$

where L_{eff} is given by Eq. (10). For $p \to p_0$ the nonlinear diffusion coefficient Eq. (12) diverges because the approximation of the discrete random walk is valid only for $p \ll p_0$, which means that $\Delta x_{\text{eff}} \gg \Delta T v$. However, the tendency of the fast increase of $D(p)$ for $p \sim p_0/2$ explains the quick growth of the diffusion coefficient from D_0 to the typical values determined by BM analysis in Sec. VII A.

We now calculate the collision time τ_{coll} for $p \ll p_0$. Assume that, at some moment in time, two cells collide. Similarly to Ref. [24], we introduce an "extended collision" time τ_{extend} as the average time over which two initially collided cells will be separated by the distance Δx_{eff}. The problem of finding τ_{extend} can be formulated as a mean escape time problem for the motion of a pointwise particle in a domain of size Δx_{eff} with the reflecting boundary condition $\partial_x p = 0$ at one boundary (corresponding to the collision between cells) and the absorbing boundary condition $p = 0$ at the second boundary (corresponding to the escape of cells from the extended collision because, if the distance between cells is above Δx_{eff}, then collisions with other neighboring cells are equally probable [24]).

We define

$$m \equiv \frac{\Delta x_{\text{eff}}}{\Delta T v}, \tag{13}$$

where m is the value of Δx_{eff} in units of elementary jumps $\Delta T v$. In the limit $\Delta x_{\text{eff}} \gg \Delta T v$, the average number of jumps until the cells will be separated by Δx_{eff} is m^2, giving the average time τ_{extend} (the extended collision duration), during which the extended collision lasts

$$\tau_{\text{extend}} = m^2 T. \tag{14}$$

During that time the number of elementary collisions between two cells is m (similar to the number of times that a random walker passes through the initial point during m^2 jumps). It gives the cumulative collision time $\tau_{\text{coll,cumulative}}$ (a sum of all times when a cell is not moving because of collision) during the extended collision duration τ_{extend} as follows:

$$\tau_{\text{coll,cumulative}} = m \Delta T. \tag{15}$$

It follows from Eqs. (11), (13), (14), and (15) that the pairwise collision time $\tau_{\text{pair}} \equiv T \tau_{\text{coll,cumulative}}/\tau_{\text{extend}}$ per period T is given by

$$\tau_{\text{pair}}/T = \frac{\Delta T^2 v}{\left(\frac{L}{p} - L - \frac{vT}{2}\right)T}. \tag{16}$$

Figure 16 shows τ_{pair}/T versus p for $T = 8$. It is seen that for $p \lesssim 0.12$, the analytical approximation Eq. (16) is quite accurate. For $0.12 \lesssim p \lesssim 0.2 = p_0$, the number of effective jumps in the discrete random walk approximation, necessary to span Δx_{eff}, is no longer large compared with 1. But the approximation of Eq. (16) is based on continuous random walk (limit of infinitely small jumps). Thus, for $0.12 \lesssim p \lesssim 0.2 = p_0$, the approximation in Eq. (16) breaks down and the effect of discreteness of the number of effective jumps needs to be taken into account, which is beyond the scope of this paper. The solid line in Fig. 16 represents the simplest reduction of Eq. (16) for $p \ll L/L_{\text{eff}}$, when the effective size L_{eff} can be neglected, giving

$$\tau_{\text{pair}}/T = \Delta T^2 v p L^{-1} T^{-1}. \tag{17}$$

Also note that for $p < p_0$ the difference between the pairwise and the total collision times is negligible because clustering is insignificant for these densities so Eq. (16) is equally good for the description of both pairwise and total collision times. For $p > p_0$, the fluctuations of T result in a more efficient exploration of the space by cells that increases τ_{pair} in comparison with Eq. (9), explaining the difference between the solid curve and dashed curve in Fig. 15(a).

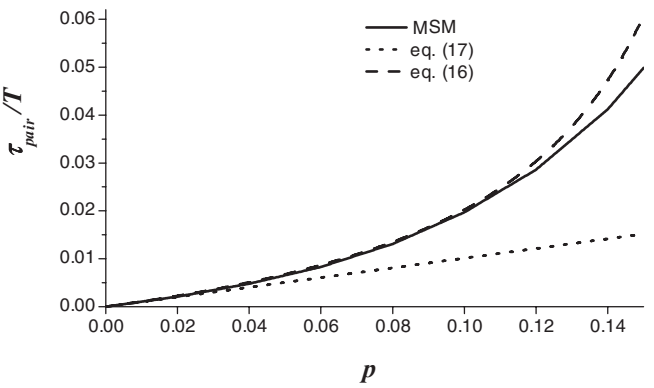

FIG. 16. Dependence of τ_{pair}/T on the density p from MSM simulations (solid line), Eq. (16) (dashed line), and Eq. (17) (dotted line).

We also performed simulations with decreased values of $\triangle x$ to demonstrate that $\triangle x = 0.05$ is already small enough to provide a good approximation of $\tau_{\text{pair}}(p)$ in comparison with the continuous limit $\triangle x \to 0$. Figure 15(a) shows convergence of $\tau_{\text{pair}}(p)$ to the analytical expression of Eq. (9) with the decrease of $\triangle x$.

The same MSM simulations were used to calculate the total collision time per period T. In simulations we distinguish two types of the total collision time. The first type is the total collision time τ itself (total jam time per period T), which includes both pairwise jams and indirect jams. The second type is the average time (per period T) cells spend without movement, which includes pairwise jams, indirect jams, and jams due to finite values of $\triangle x$. The third contribution occurs when two cells are attached to each other and move in the same direction. If a cell that follows another one is chosen by the MSM algorithm, then its movement is prevented by the second cell. This artificial effect is due to discretization and finite value of $\triangle x$. It disappears for $\triangle x \to 0$ so both types of the total collision time are the same in that limit. Figure 15(b) shows the time cells spend without movement per period T versus τ. It demonstrates that these total collision times (normalized to T) are very close to each other for $\triangle x \to 0.005$.

In contrast to the pairwise collision time used in Eq. (9), the Eq. (8) includes an extra term, $\tau O(\tau/T)$, which corresponds to the total jam time τ per period T. The dashed line in Fig. 17 shows the dependence of MSM simulations with the noise in T on $\tau(p)$ (with the standard $\Delta T_0 = 0.9$). We now approximate the term $\tau O(\tau/T)$ in Eq. (8) in the simplest possible way, neglecting noise in T as $O(x) = c_1 x$ with $c_1 = 2$, which yields

$$\tau_{\text{approx}}(p) = \tau_{\text{pair}}(p)\left[1 + 2\frac{\tau_{\text{pair}}(p)}{T}\right]\Theta(p - p_0), \quad (18)$$

where $\tau_{\text{pair}}(p)$ is given by Eq. (9). The Θ function reflects the neglect of noise in T similar to Eq. (9). We choose $c_1 = 2$ here to ensure correct asymptotic value $\tau_{\text{approx}}(1) = T$ for $p = 1$ because all cells are jammed all the time in that case. Figure 17 shows a good fit between Eq. (18) (solid line 2) and the total collision time per period τ/T from MSM simulations with $\Delta T_0 = 0.9$ (dashed line 1). It suggests that the addition of noise in Eq. (18) might result in a very good fit. Exact analytical theory is needed in order to verify this hypothesis, which is quite a challenging problem and which is outside of the scope of this paper.

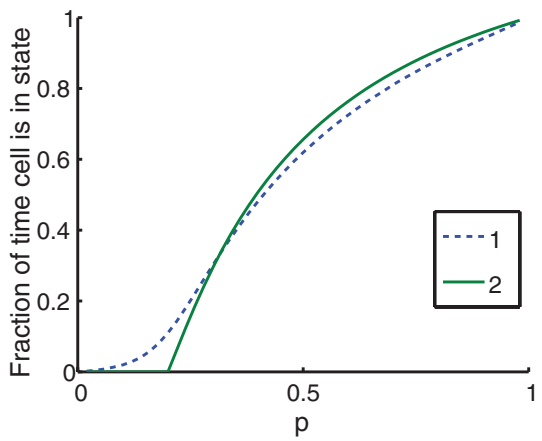

FIG. 17. (Color online) Plots of the total collision time per period τ/T from MSM simulations (dashed line 1) and from the analytic approximation of τ_{approx} (18) (solid line 2). The lattice size is $\triangle x = 0.005$ and $T = 8$.

IX. CONCLUSIONS AND DISCUSSION

A connection was established in this paper between a stochastic 1D model (MSM) of microscopic motion of the system of regularly reversing self-propelled rod-shaped cells and a nonlinear diffusion equation describing the macroscopic behavior of this system. Stochastic dynamics averaged over an ensemble was shown to be in a very good agreement with the numerical solutions of the nonlinear diffusion equation (1), where the diffusion coefficient was obtained using BM analysis. Critical density p_0 was found such that for $p < p_0$ the cellular diffusion is dominated by the diffusion (random walk) of individual cells while for $p > p_0$ the diffusion is dominated by the collisions between cells. p_0 was determined Eq. (4) from the condition that cells do not jam with each other in the no-noise limit. We found that the role of noise in the reversal period is crucial. Without noise, BM analysis cannot reproduce the MSM dynamics, which means that nonlinear diffusion is not a good approximation for it. However, even a relatively small level ($\Delta T_0/T \simeq 0.1$) of noise produces excellent agreement between BM-based nonlinear diffusion and MSM simulations. The primary role of a small noise is to ensure randomization of collisions among different cells.

An analytical approximation of the pairwise collision time τ_{pair} [Eq. (9)] and semianalytical fit for the total jam time per reversal period $\tau_{\text{approx}}(p)$ [Eq. (18)] have also been obtained. Frequent collisions for $p > p_0$ are responsible for the nonlinear diffusion of the cellular density. For $p < p_0$, cells tend to spread out so they collide only if fluctuations of the reversal time are taken into account. Without such fluctuations there are no collisions and no cellular transport is possible because cells experience periodic motion in space and time. There still remains quite a challenging problem of developing a full statistical theory of 1D self-propelled rod dynamics with reversals that would be applicable for all densities. Such a theory would require a detailed description of formation and interaction of large cellular clusters.

It was also shown that the nonlinear diffusion coefficient $D(p)$ used to describe the macroscopic process, changes depending on the reversal period. Small and large reversal periods yield diffusion coefficients that favor high and low cell-density diffusion, respectively, as is shown in Fig. 8. Since the dynamics of the system is determined by the dimensionless parameters vT/L (the ratio of distance traveled by cells between reversals and the cell length) and p, increase of the speed at which cells move is equivalent to the increase of the reversal period. Thus, cell populations with small T are able to spread out effectively at high densities while large T promotes cell population swarming at smaller densities.

An interesting problem to be studied in future work is to determine the optimal choice of reversal time T maximizing the swarming rate of an *M. xanthus* colony using a nonlinear diffusion equation and then to compare it with the one obtained in Ref. [5] using a stochastic model.

ACKNOWLEDGMENTS

The work of P.L. was supported by NSF Grant No. DMS 0719895 and a UNM RAC grant. R.G. was supported by the University of Notre Dame's CAM Fellowship and partially supported by NSF Grant No. DMS 0931642. M.A. was partially supported by NSF Grant No. DMS 0719895.

APPENDIX A: BOLTZMANN-MATANO ANALYSIS

In this appendix we review BM analysis (see Ref. [33] for details) for the readers' convenience. Assume that the process we are studying can be modeled using the nonlinear diffusion equation (1) with some unknown nonlinear diffusion coefficient, $D(p)$.

Boltzmann-Matano analysis allows us to recover $D(p)$ from the 1D dynamics of the cellular density p with the stepwise initial condition

$$p(x,0) = \begin{cases} p_L, & \text{if} \quad x < x_M \\ p_R & \text{if} \quad x > x_M \end{cases} \tag{A1}$$

at infinite 1D domain. Here we assume that $p_L > p_R$.

The special property of the stepwise initial condition is that it has no spatial scale (the spatial size of the system is infinite and the spatial scale of the jump at $x = x_M$ is zero). The only possible solution then has a self-similar form, $p(\zeta)$, which was found by Boltzmann in 1894. Here

$$\zeta = (x - x_M)/t^{1/2}, \tag{A2}$$

which is motivated by a self-similar solution of a heat equation (for $D = \text{const}$). x_M is a reference point also known as the Matano interface. Assuming that $p(\zeta)$ does not depend on t explicitly, we obtain that $\frac{\partial}{\partial t} p(\zeta) = -\frac{1}{2}\frac{\zeta}{t}\frac{\partial}{\partial \zeta} p(\zeta)$ and $\frac{\partial}{\partial x} p(\zeta) = \frac{1}{t^{1/2}}\frac{\partial}{\partial \zeta} p(\zeta)$, which allows us to reduce Eq. (1) to

$$-\frac{\zeta}{2}\frac{\partial}{\partial \zeta} p = \frac{\partial}{\partial \zeta}\left[D(p)\frac{\partial}{\partial \zeta} p \right]. \tag{A3}$$

Since the solutions to a nonlinear diffusion equation with stepwise initial conditions are monotonic, it follows that for any given fixed time the function $p(x)$ is invertible with respect to x. Below we use the notation $x(p)$ for the inverse of $p(x)$. Integrating both sides of Eq. (A3) with respect to ζ yields

$$-\frac{1}{2t^{1/2}} \int_{p_L}^{p} (x(p) - x_M) dp = D(p) p_\zeta,$$

where the left-hand side follows from

$$\int_{-\infty}^{\zeta} \zeta \frac{\partial p}{\partial \zeta} d\zeta = \int_{p_L}^{p} \zeta(p) dp = \frac{1}{t^{1/2}} \int_{p_L}^{p} (x(p) - x_M) dp.$$

Since $\frac{\partial p}{\partial \zeta} = t^{1/2}\frac{\partial p}{\partial x}$, the equation can be rewritten as

$$D(p) = -\frac{1}{2t}\left[\frac{\partial p}{\partial x}\right]^{-1} \int_{p_L}^{p} (x(p) - x_M) dp,$$

which gives the Boltzmann description of the diffusion equation. It is now possible to calculate the appropriate value of the interface, x_M, to ensure that the diffusion calculation is consistent. Specifically, since mass diffuses from left to right across the interface, there is a mass conservation equation where the mass lost on the left of the interface should equal the mass gained on the right of the interface,

$$\int_{-\infty}^{x_M} [p_L - p(x)] dx = \int_{x_M}^{\infty} [p(x) - p_R] dx.$$

Again inverting $p(x)$, we can calculate the area under of the integrals in terms of $x(p)$ to get the following equivalent expression:

$$\int_{p_M}^{p_L} [x(p) - x_M] dp = \int_{p_R}^{p_M} [x_M - x(p)] dp,$$

which simplifies to

$$\int_{p_L}^{p_R} [x(p) - x_M] dp = 0.$$

Mass conservation occurs precisely when

$$x_M = \frac{\int_{p_L}^{p_R} x(p) dp}{p_L - p_R}, \tag{A4}$$

which is Matano's result to determine x_M if it is unknown in advance.

In our simulations we know x_M in advance so, in fact, we use Boltzmann analysis but not BM analysis (except the additional tests discussed in Appendix B). Moreover, in our simulations $p_L = p_{\max}$ and $p_R = 0$.

APPENDIX B: ACCURACY OF BOLTZMANN-MATANO ANALYSIS

Boltzmann-Matano analysis, described in Appendix A, is defined on an infinite spatial interval with stepwise initial conditions only. Assume now that we apply BM analysis for the top-hat initial conditions as described in Sec. II. In that case, BM analysis is only approximate because initial conditions include spatial scale, x_{width}, which is the spatial width of the top hat. A self-similar solution of Appendix A does not agree with the top hat. That solution is only approximately valid in the neighborhood of each of two steps of the top hat. Because of spatial symmetry it is enough to consider any of these two steps. To estimate the accuracy of the BM analysis in that case we note that if the density at $x = 0$ (middle of the top hat) remains nearly constant, then BM analysis is still applicable (except small unavoidable corrections because for any $t > 0$ density is never exactly constant). Assuming that the diffusion coefficient $D(p) \sim 1$, we roughly estimate that the width of the initial top hat doubles with time when $D(p)t_0/x_{\text{width}}^2 \sim 1$, which gives $t_0 \sim 10^6$ for $x_{\text{width}} = 1000$. For $t \ll t_0$ a change of density in the middle of top hat is small in agreement with Fig. 7. A similar limitation of the BM analysis is that the total spatial width of the simulation domain must exceed the width of the top hat several times to ensure that the cellular density remains low at boundaries as seen in Fig. 7.

As additional test of BM analysis we varied the domain length and width of the initial top-hat distribution calculating diffusion coefficient by BM analysis from MSM simulations [Fig. 18(a)]. We observed that small top-hat width $\sim$100 is not enough for applicability of BM analysis [dash-dotted curve in Fig. 18(a)] while top-hat widths $\gtrsim$1000 total domain lengths $\gtrsim$4000 are far enough for such applicability. Figure 18(b)

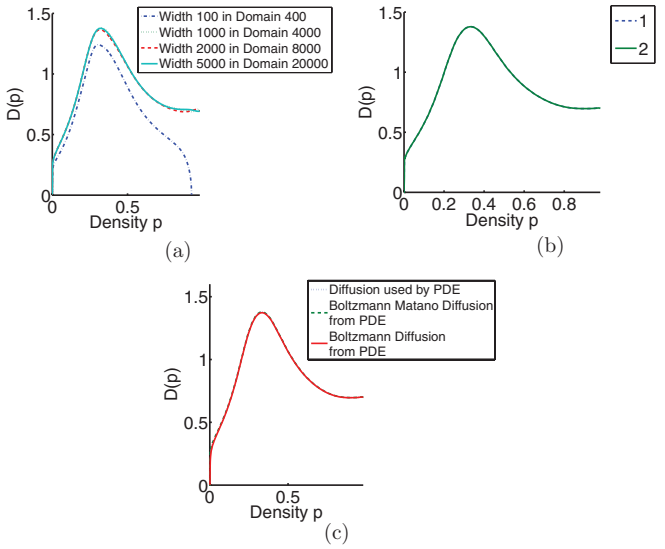

FIG. 18. (Color online) (a) The nonlinear diffusion coefficient $D(p)$ determined from BM analysis of MSM simulations with different initial top-hat widths. The density profiles at $t_D = 500$ were used for all curves. It is seen that curves for the widths 1000 and above are almost undistinguishable. (b) $D(p)$ obtained from MSM simulations (dashed line, curve 1) and PDE simulation (solid line, curve 2) for the top-hat initial conditions of width 600. Density profile at time $t_D = 1000$ is used for BM analysis. (c) Comparison of BM analysis with Boltzmann analysis from PDE density profile at $t_D = 500$. Dashed line is $D(p)$ used to produce density profiles from PDE simulations. All curves at (b) and (c) are almost indistinguishable.

compares $D(p)$ obtained from PDE simulation (solid line) and MSM simulations (dashed line) for the top-hat initial conditions of width 600. Difference between these curves is almost indistinguishable. This indicates that our statistical ensemble in MSM simulations is large enough to avoid influence of noise in the data on the diffusion curve. We also tested MSM data with and without the Gaussian filter and obtained the same diffusion curves. Larger widths were also tested and proven to match very well, but the results are not displayed here. From these observations, we conclude that the generated diffusion curves are independent of the width of the top hat if the top hat is wide enough. This means that the center and boundaries of the spatial domain have constant density in time.

We would like to point out, to avoid confusion, that the BM analysis is needed only to determine the diffusion curves at reasonably small times ($t \ll t_0$). After that, PDE simulations can be run with these diffusion curves for much longer times (when density is changing both at the middle of the top hat and at the boundaries). For these much longer times a very good agreement has been also observed between MSM simulations and PDE simulations (see, e.g., Figs. 4 and 8).

As discussed in Sec. VII, another limitation of the BM analysis is the loss of numerical precision near $p(x) = \text{const}$ because it requires calculating $[dp(x)/dx]^{-1}$. Figures 7(a), 7(b), and 18 show jumps of $D(p)$ near $p = 1$ due to such a loss of numerical precision, which can be fixed by the polynomial extrapolation. This is, however, not necessary because these jumps do not change the results of the PDE simulations in a significant way.

We also tested BM analysis versus Boltzmann analysis as shown in Fig. 18(c). Although x_M is known from the top-hat initial conditions, for the finite width of the top-hat one can ask whether allowing x_M to be located around the step of the top hat could improve the accuracy of the BM analysis to determine $D(p)$. In this sense, x_M could be viewed as an additional fitting parameter needed for accommodating the top-hat width. Figure 18(c) compares the diffusion curves obtained using BM analysis, Boltzmann analysis, and an exact diffusion curve. The difference in accuracy between the BM analysis and Boltzmann analysis is very small. It appears that the advantage of using BM analysis versus Bolztmann analysis is insignificant in our case.

[1] D. Kaiser, Curr. Bio. **17**, 561 (2007).
[2] Y. Wu, D. Kaiser, Y. Jiang, and M. Alber, PLoS Comput. Biol. **3**, e253 (2007).
[3] A. Kaiser and C. Crosby, Cell Motility **3**, 227245 (1983).
[4] R. Yu and D. Kaiser, Mol. Microbiol. **63**, 454 (2007).
[5] Y. Wu, D. Kaiser, Y. Jiang, and M. Alber, Proc. Natl. Acad. Sci. USA **106**, 1222 (2009).
[6] T. Mignot, J. Merlie, and D. Zusman, Science **310**, 855 (2005).
[7] D. E. Whitworth, ed., Myxobacteria *Multicellularity and Differentiation* (ASM Press, Washington, DC, 2008).
[8] Y. Wu, Y. Jiang, D. Kaiser, and M. Alber, Phys. Biol. **8**, 055003 (2011).
[9] B. D. Blackhart and D. Zusman, Proc. Natl. Acad. Sci. USA **82**, 8767 (1985).
[10] D. Kaiser, in *Myxobacteria, Multicellularity, and Differentiation*, edited by D. E. Whitworth (ASM Press, Washington, DC, 2008), p. 93.
[11] E. M. Mauriello, T. Mignot, Z. Yang, and D. R. Zusman, Microbiol. Mol. Biol. Rev. **74**, 229 (2010).
[12] C. W. Harvey, F. Morcos, C. R. Sweet, D. Kaiser, S. Chatterjee, X. Liu, D. Z. Chen, and M. Alber, Phys. Biol. **8**, 026016 (2011).
[13] O. A. Igoshin, A. Mogilner, R. D. Welch, D. Kaiser, and G. Oster, Proc. Natl. Acad. Sci. USA **98**, 14913 (2001).
[14] O. A. Igoshin and G. Oster, Math Biosci. **188**, 221 (2004).
[15] O. A. Igoshin, R. D. Welch, D. Kaiser, and G. Oster, Proc. Natl. Acad. Sci. USA **101**, 4256 (2004).
[16] M. S. Alber, Y. Jiang, and M. A. Kiskowski, Physica D **191**, 343 (2004).
[17] B. V. Reddy and S. N. Khanna, Phys. Rev. Lett. **93**, 068301 (2004).
[18] M. A. Kiskowski, Y. Jiang, and M. S. Alber, Phys. Biol. **1**, 173 (2004).
[19] D. Kaiser, Curr. Biol. **17**, R561 (2007).
[20] M. R. D'Orsogna, M. A. Suchard, and T. Chou, Phys. Rev. E **68**, 021925 (2003).
[21] T. J. Newman and R. Grima, Phys. Rev. E **70**, 051916 (2004).

[22] M. Alber, N. Chen, T. Glimm, and P. M. Lushnikov, Phys. Rev. E **73**, 051901 (2006).
[23] M. Alber, N. Chen, P. M. Lushnikov, and S. A. Newman, Phys. Rev. Lett. **99**, 168102 (2007).
[24] P. M. Lushnikov, N. Chen, and M. Alber, Phys. Rev. E **78**, 061904 (2008).
[25] M. Alber, R. Gejji, and B. Kazmierczak, Appl. Math. Lett. **22**, 1645 (2009).
[26] C. S. Patlak, Bull. Math. Biophys. **15**, 311 (1953).
[27] E. F. Keller and L. A. Segel, J. Theor. Biol. **26**, 399 (1970).
[28] M. P. Brenner, P. Constantin, L. P. Kadanoff, A. Schenkel, and S. C. Venkataramani, Nonlinearity **12**, 1071 (1999).
[29] P. M. Lushnikov, Phys. Lett. A **374**, 1678 (2010).
[30] S. Dejak, P. Lushnikov, Y. Ovchinnikov, and I. Sigal (to be published).
[31] P. J. Murray, C. M. Edwards, M. J. Tindall, and P. K. Maini, Phys. Rev. E **80**, 031912 (2009).
[32] R. Welch and A. D. Kaiser, Proc. Natl. Acad. Sci. USA **98**, 14907 (2001).
[33] M. E. Glicksman, *Diffusion in Solids: Field Theory, Solid-State Principles, and Applications* (Wiley, New York, 2000).
[34] L. D. Landau and E. M. Lifshitz, *Statistical Physics*, 3rd ed. (Butterworth-Heinemann, London, 1980).
[35] M. Nixon and A. Aguado, *Feature Extraction and Image Processing* (Academic Press, New York, 2008).
[36] C. W. Gardiner, *Handbook of Stochastic Methods for Physics, Chemistry, and the Natural Sciences* (Springer-Verlag, Berlin, 2004).
[37] P. C. Hemmer, Physica **27**, 79 (1961).
[38] S. Ispolatov and P. L. Krapivsky, Physica A **252**, 165 (1998).