Filling the gaps: A robust description of adhesive birth-death-movement processes

Stuart T. Johnston,[1,*] Ruth E. Baker,[2] and Matthew J. Simpson[1]
[1]*School of Mathematical Sciences, Queensland University of Technology, Brisbane, Australia*
[2]*Mathematical Institute, University of Oxford, Oxford, United Kingdom*
(Received 28 September 2015; revised manuscript received 8 March 2016; published 21 April 2016)

Existing continuum descriptions of discrete adhesive birth-death-movement processes provide accurate predictions of the average discrete behavior for limited parameter regimes. Here we present an alternative continuum description in terms of the dynamics of groups of contiguous occupied and vacant lattice sites. Our method provides more accurate predictions, is valid in parameter regimes that could not be described by previous continuum descriptions, and provides information about the spatial clustering of occupied sites. Furthermore, we present a simple analytic approximation of the spatial clustering of occupied sites at late time, when the system reaches its steady-state configuration.

DOI: 10.1103/PhysRevE.93.042413

I. INTRODUCTION

Birth, death, and movement of individuals are key components of collective behavior, relevant to tissue repair [1–5] and polymer aggregation [6]. Lattice-based random walks are often used to describe these processes [2,7–10]. For example, Deroulers *et al.* [9] use a random walk to model the migration of glioma cells, while Hackett-Jones *et al.* [10] use a random walk to describe cellular aggregation patterns arising from nonlocal interactions. Mackie *et al.* [6] are interested in molecular aggregation and use random walks to examine the formation of micelles. In the context of active transport, Illien *et al.* [11] use lattice-based random walks to model diffusion in a crowded single-file environment. While these stochastic random-walk models provide insight into collective behavior, performing a sufficient number of stochastic realizations to obtain representative average behavior can be computationally expensive. Furthermore, relying solely upon stochastic simulations provides little opportunity for analyzing how parameters in the simulation are related to the collective behavior.

There is significant interest in continuum descriptions that produce robust approximations of the average behavior. Additionally, there is considerable interest in analytic approaches that describe the average steady-state behavior of the random-walk model, particularly for the one-dimensional asymmetric exclusion process [8]. To address both the analytic intractability and the computational expense associated with obtaining representative average behavior of random-walk models, mean-field (MF) descriptions of lattice-based adhesive birth-death-movement (ABDM) processes have been developed [3,12,13]. Unfortunately, MF descriptions only apply in limited parameter regimes due to the neglect of spatial correlations [1,14–17]. Specifically, MF descriptions only apply when adhesion (or repulsion) is sufficiently weak, and the ratio of both the birth and death rates to the movement rate is sufficiently small [14,15]. To address this, corrected mean-field (CMF) descriptions that approximately incorporate spatial correlations have been proposed [14,15,18–20]. While CMF descriptions provide an accurate approximation of the average behavior for a wider range of parameter regimes, they are invalid when the rates of birth and death are not sufficiently small compared to the rate of movement [14]. Here we interpret ABDM processes in terms of groups of contiguous occupied and vacant lattice sites, and we present the corresponding continuum description. We note that our description is limited to spatially independent initial conditions. This implies that our description is translationally invariant and cannot describe processes that are initially spatially dependent. Our method provides accurate predictions of the average behavior in parameter regimes where MF and CMF descriptions are invalid. Furthermore, our description provides information about spatial clustering, and we give a simple analytic approximation of the spatial clustering of occupied sites when the system has reached steady state.

II. DISCRETE MODEL

We consider a periodic one-dimensional lattice-based random walk, where each site may be occupied by, at most, one agent [21]. Isolated agents undergo birth, death, and movement events at constant rates P_p, P_d, and P_m, respectively [14]. During a birth event, an agent attempts to place a daughter at a randomly selected nearest-neighbor site. This event is successful provided that the selected site is vacant. A death event removes an agent and the associated site becomes vacant. During a movement event, an agent attempts to move to a randomly selected nearest-neighbor site. This selection is unbiased if both nearest-neighbor sites are vacant. If one nearest-neighbor site is occupied, the vacant nearest-neighbor site is selected with probability $(1-\alpha)/2$, where $\alpha \in [-1,1]$ represents the strength of agent-agent adhesion ($\alpha > 0$) or repulsion ($\alpha < 0$) [12,15]. We use the Gillespie algorithm [22] to simulate the random walk. To generate representative behavior, we perform M identically prepared realizations of the random walk to calculate the average of the summary statistic of interest.

III. CHAIN-AND-GAP DESCRIPTION

Here we interpret the system as a combination of groups of contiguous occupied and vacant sites of length $i \in [1,N]$, where N is the number of sites. We refer to groups of contiguous occupied sites as *chains* and groups of contiguous

*s17.johnston@qut.edu.au

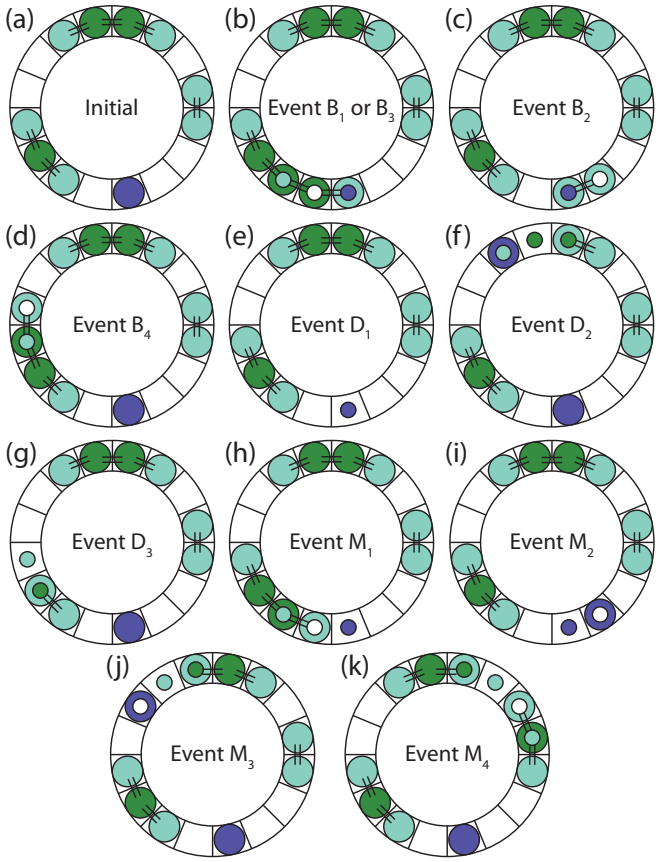

FIG. 1. (a) Initial lattice configuration containing single [purple (dark gray)], edge [blue (light gray)], and middle [green (gray)] agents. Lattice configuration after potential (b)–(d) birth, (e)–(g) death, and (h)–(k) movement events. Inset circles represent the initial lattice configuration. Dashed lines represent agents influenced by agent-agent adhesion or repulsion.

vacant sites as *gaps*. For example, Fig. 1(a) shows a lattice configuration that contains four chains: one each of length 1, 2, 3, and 4; and four gaps: two each of length 1 and 2. Instead of considering how a birth, death, or movement event affects the occupancy of an individual site, we consider how these events change the lengths of the chains and gaps. The success of birth and movement events depends on the "type" of agent, that is, whether an agent has zero, one, or two nearest-neighbor agents. We refer to these as *single*, *edge*, or *middle* agents, respectively. Single agents can move to, or place a daughter agent at, both nearest-neighbor sites, whereas edge agents can move to, or place a daughter agent at, only one nearest-neighbor site. Due to crowding, it is impossible for middle agents to undergo successful movement or birth events. Potential death events are not affected by crowding. The net birth rate is $P_p[2N_S(t)+N_E(t)]/2 - P_d[N_S(t)+N_E(t)+N_M(t)]$, where $N_S(t)$, $N_E(t)$, and $N_M(t)$ are the number of single, edge, and middle agents, respectively. We can calculate $N_S(t)$, $N_E(t)$, and $N_M(t)$ provided that we know the number of chains of length $n \in [1,N]$, which we denote $C_n(t)$. For all $n \geqslant 2$, there are two edge agents and $n-2$ middle agents per chain. The number of single agents is $C_1(t)$. We denote the number of gaps of length $m \in [1,N]$ by $G_m(t)$. For notational convenience, we henceforth refer to N_S, N_E, N_M, C_n, and G_m without explicitly noting temporal dependence.

There are 11 distinct birth, death, or movement events, denoted by B_j, $j=1,2,3,4$, D_j, $j=1,2,3$, and M_j, $j=1,2,3,4$. In Figs. 1(b)–1(k), we demonstrate how each of these events affects the configuration of agents in Fig. 1(a). For an arbitrary configuration, the influence of each event can be described as follows. We note that an increase (decrease) in C_n or G_m refers to an increase (decrease) of 1.

Event B_1: A single agent undergoes birth and places a daughter agent into a gap of length 1, which is next to a chain of length n [Fig. 1(b)]. C_1, G_1, and C_n decrease, C_{n+2} increases.

Event B_2: A single agent undergoes birth and places a daughter agent into a gap of length $m \geqslant 2$ [Fig. 1(c)]. C_1 and G_m decrease, C_2 and G_{m-1} increase.

Event B_3: An edge agent, part of a chain of length n_1, undergoes birth and places a daughter agent into a gap of length 1, which is next to a chain of length n_2 [Fig. 1(b)]. C_{n_1}, C_{n_2} and G_1 decrease, $C_{n_1+n_2+1}$ increases.

Event B_4: An edge agent, part of a chain of length n, undergoes birth and places a daughter agent into a gap of length $m \geqslant 2$ [Fig. 1(d)]. C_n and G_m decrease, C_{n+1} and G_{m-1} increase.

Event D_1: A single agent, with neighbor gaps of length m_1 and m_2, undergoes death [Fig. 1(e)]. C_1, G_{m_1}, and G_{m_2} decrease, $G_{m_1+m_2+1}$ increases.

Event D_2: A middle agent, part of a chain of length n, undergoes death [Fig. 1(f)]. C_n decreases, G_1, C_{n_1}, and C_{n_2} increase, where $n_1+n_2+1=n$. We note that the expected increase of C_{n_1} and C_{n_2} is uniform for all $n_1,n_2 \in [1,n-2]$, $n_1+n_2+1=n$.

Event D_3: An edge agent, part of a chain of length n, which is next to a gap of length m, undergoes death [Fig. 1(g)]. C_n and G_m decrease, C_{n-1} and G_{m+1} increase.

Event M_1: A single agent undergoes movement into a gap of length 1, which is next to a chain of length n [Fig. 1(h)]. The single agent is also next to a gap of length m. C_1, C_n, G_1, and G_m decrease, C_{n+1} and G_{m+1} increase.

Event M_2: A single agent undergoes movement into a gap of length $m_1 \geqslant 2$, away from a gap of length m_2 [Fig. 1(i)]. G_{m_1} and G_{m_2} decrease, G_{m_1-1} and G_{m_2+1} increase.

Event M_3: An edge agent, part of a chain of length n, undergoes movement into a gap of length $m \geqslant 2$ [Fig. 1(j)]. C_n and G_m decrease, C_1, C_{n-1}, G_1, and G_{m-1} increase.

Event M_4: An edge agent, part of a chain of length n_1, undergoes movement into a gap of length 1, which is next to a chain of length n_2 [Fig. 1(k)]. C_{n_1} and C_{n_2} decrease, C_{n_1-1} and C_{n_2+1} increase.

We obtain transition rates between possible states of the system by considering the result of each potential event and the rate at which it occurs, and, subsequently, we obtain a system of ordinary differential equations (ODEs) describing dC_n/dt and dG_m/dt, $n,m \in [1,N]$. For single agents, birth and movement events are never aborted due to crowding, and, therefore, the rate at which each of these events occurs is P_pC_1 and P_mC_1, respectively. We note that α does not influence single agents. For edge agents, birth events are aborted, on average, half the time. However, there are two edge agents for each C_n, $n \geqslant 2$, which implies that the rate at which these events occur for a

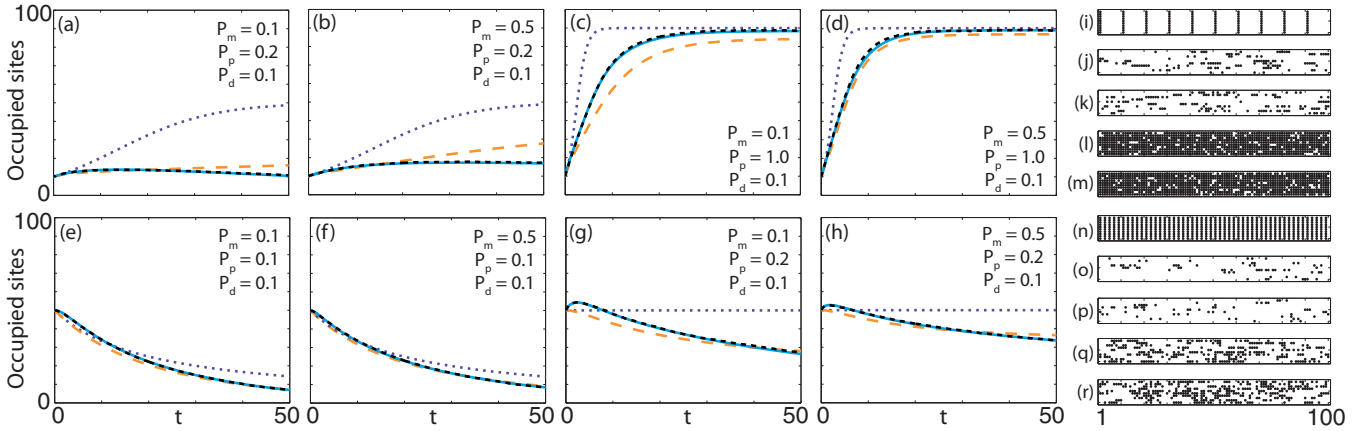

FIG. 2. Comparison between averaged discrete (black, dashed), MF (purple, dotted), CMF (orange, long dashed), and C&G (cyan, solid) results for a suite of parameter regimes and (a)–(d) 10 and (e)–(f) 50 uniformly distributed initially occupied sites. For all results $M = 1000$, $N = 100$, and $\alpha = 0$. (i)–(r) Ten discrete snapshots for (i),(n) the two initial conditions and (j)–(m) and (o)–(r) the parameter regimes in (a)–(d) and (e)–(h), respectively, at $t = 50$.

chain of length n is $P_p C_n$. We follow a similar process for the rate of movement events. However, the probability of a successful movement event for edge agents is influenced by α and hence the rate of movement for a chain of length n is $(1-\alpha)P_m C_n$ for $n \geqslant 2$. Death events are never aborted due to crowding, which implies that the rate at which death events occur for a chain of length n is $nP_d C_n$. These rates describe ABDM events in terms of the agent type, but they do not describe the rate and result of possible subevents. For example, events B_1 and B_2 are both birth events for single agents, but they occur depending on whether the gap that the daughter agent is placed into is length 1 (event B_1) or greater (event B_2). The proportion of single-agent birth events that are event B_1 is equivalent to the probability that, given that we have selected a single agent, there is a neighboring gap of length 1. We denote this probability by $P(G_1|C_1)$. Due to the periodic geometry, a single agent can be next to gaps of length $j \in A(C_1)$, where $A(C_1) = \{i \in \mathbb{Z} | 1 \leqslant i \leqslant N-3\} \cup \{N-1\}$. We make the assumption that $P(G_1|C_1) = G_1 / \sum_{j \in A(C_1)} G_j$, that is, the ratio of gaps of length 1 to all gaps that can exist next to a chain of length 1. The expected change of both C_1 and G_1, due to event B_1, is $P_p C_1 P(G_1|C_1)/2$. Event B_1 also affects C_n and C_{n+2} and, as such, we require the expected change of C_n and C_{n+2}, due to event B_1. To obtain an expression for the expected change for C_n and C_{n+2}, due to event B_1, we consider the probability that, given that we have a single agent with a neighboring gap of length 1, the other neighboring chain, with respect to the gap, is of length n (Appendix A). We can calculate the average outcome of event B_1 by considering potential n values and the associated transition rate for each n. Combining all possible events with their respective rates gives the governing system of ODEs (Appendix A).

Our current framework can be extended to describe additional details. However, in this first presentation of the chain-and-gap (C&G) framework, we focus on the most straightforward periodic one-dimensional case, with translationally invariant initial conditions. Other boundary conditions could be considered; however, this would require that the number of chains and gaps of each length is known at each lattice site, and hence the number of equations in the governing system would be $O(N^2)$ rather than $O(N)$. For example, we could consider nonperiodic geometry, with specific conditions imposed on the two boundary lattice sites. We acknowledge that our approach is currently limited to one-dimensional processes. While this restricts the type of physical applications that our approach can describe, there is considerable interest in one-dimensional exclusion processes throughout the physical and life sciences [11,12,23]. Furthermore, the influence of the spatial correlations is most pronounced in one dimension [14], and, therefore, it is logical to focus on the development of new approximations for one-dimensional processes.

IV. TRADITIONAL CONTINUUM DESCRIPTIONS

Typically, continuum descriptions of ABDM processes are obtained by considering the time rate of change of the average occupancy of sites [13–15,19]. These descriptions, which describe the number of occupied sites, $S(t)$, take the form

$$\frac{dS(t)}{dt} = P_p S(t)\left(1 - \frac{F(t)S(t)}{N}\right) - P_d S(t), \qquad (1)$$

where $F(t)$ represents the nearest-neighbor correlation [14,24]. Traditional MF descriptions implicitly assume that $F(t) \equiv 1$ [24]. For CMF descriptions, which approximate spatial correlation between sites by considering the evolution of pairs [25,26], $F(t)$ is described by a system of ODEs [14]. This system is truncated through an appropriate moment closure method, such as the Kirkwood superposition approximation [27,28].

V. RESULTS

To examine whether our C&G description provides an accurate approximation of the lattice-based ABDM process, we first calculate the evolution of the number of occupied sites. The C&G system of ODEs is solved using an adaptive Runge-Kutta method with a strict truncation error control of 10^{-8} [29]. The number of occupied sites in the C&G description is $S_{\mathrm{C\&G}} = \sum_{i=1}^{N} i C_i$. In Figs. 2(a)–2(h), we present occupancy

TABLE I. Comparison of the time taken to perform (i) a single realization of the discrete model; (ii) 1000 realizations of the discrete model, and (iii) a numerical solution of the C&G system of equations for the parameter values in Figs. 2(a)–2(h). All solutions are obtained using a single 3.0 GHz Intel i7-3540M desktop processor.

	Fig. 2(a)	Fig. 2(b)	Fig. 2(c)	Fig. 2(d)	Fig. 2(e)	Fig. 2(f)	Fig. 2(g)	Fig. 2(h)
Discrete model (single realization)	0.08 s	0.19 s	1.40 s	1.91 s	0.10 s	0.24 s	0.23 s	0.48 s
Discrete model (1000 realizations)	76.7 s	192.9 s	1401.7 s	1907.3 s	99.4 s	241.9 s	233.5 s	483.1 s
C&G description	21.1 s	28.2 s	36.8 s	39.1 s	26.8 s	27.9 s	30.3 s	27.5 s

evolution curves obtained from the averaged discrete model, and the MF, CMF, and C&G descriptions for a suite of parameter regimes and two different initial conditions. All results in Fig. 2 are without adhesion or repulsion ($\alpha = 0$), and the discrete initial conditions are presented in Figs. 2(i) and 2(n). For all parameter regimes and initial conditions, the C&G description provides extremely accurate predictions, even when both the MF and CMF descriptions are invalid due to the emergence of significant spatial correlations [14]. We illustrate correlations between sites by presenting snapshots of the clustering in the discrete model in Figs. 2(j)–2(m) and Figs. 2(o)–2(r) at $t = 50$. The C&G description implicitly accounts for spatial correlations by considering contiguous occupied sites as a single object, which necessarily has neighboring vacant sites. Both the transient and steady-state behaviors of the system are correctly predicted by the C&G description. For example, in Fig. 2(b), the MF description inaccurately predicts both the transient and steady-state behaviors. The CMF model predicts the transient behavior accurately for $t < 20$, but provides inaccurate predictions for $t > 20$ and, consequently, the steady-state behavior. The C&G description correctly predicts that at steady state there are, on average, approximately 18 occupied sites, as well as accurately describing the transient behavior of the system.

We compare the time required to perform a single realization of the discrete model, 1000 realizations of the discrete model, and to solve the C&G system of equations in Table I. We note that in all cases it is significantly faster to perform a single realization of the discrete model. However, it is significantly slower to perform sufficiently many realizations to obtain meaningful average behavior, compared to obtaining the numerical solution to the C&G system of equations.

A. Clustering

Additionally, the C&G description provides information about the clustering of both occupied and vacant sites. In Fig. 3 we compare predictions from the C&G description with the averaged discrete model for C_n and G_m for a range of parameter regimes and three different initial conditions. As neither the MF or CMF description contains this information, we are unable to provide comparisons with the predictions from these descriptions. We observe that the C&G description provides accurate predictions for all cases, describing the monotonically decreasing relationship between both C_n and n, and G_m and m. An increase in P_p reduces C_1 and increases the number of longer chains. As expected, we observe the opposite behavior for the gaps. We note that the C&G description provides slightly less accurate predictions for the lowest number of initially occupied sites [Figs. 3(c) and 3(f)], a result that is consistent with other continuum descriptions [14].

In Figs. 4(a)–4(c), we present occupancy evolution profiles, C_n, and G_m, for different α. The number of occupied sites decreases with α due to the clustering that arises in the presence of adhesion, which we observe in Figs. 4(b) and 4(c), where there are significantly fewer chains and gaps of short length compared to when $\alpha \leqslant 0$. Snapshots of the discrete model in Figs. 4(d)–4(g) confirm this, as longer chains are present when $\alpha > 0$. For all adhesion and repulsion values, the C&G predictions match the averaged discrete model extremely well.

B. Spatial correlations

Traditional continuum descriptions of ABDM processes describe the evolution of the number of occupied sites, $S(t)$. If there is no adhesion or repulsion present in the discrete model, traditional continuum descriptions are described by Eq. (1) [14,24]. Traditional MF descriptions implicitly assume the spatial correlation between the occupancy of nearest-neighbor lattice sites, $F(t) \equiv 1$, while CMF descriptions approximate $F(t)$ by explicitly considering the dynamics of pairs of lattice sites [14]. In the CMF description,

$$F(t) = \frac{\rho^{(2)}(\sigma_i,\sigma_{i+1})}{\rho^{(1)}(\sigma_i)\rho^{(1)}(\sigma_{i+1})},$$

where $\rho^{(2)}(\sigma_i,\sigma_{i+1})$ is the probability that both sites i and $i+1$ are occupied and $\rho^{(1)}(\sigma_i)$ is the probability that site i is occupied [14]. It is possible to reconstruct $F(t)$ from the C&G description as we can express the probabilities that

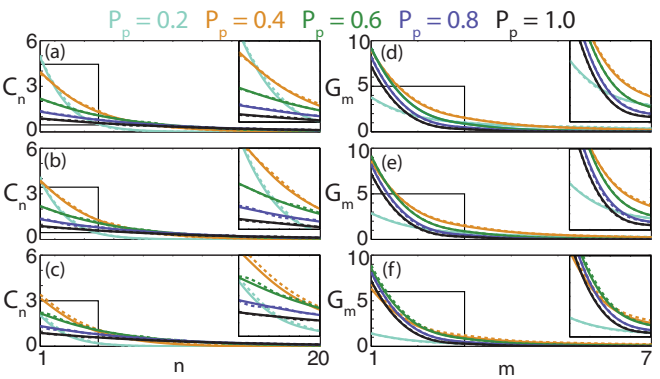

FIG. 3. Comparison between averaged discrete (dashed) and C&G (solid) results for (a)–(c) C_n and (d)–(f) G_m for a suite of P_p values at $t = 50$. Initially, we have (a),(d) 50, (b),(e) 25, and (c),(f) 10 uniformly distributed occupied sites. For all results, $P_m = P_d = 0.1$, $\alpha = 0$, $M = 1000$, and $N = 100$. Inset boxes highlight regions of particular interest.

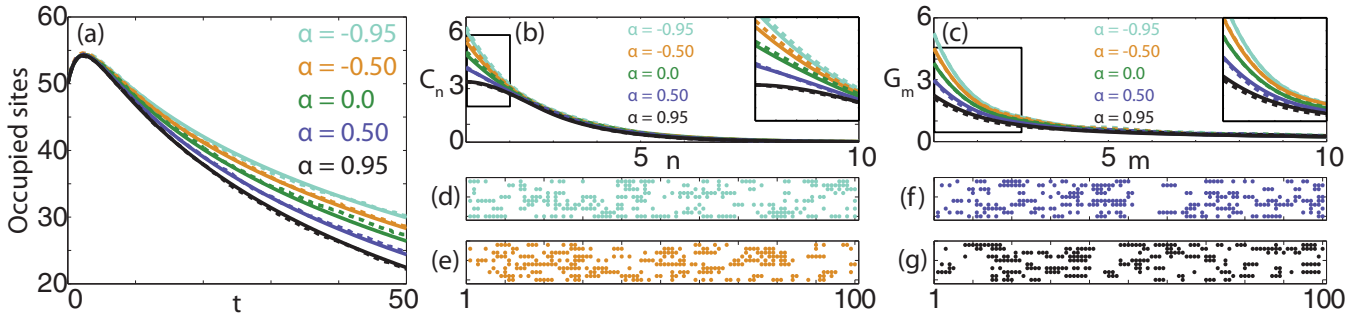

FIG. 4. Comparison between averaged discrete (dashed) and C&G (solid) results for (a) occupied sites, (b) C_n, and (c) G_m for a range of adhesion and repulsion values. (d)–(g) Ten snapshots of the discrete model for (d) $\alpha = -0.95$, (e) $\alpha = -0.5$, (f) $\alpha = 0.5$, and (g) $\alpha = 0.95$ at $t = 50$. Initially, we have 50 uniformly distributed occupied sites. For all results, $P_m = P_d = 0.1$, $P_p = 0.2$, $M = 1000$, and $N = 100$. Inset boxes highlight regions of particular interest.

certain sites are occupied. Since we consider problems with translationally invariant initial conditions, the probability that a site is occupied is independent of position, and hence

$$\rho^{(1)}(\sigma) = \frac{1}{N}\sum_{i=1}^{N} iC_i.$$

The probability that two nearest-neighbor sites are occupied follows a similar argument. Every chain of length $2 \leqslant n \leqslant N-1$ consists of $n-1$ pairs of occupied nearest-neighbor sites, while a chain of length N consists of N pairs of occupied nearest-neighbor sites. As our domain is periodic, there are N possible pairs of nearest-neighbor sites. The probability that two nearest-neighbor sites are occupied is therefore

$$\rho^{(2)}(\sigma_i,\sigma_{i+1}) = \frac{1}{N}\left[\sum_{j=2}^{N-1}(j-1)C_j + NC_N\right].$$

In Fig. 5, we demonstrate that the CMF description fails to describe the nearest-neighbor correlation present in the discrete model, whereas the C&G description correctly predicts the dynamics of the nearest-neighbor correlation. Furthermore, we demonstrate that both the CMF and C&G estimates of the nearest-neighbor correlation are significantly different from the traditional implicit assumption that $F(t) \equiv 1$. The CMF description relies on an approximation to obtain a closed system of equations describing the correlations, which may result in inaccurate predictions, whereas the C&G description accounts for all possible chain sizes and reconstructs the correlations from this information.

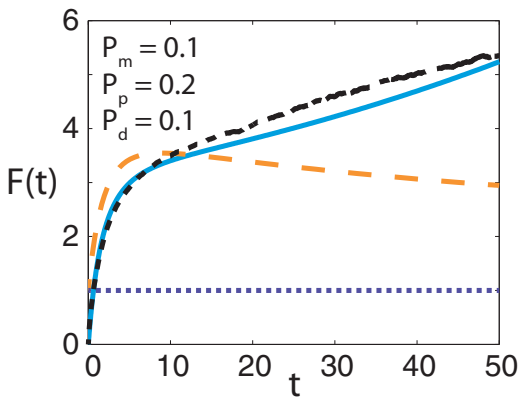

FIG. 5. Comparison of $F(t)$ for the averaged discrete model (black, dashed) and the MF (purple, dotted), CMF (orange, long dashed), and C&G (cyan, solid) descriptions. Here $P_m = 0.1$, $P_p = 0.2$, $P_d = 0.1$, $\alpha = 0$, $N = 100$, and $M = 10^4$. Initially we have 10 uniformly distributed occupied lattice sites.

C. Steady-state approximation

In Fig. 6(a), we present C_n obtained from the C&G description at late time (Appendix C) on a $\log_{10}$ scale. These plots are approximately linear, implying $C_n \approx \beta^{n-1}C_1, \beta > 0$. At steady state, the net birth rate is zero and $P_p(2N_S + N_E)/2 = P_d(N_S + N_E + N_M)$. If we express N_S, N_E, and N_M in terms of C_n and make a power series approximation in terms of β, we obtain $\beta = 1 - P_d/P_p$. For the power series to converge, we require $|\beta| < 1$. As expected, this implies that nontrivial steady states exist only for $P_p > P_d$. Full details of this argument are presented in Appendix C. In Fig. 6(b), we compare the averaged discrete model at late time and the steady-state approximation, and we observe that the approximation is extremely accurate.

VI. CONCLUSIONS

We present a powerful approach to describe lattice-based ABDM processes in terms of groups of contiguous occupied

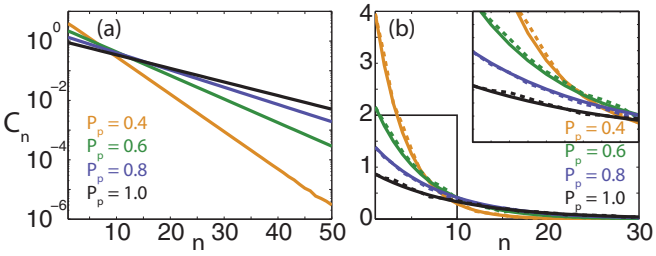

FIG. 6. (a) C_n, calculated from the C&G description, on a $\log_{10}$ scale at steady state for different P_p values. (b) Steady-state approximation for C_n (solid) compared with the averaged discrete results (dashed). For all results, $P_m = P_d = 0.1$, $\alpha = 0$, $M = 1000$, and $N = 100$. Inset box highlights a region of particular interest.

and vacant sites. Our C&G description provides more accurate predictions than either traditional or corrected mean-field descriptions, and, unlike those descriptions, it does not require that $P_p/P_m \ll 1$ and $P_d/P_m \ll 1$ to give accurate predictions [14]. Additionally, our description provides predictions about the clustering present in the system, unlike previous continuum descriptions, and these predictions match the averaged results from the discrete model extremely well, even with significant adhesion or repulsion. Furthermore, we derive a simple analytic approximation of the spatial clustering of occupied sites when the system has reached steady state, and we demonstrate that this approximation is accurate.

We could extend the description presented here in several ways. One extension could be to introduce spatial dependence [11], which would allow us to simulate processes such as scratch assays. Scratch assays are a common experimental procedure used to study collective cell behavior, where the initial distribution of cells depends on spatial location [2]. This extension would require $O(N^2)$ ODEs, rather than the $O(N)$ ODEs in the description presented here, to describe the location and length of the chains and gaps, and we leave the description of these processes for future work. Alternatively, we could introduce a chemical species and couple the ABDM parameters to the chemical concentration to model cellular nutrient uptake [30]. As the C&G description describes the dynamics of every possible chain and gap, it would be relatively straightforward to consider nonconstant parameters (P_m, P_p, P_d, and α) that change with, for example, time, chain, or gap length. Another extension would be to derive and analyze the partial differential equation approximation of the C&G description, similar to the work carried out by Markham *et al.* [31]. This extension would require a Taylor series expansion of C_n and G_m in an appropriate limit, such that n and m can be treated as continuous. We leave this extension for future analysis.

The calibration of traditional continuum descriptions to experimental ABDM data to obtain parameter estimates results in incorrect estimates when spatial correlations become significant [32]. Calibrating stochastic models to experimental data is significantly more computationally expensive [33], compared to dealing with continuum approximations. Non-Bayesian data calibration techniques have also been proposed [2]. However, these approaches can also be computationally expensive as they require the calculation of average discrete behavior across a potentially large parameter space. It would therefore be of interest to investigate the accuracy of parameter estimates obtained from calibrating the C&G description to experimental data in the future.

ACKNOWLEDGMENTS

This work is supported by the Australian Research Council (FT130100148). We appreciate the assistance of Sean McElwain, and the support provided by the High Performance Computing Centre at QUT. We also thank the referees for their helpful comments.

APPENDIX A: CHAIN-AND-GAP EQUATIONS

If we consider all possible events that change the number of chains and gaps of length $n \in [1,N]$, where N is the number of lattice sites, and the rate at which each event occurs, we obtain the following system of equations.

1. Chains

$$\frac{dC_1}{dt} = P_m\left[(1-\alpha)C_2 + \sum_{i=2}^{N-2}\{(1-\alpha)C_i\} - \frac{1}{2}\left\{Y(C_1,G_1,C_1) + \sum_{j=1}^{N-3}[Y(C_1,G_1,C_j)]\right\} - \frac{1-\alpha}{2}\sum_{i=2}^{N-3}\{Y(C_i,G_1,C_1) + \sum_{j=1}^{N-i-2}[Y(C_i,G_1,C_j)]\}\right] + P_p\left[-C_1 - \frac{1}{2}\sum_{i=1}^{N-3}\{Y(C_i,G_1,C_1)\}\right] + P_d\left[-C_1 + 2\sum_{i=2}^{N-1}\{C_i\}\right],$$

$$\frac{dC_k}{dt} = P_m\left[(1-\alpha)(C_{k+1}-C_k) + \frac{1}{2}\{Y(C_1,G_1,C_{k-1}) - Y(C_1,G_1,C_k)\} + \frac{1-\alpha}{2}\sum_{i=2}^{N-k-2}\{Y(C_i,G_1,C_{k-1}) - Y(C_i,G_1,C_k)\} + \frac{1-\alpha}{2}Y(C_{N-k-1},G_1,C_{k-1})\right] + P_p\left[C_{k-1} - C_k - \frac{1}{2}\sum_{i=1}^{N-k-1}\{Y(C_{k-1},G_1,C_i)\} - \frac{1}{2}\sum_{i=1}^{N-k-2}\{Y(C_i,G_1,C_k)\} + \frac{1}{2}\sum_{i=1}^{k-2}\{Y(C_i,G_1,C_{k-i-1})\}\right] + P_d\left[-kC_k + 2\sum_{i=k+1}^{N-1}\{C_i\}\right] \text{ for } k = 2,\ldots,N-3,$$

$$\frac{dC_{N-2}}{dt} = P_m\left[\frac{1}{2}Y(C_1,G_1,C_{N-3}) - (1-\alpha)C_{N-2}\right]$$
$$+ P_p\left[C_{N-3} - C_{N-2} - \frac{1}{2}Y(C_{N-3},G_1,C_1) + \frac{1}{2}\sum_{i=1}^{N-4}\{Y(C_i,G_1,C_{N-i-3})\}\right] + P_d[-(N-2)C_{N-2} + 2C_{N-1}],$$
$$\frac{dC_{N-1}}{dt} = P_p\left[C_{N-2} - C_{N-1} + \frac{1}{2}\sum_{i=1}^{N-3}Y(C_i,G_1,C_{N-i-2})\right] + P_d[-(N-1)C_{N-1} + NC_N],$$
$$\frac{dC_N}{dt} = P_p[C_{N-1}] + P_d[-NC_N],$$

where

$$Y(C_i,G_1) = C_i\left[P(G_1|C_i) + \sum_{j=1}^{N-i-2}P(G_j|C_i)P(G_i|C_i,G_j)\right]$$

represents the number of chains of length i that are next to a gap of length 1, and

$$Y(C_i,G_1,C_j) = C_i\left[P(G_1|C_i)P(C_j|C_i,G_1) + \sum_{k=1}^{N-i-j-1}P(G_k|C_i)P(G_1|C_i,G_k)P(C_j|C_i,G_k,G_1)\right]$$

represents the number of chains of length i that are separated from a chain of length j by a chain of length 1;

$$P(G_j|C_i) = \frac{G_j}{\sum_m G_m},\ m \in A_G(C_i)$$

represents the probability that, provided we have selected a chain of length i, there is a neighboring gap of length j, where

$$A_G(C_i) = \{m \in \mathbb{Z} | 1 \leqslant m \leqslant N-j-2\} \cup \{N-j\}$$

are the possible gap lengths, given that we have selected a chain of length i;

$$P(G_k|C_i,G_j) = \frac{G_k}{\sum_m G_m},\ m \in A_G(C_i,G_j)$$

represents the probability that, provided we have selected a chain of length i with a neighboring gap of length j, there is a neighboring gap of length k, where

$$A_G(C_i,G_j) = \{m \in \mathbb{Z} | 1 \leqslant m \leqslant N-i-j-1\}$$

are the possible gaps, given that we have selected a chain of length i with a neighboring gap of length j;

$$P(C_k|C_i,G_j) = \frac{C_k}{\sum_m C_m},\ m \in A_C(C_i,G_j)$$

represents the probability that, provided we have selected a chain of length i with a neighboring gap of length j, there is a neighboring chain of length k, where

$$A_C(C_i,G_j) = \{m \in \mathbb{Z} | 1 \leqslant m \leqslant N-i-j-1\}$$

are the possible chains, given that we have selected a chain of length i with a neighboring gap of length j;

$$P(C_l|C_i,G_j,G_k) = \frac{C_l}{\sum_m C_m},\ m \in A_C(C_i,G_j,G_k)$$

represents the probability that, provided we have selected a chain of length i with neighboring gaps of length j and k, there is a neighboring chain of length l, where

$$A_C(C_i,G_j,G_k) = \{m \in \mathbb{Z} | 1 \leqslant m \leqslant N-i-j-k-2\} \cup \{N-i-j-k\}$$

are the possible chains, given that we have selected a chain of length i with neighboring gaps of length j and k. We note that the denominator in the probability expressions can be zero. However, this occurs if and only if the numerator is zero and, in terms of probability, is intuitive to interpret as zero.

2. Gaps

$$
\begin{aligned}
\frac{dG_1}{dt} &= P_m\Bigg[-\frac{1}{2}\sum_{i=1}^{N-3} Y(C_1,G_1,C_i) + (1-\alpha)\sum_{i=2}^{N-2}\{C_i\} - \frac{1}{2}\sum_{i=1}^{N-3}\{Y(C_1,G_1,C_i)\} - \frac{1-\alpha}{2}\sum_{i=2}^{N-3}\sum_{j=1}^{N-i-2}\{Y(C_i,G_1,C_j)\} \\
&\quad + \frac{1}{2}\sum_{i=1}^{N-4}\{Y(C_1,G_2,C_i)\} + \frac{1-\alpha}{2}\sum_{i=2}^{N-4}\sum_{j=1}^{N-i-3}\{Y(C_i,G_2,C_j)\} + (1-\alpha)Y(C_{N-2},G_2)\Bigg] + P_p\,[G_2 - G_1] \\
&\quad + P_d\Bigg[NC_N + \sum_{i=3}^{N-1}\{(i-2)C_i\} - \sum_{i=2}^{N-3}\sum_{j=1}^{N-i-2}\{Y(C_i,G_1,C_j)\} - 2Y(C_{N-1},G_1) - \sum_{i=1}^{N-3}\{Y(C_1,G_1,C_i)\}\Bigg], \\
\frac{dG_k}{dt} &= P_m g\Bigg[\frac{1}{2}\sum_{i=1}^{N-k-1}\{Y(C_1,G_{k-1},C_i)\} - \frac{1}{2}\sum_{i=1}^{N-k-2}\{Y(C_1,G_k,C_i)\} - \frac{1}{2}\sum_{i=1}^{N-k-2}\{Y(C_1,G_k,C_i)\} \\
&\quad - \frac{1-\alpha}{2}\sum_{i=2}^{N-k-2}\sum_{j=1}^{N-k-i-1}\{Y(C_i,G_k,C_j)\} + \sum_{i=1}^{N-k-3}\{Y(C_1,G_{k+1},C_i)\} \\
&\quad + \frac{1-\alpha}{2}\sum_{i=2}^{N-k-3}\sum_{j=1}^{N-k-i-2}\{Y(C_i,G_{k+1},C_j)\} + (1-\alpha)[Y(C_{N-k-1},G_{k+1}) - Y(C_{N-k},G_k)]\Bigg] + P_p[G_{k+1} - G_k] \\
&\quad + P_d\Bigg[\sum_{i=2}^{N-k-1}\sum_{j=1}^{N_k-1}\{Y(C_i,G_{k-1},C_j)\} + 2Y(C_{N-k+1},G_{k-1}) - \sum_{i=2}^{N-k-2}\sum_{j=1}^{N-k-i-1}\{Y(C_i,G_k,C_j)\} \\
&\quad - 2Y(C_{N-k},G_k) - \sum_{i=1}^{N-k-1} Y(C_1,G_k,C_i) + \frac{1}{2}\sum_{i=1}^{k-2}\{Y(C_1,G_i,G_{k-i-1})\}\Bigg] \quad \text{for } k = 2,\ldots,N-3, \\
\frac{dG_{N-2}}{dt} &= P_m\Bigg[\frac{1}{2}Y(C_1,G_{N-3},C_1) - (1-\alpha)Y(C_2,G_{N-2})\Bigg] + P_p[G_{N-1} - G_{N-2}] \\
&\quad + P_d\Bigg[2Y(C_3,G_{N-3}) - 2Y(C_2,G_{N-2}) + \frac{1}{2}\sum_{i=1}^{N-4} Y(C_1,G_i,G_{N-i-3})\Bigg], \\
\frac{dG_{N-1}}{dt} &= P_p[-G_{N-1}] + P_d\Bigg[2Y(C_2,G_{N-2}) - G_{N-1} + \frac{1}{2}\sum_{i=1}^{N-3} Y(C_1,G_i,G_{N-i-2})\Bigg], \\
\frac{dG_N}{dt} &= P_d[G_{N-1}],
\end{aligned}
$$

where

$$Y(C_1,G_i,G_j) = C_1[P(G_i|C_1)P(G_j|C_1,G_i) + P(G_j|C_1)P(G_i|C_1,G_j)]$$

represents the number of chains of length 1 next to both a gap of length i and a gap of length j.

APPENDIX B: SMALL BIRTH AND DEATH RATES

Baker and Simpson [14] demonstrate that both MF and CMF descriptions provide predictions that match the average discrete behavior well, provided that $P_p/P_m \ll 1$ and $P_d/P_m \ll 1$. In Fig. 7, we show that the C&G description provides similarly accurate predictions of the average discrete behavior in these parameter regimes.

APPENDIX C: STEADY-STATE APPROXIMATION

We solve the C&G model until late time, that is, where $dC_n/dt \approx 0$ and $dG_n/dt \approx 0$. We note that the time required for $dC_n/dt \approx 0$ and $dG_n/dt \approx 0$ is dependent on the parameters and the initial condition. In practice, we solve the C&G model until the solution is approximately constant with respect to time, rather than algebraically solving for the steady state. For the results presented in Fig. 6, the C&G model was approximately at steady state.

At steady state, the net birth rate must be zero, and hence

$$\frac{P_p}{2}(2N_S + N_E) = P_d(N_S + N_E + N_M).$$

We know that $N_S = C_1$ and that there are two edge agents and $n-2$ middle agents for chains of $n \geqslant 2$, except the

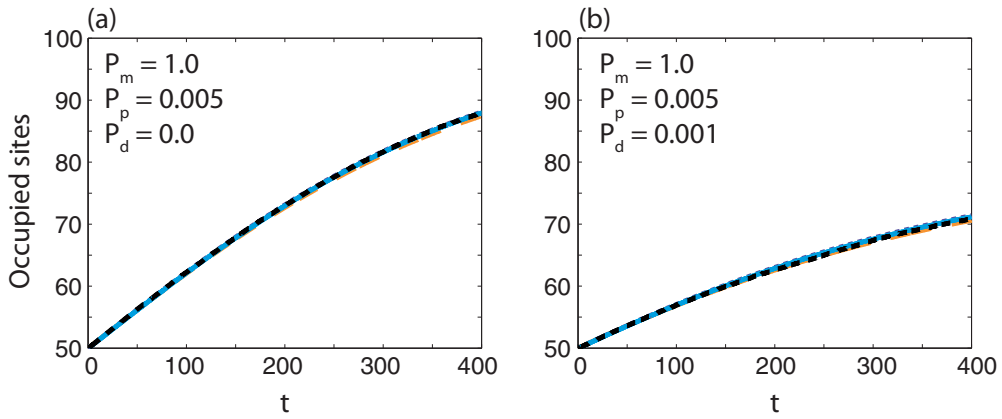

FIG. 7. Comparison between averaged discrete (black, dashed), MF (purple, dotted), CMF (orange, long dashed), and C&G (cyan, solid) results for parameter regimes where $P_p/P_m \ll 1$ and $P_d/P_m \ll 1$. For all results, $M = 1000$, $N = 100$, and $\alpha = 0$. Initially, we have 50 occupied lattice sites.

special case in which $n = N$ and all agents are middle agents. Therefore,

$$N_E = 2\sum_{i=2}^{N-1} C_i,\ N_M = NC_N + \sum_{i=2}^{N-1}(i-2)C_i.$$

If we rewrite the net birth rate in terms of C_n, we obtain

$$\frac{P_p}{2}\left[2C_1 + 2\sum_{i=2}^{N-1} C_i\right] = P_d\left[C_1 + 2\sum_{i=2}^{N-1} C_i + NC_N + \sum_{i=2}^{N-1}(i-2)C_i\right],$$

and, simplifying,

$$P_p\left[\sum_{i=1}^{N-1} C_i\right] = P_d\left[\sum_{i=1}^{N} iC_i\right].$$

Given that C_n is linear on a $\log_{10}$ scale at late time, we make the assumption that $C_n = \beta^{n-1}C_1$, $\beta > 0$, and, subsequently,

$$P_pC_1\left[\sum_{i=1}^{N-1} \beta^{i-1}\right] = P_dC_1\left[\sum_{i=1}^{N} i\beta^{i-1}\right].$$

In the limit $N \to \infty$, both the left- and right-hand sides of the net birth rate equation have a power series representation. As N is finite in our model, we therefore have an approximation, namely

$$\sum_{i=1}^{N-1} \beta^{i-1} \approx \frac{1}{1-\beta}$$

and

$$\sum_{i=1}^{N} i\beta^{i-1} \approx \frac{1}{(1-\beta)^2},$$

which requires $|\beta| < 1$ for convergence. The net birth rate can therefore be expressed as

$$\frac{P_pC_1}{1-\beta} = \frac{P_dC_1}{(1-\beta)^2},$$

which can be rearranged to show that

$$\beta = 1 - \frac{P_d}{P_p}.$$

In Fig. 6(b), we compare this approximation with β calculated from the C&G model when $dC_n/dt \approx 0$ and $dG_m/dt \approx 0$, and we observe that the approximation matches the solution of the C&G model very well.

[1] R. Grima, Curr. Top. Dev. Biol. **81**, 435 (2008).
[2] S. T. Johnston, M. J. Simpson, and D. L. S. McElwain, J. R. Soc. Interface **11**, 20140325 (2014).
[3] J. T. Nardini, D. A. Chapnick, X. Liu, D. Bortz *et al.* (unpublished).
[4] S. Turner and J. A. Sherratt, J. Theor. Biol. **216**, 85 (2002).
[5] S. Turner, J. A. Sherratt, K. J. Painter, and N. J. Savill, Phys. Rev. E **69**, 021910 (2004).
[6] A. D. Mackie, A. Z. Panagiotopoulos, and I. Szleifer, Langmuir **13**, 5022 (1997).
[7] E. A. Codling, M. J. Plank, and S. Benhamou, J. R. Soc. Interface **5**, 813 (2008).
[8] B. Derrida, M. Evans, V. Hakim, and V. Pasquier, J. Phys. A **26**, 1493 (1993).
[9] C. Deroulers, M. Aubert, M. Badoual, and B. Grammaticos, Phys. Rev. E **79**, 031917 (2009).
[10] E. J. Hackett-Jones, K. A. Landman, and K. Fellner, Phys. Rev. E **85**, 041912 (2012).
[11] P. Illien, O. Bénichou, C. Mejía-Monasterio, G. Oshanin, and R. Voituriez, Phys. Rev. Lett. **111**, 038102 (2013).
[12] K. Anguige and C. Schmeiser, J. Math. Biol. **58**, 395 (2009).

[13] A. E. Fernando, K. A. Landman, and M. J. Simpson, Phys. Rev. E **81**, 011903 (2010).
[14] R. E. Baker and M. J. Simpson, Phys. Rev. E **82**, 041905 (2010).
[15] S. T. Johnston, M. J. Simpson, and R. E. Baker, Phys. Rev. E **85**, 051922 (2012).
[16] R. Law and U. Dieckmann, Ecology **81**, 2137 (2000).
[17] A. M. Middleton, C. Fleck, and R. Grima, J. Theor. Biol. **359**, 220 (2014).
[18] K. J. Sharkey, C. Fernandez, K. L. Morgan, E. Peeler, M. Thrush, J. F. Turnbull, and R. G. Bowers, J. Math. Biol. **53**, 61 (2006).
[19] M. J. Simpson and R. E. Baker, Phys. Rev. E **83**, 051922 (2011).
[20] M. Taylor, P. L. Simon, D. M. Green, T. House, and I. Z. Kiss, J. Math. Biol. **64**, 1021 (2012).
[21] D. Chowdhury, A. Schadschneider, and K. Nishinari, Phys. Life Rev. **2**, 318 (2005).
[22] D. T. Gillespie, J. Phys. Chem. **81**, 2340 (1977).
[23] A. Borodin, I. Corwin, and T. Sasamoto, Ann. Probab. **42**, 2314 (2014).
[24] M. J. Simpson, K. A. Landman, and B. D. Hughes, Physica A **389**, 3779 (2010).
[25] K. J. Sharkey, I. Z. Kiss, R. R. Wilkinson, and P. L. Simon, Bull. Math. Biol. **77**, 614 (2015).
[26] R. R. Wilkinson and K. J. Sharkey, Phys. Rev. E **89**, 022808 (2014).
[27] A. Singer, J. Chem. Phys. **121**, 3657 (2004).
[28] K. Zahn, G. Maret, C. Ruß, and H. H. von Grünberg, Phys. Rev. Lett. **91**, 115502 (2003).
[29] W. H. Press, S. A. Teukolsky, W. T. Vetterling, and B. P. Flannery, *Numerical Recipes, 3rd ed.: The Art of Scientific Computing* (Cambridge University Press, Cambridge, 2007).
[30] P. Gerlee and A. R. A. Anderson, J. Theor. Biol. **250**, 705 (2008).
[31] D. C. Markham, M. J. Simpson, and R. E. Baker, Phys. Rev. E **87**, 062702 (2013).
[32] M. J. Simpson, J. A. Sharp, and R. E. Baker, Physica A **395**, 236 (2014).
[33] S. T. Johnston, M. J. Simpson, D. L. S. McElwain, B. J. Binder, and J. V. Ross, Open Biol. **4**, 140097 (2014).