Exclusion processes: short range correlations induced by adhesion and contact interactions

Gianluca Ascolani, Mathilde Badoual, Christophe Deroulers

Exclusion processes: short range correlations induced by adhesion and contact interactions

Gianluca Ascolani*

CNRS, UMR 8165, IMNC, Univ Paris-Sud, Univ Paris Diderot, F-91405 Orsay, France

Mathilde Badoual† and Christophe Deroulers‡

Univ Paris Diderot, Laboratoire IMNC, UMR 8165 CNRS, Univ Paris-Sud, F-91405 Orsay, France
(Dated: 5 October 2012)

We analyze the out-of-equilibrium behavior of exclusion processes where agents interact with their nearest neighbors, and we study the short-range correlations which develop because of the exclusion and other contact interactions. The form of interactions we focus on, including adhesion and contact-preserving interactions, is especially relevant for migration processes of living cells. We show the local agent density and nearest-neighbor two-point correlations resulting from simulations on two dimensional lattices in the transient regime where agents invade an initially empty space from a source and in the stationary regime between a source and a sink. We compare the results of simulations with the corresponding quantities derived from the master equation of the exclusion processes, and in both cases, we show that, during the invasion of space by agents, a wave of correlations travels with velocity $v(t) \sim t^{-1/2}$. The relative placement of this wave to the agent density front and the time dependence of its height may be used to discriminate between different forms of contact interactions or to quantitatively estimate the intensity of interactions. We discuss, in the stationary density profile between a full and an empty reservoir of agents, the presence of a discontinuity close to the empty reservoir. Then, we develop a method for deriving approximate hydrodynamic limits of the processes. From the resulting systems of partial differential equations, we recover the self-similar behavior of the agent density and correlations during space invasion.

I. INTRODUCTION

Since their introduction in the 1940s [1], cellular automata have become an essential tool to study collective behavior in complex systems starting from the individual level in many areas of science: fluid dynamics [2–4], reaction-diffusion problems [5], dynamics of glasses [6], epidemiology [7], traffic flow [8–10]. In the recent years, there have been many applications in biology, both at the intracellular and tissue levels [11–13].

Special cases of cellular automata are the exclusion processes which have been successfully applied to study motility problems when the concentration of agents is such that the geometric hindrance which they impose on each other cannot be neglected [14, 15]. In the exclusion processes, each lattice site is occupied by at most one agent, so that steric effects (hard-core repulsion) are incorporated from the very beginning and lead to nontrivial effects even in absence of other interactions [16, 17]. Ulterior interactions between agents may be added to study intermolecular [18], intercellular or interindividual relationship [9, 19].

In practice, these interactions are specified using *rules*, that is, expressions for the time rate of jump of each agent to another (empty) lattice site as a function of the present content of each lattice site. All what is needed to define the model is the lattice geometry, not necessarily regular, and the list of rules.

Although cellular automata were designed to be efficiently simulated on a computer, it is helpful to supplement them with a macroscopic description of the collective behavior of agents, which often takes the form of a partial differential equation (PDE). Even in cases where simulations are available (usually repeated a large number of times to take into account stochastic noise [15, 20]), the PDE may be a compact way to give an overview of the large scale behavior of the system, to distinguish the universal features of a family of related exclusion processes in the spirit of a RG-like approach [21], or to classify them [19, 22].

If one wants to study the effects of variations of the model parameters, solving PDEs is faster than performing stochastic simulations, and PDEs are also useful to retrieve analytic results. In cases where simulations of the exclusion process is intractable because the number of agents in a realistic system is too large (the human body contains $\simeq 10^{14}$ cells, the human brain $\simeq 3.10^{11}$), or because the exclusion process has to be embedded in a larger system in a multiscale approach, this macroscopic description is essential.

The proof of existence of a PDE describing an exclusion process has been the subject of quite involved mathematical developments [23–29]. Usually, to get an explicit expression for the PDE, one uses a simple approximate technique based on the Chapman-Enskog expansion [30].

* ascolani@imnc.in2p3.fr
† badoual@imnc.in2p3.fr
‡ deroulers@imnc.in2p3.fr

First, a system of coupled ordinary differential equations (ODEs) for the average number of agents in each site of the lattice, or equivalently the occupation probability or density of each site of the lattice, is derived from the rules which define the exclusion process (possibly through the use of the so-called master equation [31]). Because of agent interactions and exclusion, each equation generally involves joint probabilities or correlation functions, like the probability that two nearest-neighbor sites are both occupied at the same time. Then, a mean-field-like approximation is made to express these correlation functions as product of site occupation probabilities, as if the occupations of two sites were statistically independent.

Finally, assuming that the occupation probability of each site is a regular function of the position when the lattice step tends to zero, which amounts to say that the typical length scale over which this probability varies is much longer than one lattice step, a Taylor expansion of this function is substituted into the system of ODEs. Truncating the result to lowest non-vanishing order, one is left with a PDE for the density of agents as a function of space and time. For completeness, let us mention that such a derivation of a macroscopic model from a discrete, microscopic model can be done in many other settings, e.g., to quote a few that were used in biological modeling, the cellular Potts model [32–34], cellular automata on a disordered lattice [35], lattice-gas cellular automata [36], and discrete models with forces [37].

In the case of exclusion processes where agents can only jump to nearest neighboring sites and where the rules involve only a short-range interaction between agents [38], the PDE often takes the form of a nonlinear diffusion equation, and the diffusivity depends on the local density [26–28, 39–41]. The interested reader will find explicit expressions of the nonlinear diffusivity for a large number of such exclusion processes in two recent works by Fernando *et al.* [19] and by Penington *et al.* [42].

This simple, mean-field-like approximation works remarkably well when the large-scale behavior of the system is of diffusive nature, i.e., when the nonlinear diffusion coefficient is positive for all local densities of agents. On the contrary, a negative value of the diffusivity is the sign that the microscopic dynamics tend to form aggregates or is subject to demixion. In that case, the average occupation number of sites varies on the length scale of one or a few lattice steps, hence the hypothesis of regularity of the density as a function of position when the lattice step vanishes is inconsistent, this function cannot satisfy a PDE, and the approach breaks down.

However, even when the diffusivity is always positive, the agreement between the density of agents predicted from the PDE and the density obtained through an average over many simulations of the exclusion process is not perfect [43, 44], and this has been proven to be due to correlations between the occupations of neighboring sites [43]. Usually, the next step beyond the mean-field approximation is the so-called pair approximation, when one keeps track, at the macroscopic level, of both the

probability of occupation of each site and of joint probabilities that two sites are occupied at the same time (hereafter called two-point correlation functions). This has been used, in the case of macroscopically spatially uniform systems, in condensed matter physics [45], to study random walks [46], reaction-diffusion problems [47], epidemic models [48], ecology [49, 50], and multicellular systems [51]. Recently, Simpson and Baker extended that approach to one-dimensional systems that are macroscopically non uniform, as during an invasion process. They found an excellent agreement of the macroscopic model with stochastic simulations using two-point correlation functions of site distances by up to two lattice steps [52] or ten lattice steps [53].

In this work, we use a similar approach to study spatially non-uniform systems in two dimensions, keeping the macroscopic model simple by using only nearest-neighbor two-point correlation functions, and casting it in the form of coupled PDE. While the model is simple, its results agree much better with stochastic simulations than a PDE for the density alone. Its mathematical expression as PDE allows us to analyze the self-similarity (or scaling) properties of its solutions in the context of invasion of space from a source of cells, as in wound-healing-like [54] or migration assay [55] experiments. We show that the scaling properties of correlations may help to distinguish between several microscopic mechanisms not only in theory, but also in experiments.

We have in mind applications to living cell migration processes, which are essential in a number of biological contexts like development, repair, tumor and cancer progression; therefore, we restrict ourselves to the family of exclusion processes where the rate of movement of one cell depends only on the present contacts before moving (which may be preserved or lost) and not on the future contacts (contacts with cells which will be nearest-neighbor only after the move) — "direction then interactions" in the terminology of [19]. This is a realistic setting to study contact interactions (adhesion [20, 44, 54, 56, 57] and cell-cell communication phenomena [43, 55, 58, 59]), disregarding, e.g., chemotaxis or quorum sensing. But the method can be extended to more general exclusion processes. A potential application is personalized treatments of invasive tumors such as glioma [60], where computer simulations of a mathematical model fed by patient-specific parameters will help providing the best therapeutic strategy, guide surgical resection, radiotherapy or chemoterapy, and so on. There, giving an accurate prediction of the amount of infiltrated cells in each part of the brain will be essential.

This family of exclusion processes is introduced in Section II. In Section III, we derive the usual mean-field macroscopic approximation of them, as well as our improved macroscopic models. We compare them to stochastic simulations of the exclusion processes. In Section IV, we go to the continuous space limit and study the self-similarity behavior of the solutions. Finally, we give a discussion and conclusions in Section V.

II. THE MODEL

In our exclusion processes [15, 61], agents move on a fixed d-dimensional lattice, each site of which may contain 0 or 1 agent. For simplicity, we will apply our results only to the bidimensional hexagonal tiling (triangular lattice), but they can be extended straightforwardly. In a move, an agent can only jump to an empty nearest neighbor site. Of course, the reality of biological movement is much more complicated than this (for instance, cells deform, make protrusions etc.), but the aim is to gain access to the macroscopic collective behavior for which we believe that a too detailed description may be irrelevant because many microscopic degrees of freedom will be "forgotten" at large scales and will make simulations and computations very difficult if possible at all.

A. Jump rates and interactions

The definition of a process is completed with the specification of the rate (probability of occurrence per time unit) of each jump of an agent, which is assumed not to depend explicitly on time or position, but only of the content of the lattice sites.

Let i and j be two lattice sites. We denote $\mathcal{V}(i)$ — for *vicinity* — the set of the nearest-neighbors of i on the lattice, $V(i)$ the numbers of these sites, and $v(i)$ the number of sites among them which are occupied ($0 \leq v(i) \leq V(i)$, $V(i) = V = 6$ on the hexagonal tiling; $v(i)$ may vary with time, but not $V(i)$). Likewise, $\mathcal{M}(i,j)$ is the set of lattice sites which are nearest neighbors of both i and j (but distinct from i and j) — $\mathcal{M}$ for *maintained* contact —, $M(i,j)$ their numbers (2 on the hexagonal tiling), $m(i,j)$ the number of full sites between them, $\mathcal{N}(i,j)$ is the set of nearest-neighboring sites of i which are neither j nor a nearest-neighbor of j — $\mathcal{N}$ for *not-maintained* —, $N(i,j)$ their number and $n(i,j)$ the number of full sites between them, Fig. 1.

For simplicity, we study only some of the processes where the rate of any jump, say from site i to site j, depends only on $V(i)$, $m(i,j)$, $n(i,j)$, $M(i,j)$ and $N(i,j)$, but our approach can be extended. We consider the simplest situation, where there is no other interaction than exclusion (hard-core repulsion), to be the case where each agent has the same probability to attempt a jump in a given time interval, irrespective of the actual occupancy state of the surrounding sites; as a consequence, the rate of jumping of an agent is proportional to the inverse of the number of possible jumps it can do (number of empty surrounding sites). Moreover, all possible jumps in the lattice have the same probability to occur, and at large scales, the occupation probability obeys a linear diffusion equation while the mean quadratic distance of each agent to its departure point grows linearly with time, as if agents would not interact at all and do a simple random walk [14, 15]. In order to facilitate the comparison with results obtained on different lattices, we denote
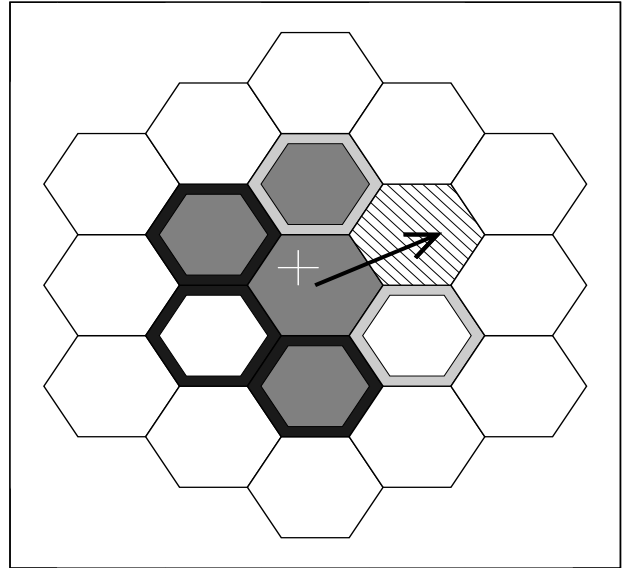

FIG. 1. Analytical computations and numerical simulations are done in a hexagonal tiling. The distance between two nearest neighbor sites is λ. Gray filled sites are occupied by cells; empty sites are filled in white. The cell marked with a white cross symbol attempts to move in the direction of the arrow, if the hatched site is empty. The neighbor sites $\mathcal{N}$ and $\mathcal{M}$ of the marked cell have black and light gray border respectively. Cells in the sites with a black border form non-maintained links, which break during the jump of the marked cell. Cells in the common neighbor sites between the marked cell and the hatched site form maintained links preserved during the jump of the marked cell.

the rate of jump from site i to site j as $T_{i,j}/V(i)$ with, by choice, $T_{i,j} = 1$ when there is no other interaction than exclusion. The processes we consider (special forms of $T_{i,j}$ due to interactions) are listed below, along with their biological motivation.

Adhesion model. To study the influence of cell-cell adhesion on cellular migration, Khain *et al.* introduced a model [56] where

$$T_{i,j} = (1-q)^{m(i,j)+n(i,j)}, \qquad (1)$$

$q \in [0,1]$ being a constant parameter to quantify the strength of adhesion (from 0, no adhesion, like in standard Simple Symmetric Exclusion Processes (SSEP) to 1, impossible movement). It is assumed that adhesion is instantaneously gained or lost, i.e., that the time scale of a possible dynamics of adhesion, like the recruitment of proteins to build up or strengthen focal adhesions, is much shorter than the time scale of migration.

Gap junctional model. To explain experimental results about the influence of gap junction communications between cells on the migration of some tumoral astrocytes [55, 58], Aubert *et al.* introduced an exclusion process where

$$T_{i,j} = 1 - p + (2p - 1)\min[m(i,j),1]. \qquad (2)$$

The parameter $p \in [0, 1]$, assumed constant and common to all cells, allows to interpolate between SSEP for $p = 1/2$, maximal effect of gap junctions for $p = 1$, where no cell will move unless it has a neighbor cell and will keep contact with it, and $p = 0$, where, to the contrary, no cell will move if it is not able to break all existing contacts (this case is probably of little relevance to biology, but was studied in detail in the context of the glassy dynamics [62–65]). Gap junctions are short channels passing through two touching cell membranes, which enable the passage of small molecules and ions. They are formed by two connected hemi-channels, one on each cell membrane. Different types of gap junctions arise from various genic expressions, and a gap junction can be functional or closed. The parameter p should be a way to take into account such variability. Here again, it is assumed that the time scale of establishing effective gap junctions is much shorter than the time scale of migration and, additionally, that gap junctions are formed and open with any nearest neighbor cells and that their number has no influence (i.e., maintaining communication with one cell has the same effect as maintaining it with many).

Linear model. To gain an overview, we finally introduce a more general model, where

$$T_{i,j} = \alpha + \beta\, m(i,j) + \gamma\, n(i,j), \qquad (3)$$

$\alpha \geq 0$, β, and $\gamma \leq 0$ being constant real numbers, Fig. 1. We choose a linear expression for $T_{i,j}$ to keep computations simple: the purpose of this model is not to be faithful to experiments, but to be illustrative. Choosing $\gamma = 0$ and $\beta > 0$, one gets a behavior similar to the gap-junctional model (the jump rate being now dependent on the number of cells with which contact is maintained). Choosing $\beta = \gamma < 0$, one gets a behavior similar to the adhesion model (for small q, $\alpha = 1$ and $\beta = \gamma = -q$, the behavior should be quantitatively the same as for the adhesion model).

B. Choice of boundary conditions

The models previously proposed can be studied in different geometries, that is different boundary conditions and different numbers and disposition of cell sources and cell sinks, to address various aspects of the exclusion processes such as: relaxation to the equilibrium, steady state analysis, approach to the steady state, or perennial out-of-equilibrium conditions. One of the geometrical setups we will focus on is directly inspired by the set-up of the cancer cell migration process experiment discussed in [43, 55, 58]. The experiments consist in placing an aggregate of cancer cells (a so-called spheroid) in a Petri dish with an agar substrate containing suitable nutritional needs for the sustainment of the cells. Initially piled in the spheroid, the cells slowly exit it and start to migrate in the outside region where they avoid overlapping. In the same way, we study the evolution of the system starting from a completely empty initial condition except for a small central region where all the cells are placed. In the proposed geometries, the spheroid is represented as a source of cells which can never empty out and where no empty sites are allowed at any time. When a cell leaves the source to enter the system, the free tile in the source, previously occupied by the cell, is immediately filled up with a new cell.

To avoid dealing with infinite lattices, we decided to add a sink region. It is an empty reservoir where cells are taken away from the system and act as if they were driven by a strong apoptotic signal putting them to death with an infinite apoptosis rate. Therefore, any cell entering the sink is destroyed. When the sink or the source extend over a set of contiguous sites that create a closed path, they become borders separating the space into independent subregions. In this work, for simplicity, we consider geometries where there is only one region of interest, always enclosed between a source and a sink.

Although cells interact only with their nearest neighbors and travel during each elementary jump a length λ equal to the distance between two neighbor cells, the boundaries may have influence on a long distance because of the exclusion rule. Actually, starting from an initially empty lattice, except for the source, there will be different time regimes: first a non-stationary period of time, when the population of cells invades the free space and the sink has not yet been reached, then an out-of-equilibrium steady state with a constant current of cells from the source to the sink (up to stochastic fluctuations). The first period of time is relevant for the migration experiments on Petri dishes.

Given the set of rules specified in the previous section, we analyze the dynamic evolution toward the steady state in two types of geometrical configurations. One on a cylindrical surface which has more theoretical advantages, and one on a planar bounded surface to mimic more closely in-vitro experiments of cells migration in culture dishes, see Fig. 2. In the cylindrical configuration, the geometry consists of a two dimensional rectangular space with regular hexagonal tiling having periodic boundary condition along one direction and the other two sides in contact with two reservoirs: a source and a sink respectively, Fig. 2(a). This geometry is particularly convenient. On one hand, it is invariant under translations along the direction with periodic boundary conditions. Thus, it makes it possible to describe the properties of the two dimensional system such as density and correlations in terms of just one single spatial variable: for example, the distance from one reservoir. On the other hand, this "quasi-1D" geometry does not generate a loss of generality. More complex geometries can be equally addressed with the techniques of this work, and they do not introduce further physical effects. At least at large times, the results are fundamentally the same up to some numerical changes (see Sec. (III D) for results on a different geometry and Sec. (IV C) for a justification). In the cylindrical geometry, the shape of the
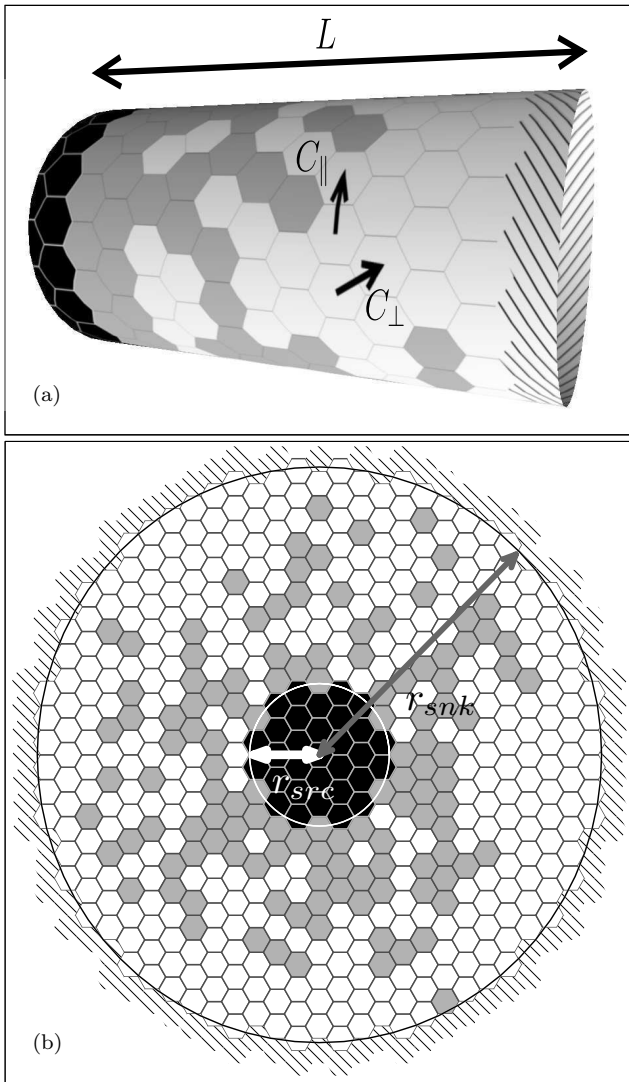
FIG. 2. Geometrical disposition of the reservoirs. The black sites represent the cell source, the hatched sites are the empty reservoir, the empty sites are white and the cells are in gray. (a) cylindrical geometry of length L with two small arrows connecting neighbor sites showing the two main directions of the correlation. (b) radial geometry with two concentric circumferences showing the borders of the source and the sink with radii r_{src} in white and r_{snk} in gray respectively.

reservoirs are fixed and the degrees of freedom are the two sizes of the cylindrical surface: the length of the circumference and the distance between the reservoirs. From numerical simulations, we have seen that the results are independent of the length of cylindrical circumference, if this is 10 or more lattice tiles λ (results not shown). The distance between the reservoirs produces some differences in the steady state, and they will be discussed in the next section. It is important to remark that these considerations hold true when the lattice is oriented as in Fig. 2(a) as well as when it is rotated by 90 degrees. The second geometrical configuration is a planar regular

hexagonal tiled lattice, hereafter called radial geometry. In it, we define an origin O represented by a generic tile and two radii from O: r_{src} and r_{snk}. The tiles with distance from O less than or equal to r_{src} function as source and the tiles with distance from O bigger than r_{snk} function as sink, Fig. 2(b). The only constraint that exists between the two degrees of freedom, r_{src} and r_{snk}, is $r_{src} < r_{snk}$. Nevertheless, numerical results prove that for radii larger than two lattice steps λ, their specific values become weakly influential on the dynamical evolution of the system (results not shown). Also, in this geometrical configuration, the steady state can depend on the distance between the reservoirs.

III. EVOLUTION EQUATIONS ON THE LATTICE

In [60], the authors analyzed similar systems in the mean-field approximation, and in [43], they commented about the discrepancy at the steady state between the numerical and the analytical results for particular values of their interaction parameter close to the sink. On the other hand, here, we address the problem beyond the mean-field approximation to investigate the behaviors of the correlation at short distance among cells defined as:

$$C_2(i,j) = \langle \eta(i)\eta(j) \rangle - \langle \eta(i) \rangle \langle \eta(j) \rangle, \quad (4)$$

where i and j are two nearest neighbor sites, η is the number of cells, and the angular brackets stand for the average over all possible configurations.

In addition, considering the presence of correlations between nearest neighbors results in a more accurate prediction for the density profile in comparison to that obtained in the mean-field approximation. For example, this is evident in the adhesion model when the interactions are sufficiently strong, as shown in Fig. 3.

It is important to stress that the previously stated rules and the results shown in the next sections will hold true not just as a consequence of the particular choice of the hexagonal tiling. Indeed, the phenomenon of a correlation wave with similar behaviors is a much more general result, and it will be present in other kinds of lattices, for example: triangular, square, and random tilings, under the condition of a more general definition of nearest neighbors like, for example, "a tile is a nearest neighbor of another if they have a common vertex".

A. Master equation and evolution equations for the correlation functions — the general case

Let us define a lattice as a partition of a n-dimensional space $\Omega \subseteq \mathbb{R}^n$ in z non-overlapping subsets, each of them representing a tile of the lattice identified by an index $i \in \mathbb{N}$. Let η_i be the number of cells in lattice site i; $\eta_i \in \{0,1\}$ so that $\eta_i^2 = \eta_i$, and the total number of cells in the system at time t is: $Z = \sum_{i=1}^z \eta_i$. The generic
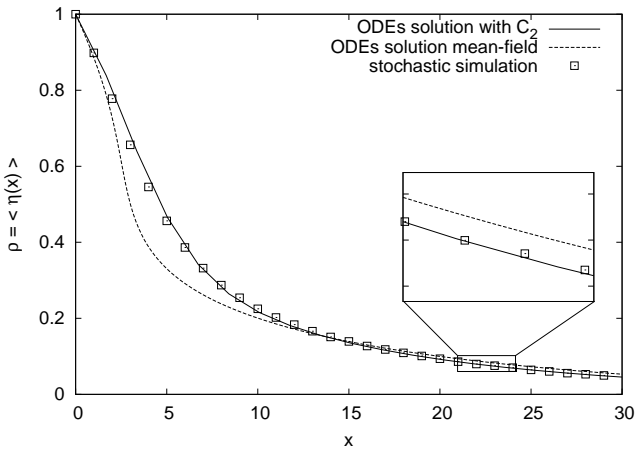
FIG. 3. Density profile for the adhesion model with $q = 0.4$ on a cylindrical geometry after 80 time steps. The dots represent the results of the stochastic simulation of Sec. (III D), and the error bars are smaller than the size of the symbols. The continuous line shows the density profile obtained by considering the two-point nearest neighbor correlation functions as explained in Sec. (III B) and Sec. (III C). The dashed line is the density profile in the mean-field approximation.

configuration of the occupancy states is given by vector $\boldsymbol{\eta} = (\eta_1, \eta_2, \ldots, \eta_z)$ of all the numbers η_i. We denote $P(\boldsymbol{\eta}, t)$ the probability that the the process is in the configuration $\boldsymbol{\eta}$ at time t.

If $\hat{W}_{i,j}$ is the operator that permutes the contents of sites i and j:

$$\hat{W}_{i,j} : (\eta_1, ..., \eta_i..., \eta_j, ..., \eta_z) \rightarrow (\eta_1, ..., \eta_j, ..., \eta_i, ..., \eta_z),$$

we can express the evolution equation for $P(\boldsymbol{\eta}, t)$ (the master equation) as

$$\partial_t P(\boldsymbol{\eta}, t) = \sum_{i=1}^{z} \frac{1}{V(i)} \sum_{j \in \mathcal{V}(i)} \left[\eta_j (1 - \eta_i) T_{j,i} \left(\hat{W}_{i,j} \boldsymbol{\eta} \right) \times \right.$$
$$\left. P \left(\hat{W}_{i,j} \boldsymbol{\eta}, t \right) - \eta_i (1 - \eta_j) T_{i,j}(\boldsymbol{\eta}) P(\boldsymbol{\eta}, t) \right] \quad (5)$$

where the first term inside the brackets represents the flux of probability from the configurations which contribute to fill the site i, so that $\eta_j(1 - \eta_i) T_{j,i} \left(\hat{W}_{i,j} \boldsymbol{\eta} \right) \mathrm{d}t$ is the chance of a cell to move from j to i in the infinitesimal time step $\mathrm{d}t$ provided that the system is in the state $\hat{W}_{i,j} \boldsymbol{\eta}$. Similarly, the second term represents the flux of probability to the configurations where the site i is empty. This expression is valid for any exclusion process; we assume from now on that $T_{i,j}(\boldsymbol{\eta})$ vanishes if i and j are not nearest neighbors on the lattice (i.e., if $j \notin \mathcal{V}(i)$).

We define the average of generic quantity $A(\boldsymbol{\eta})$ at time t as a sum over all possible configurations of the process:

$$\langle A \rangle = \sum_{\boldsymbol{\eta}} A(\boldsymbol{\eta}) P(\boldsymbol{\eta}, t). \quad (6)$$

In particular, we are interested in the average multi-point density, or correlation function, on n distinct lattice sites $l_1, l_2, \ldots, l_n$:

$$\rho_n(l_1, \ldots, l_n, t) = \langle \eta_{l_1} \eta_{l_2} \ldots \eta_{l_n} \rangle \quad (7)$$

— the density $\langle \eta_i \rangle$ of cells on a single lattice site i will simply be written $\rho(i, t)$ — and in the connected correlation function:

$$C_n(l_1, \ldots, l_n, t) = \langle [\eta_{l_1} - \rho(l_1, t)][\eta_{l_2} - \rho(l_2, t)] \ldots$$
$$[\eta_{l_n} - \rho(l_n, t)] \rangle \quad (8)$$

which vanishes if the occupation numbers on the sites are statistically independent.

Inserting the master equation Eq. (5) into the expression of the time derivative of $\rho(i, t)$ yields the evolution equation:

$$\partial_t \rho(i, t) = \frac{1}{V(i)} \left\langle \sum_{j \in \mathcal{V}(i)} [T_{j,i}(\boldsymbol{\eta}) \eta_j (1 - \eta_i) + \quad (9) \right.$$
$$\left. - T_{i,j}(\boldsymbol{\eta}) \eta_i (1 - \eta_j)] \right\rangle. \quad (10)$$

Using the general definition in Eq. (7) for $n = 2$, the equation for the time evolution of the two-point density function at the two generic sites i and j is [66]:

$$\partial_t \rho_2(i, j, t) = \frac{1}{V(i)} \left\langle \sum_{\substack{k \in \mathcal{V}(i) \\ k \neq j}} [T_{k,i}(\boldsymbol{\eta}) \eta_k (1 - \eta_i) + \right.$$
$$\left. - T_{i,k}(\boldsymbol{\eta}) \eta_i (1 - \eta_k)] \eta_j \right\rangle + i \leftrightarrow j, \quad (11)$$

where $i \leftrightarrow j$ is equal to the first term on the right-hand side of the equation above, with the roles of i and j exchanged. In Eq. (11), the constraints $k \neq j$ are added to ensure that all configurations included in the counting after a jump of a cell have both sites i and j occupied. The equations for the two point connected correlation function immediately follow from Eqs. (8, 9, 11):

$$\partial_t C_2(i, j, t) = \frac{1}{V(i)} \left\langle \sum_{\substack{k \in \mathcal{V}(i) \\ k \neq j}} [T_{k,i}(\boldsymbol{\eta}) \eta_k (1 - \eta_i) + \right.$$
$$\left. - T_{i,k}(\boldsymbol{\eta}) \eta_i (1 - \eta_k)] (\eta_j - \rho(j, t)) \right\rangle + i \leftrightarrow j. \quad (12)$$

In the same way, we can express the evolution equation of connected correlation functions for any n.

B. Expressions for our models

Let us give the expression of $T_{i,j}(\boldsymbol{\eta})$ for the models we introduced in Sec. (II), on the hexagonal tiling. To

explicitly compute the averaged quantities of the previous subsection, we need to express $m(i,j)$ and $n(i,j)$ in terms of $\boldsymbol{\eta}$:

$$m(i,j) = \sum_{k \in \mathcal{M}(i,j)} \eta_k, \qquad n(i,j) = \sum_{k \in \mathcal{N}(i,j)} \eta_k. \quad (13)$$

Consequently, the transition rate for the adhesion model in Eq. (1) becomes:

$$T_{i,j}(\boldsymbol{\eta}) = (1-q)^{\eta_k + \eta_l + \eta_r + \eta_s + \eta_w},$$

where k and l are the common neighbors of i and j, and r, s and w are the neighbors of i which are not neighbors of the future position j. In the gap junctional model, when intercellular communication through gap junctions drives the system dynamics, Eq. (2) can be rewritten as:

$$T_{i,j}(\boldsymbol{\eta}) = p(\eta_k + \eta_l - \eta_k \eta_l) + (1-p)(1-\eta_k)(1-\eta_l),$$

where the sites l and k identify the two common neighbors of i and j. In this way, $T_{i,j}$ will be equal to p if both, or just one of the sites k and l are occupied by other cells, and it will be equal to $1-p$, if both k and l are empty. When $\eta_i = 1$, the cell in i will share gap junctional links with all the nearest neighbor sites occupied by other cells, but only the gap junctions in the direction of the site k or l will be maintained functional during the cell transition to the site j. Finally, in the linear model the transition rate Eq. (3) is:

$$T_{i,j}(\boldsymbol{\eta}) = \alpha + \beta(\eta_k + \eta_l) + \gamma(\eta_r + \eta_s + \eta_w),$$

where the indices k, l, r, s and w have the same meaning as in the adhesion model.

For the gap junctional model and for the linear model with $\gamma = 0$, the transition rate is invariant under the permutation of indices i and j: $T_{i,j}(\boldsymbol{\eta}) = T_{j,i}(\boldsymbol{\eta})$. For the adhesion model, it is not invariant because the sites belonging to $\mathcal{N}(i,j)$ in the expression for $T_{i,j}(\boldsymbol{\eta})$ are different from the sites belonging to $\mathcal{N}(j,i)$ in $T_{j,i}(\boldsymbol{\eta})$. When $T_{i,j}(\boldsymbol{\eta})$ is symmetric with respect to the indices i and j, the probability distribution of the configurations at equilibrium is the uniform distribution, so that, at equilibrium, the occupation of one site is statistically independent from the others.

In the presence of sources and sinks, the general evolution equations for the densities and the correlation given in Eqs. (9, 11, 12) no longer hold. Indeed, if one or more points of the multi(single)-point density functions are at one lattice step from, or belong to a reservoir, it is necessary to take into consideration that some changes in the configurations of the positions of cells from, or toward the reservoirs are impossible and must be excluded. For example, when a multi(single)-point density function has a point at distance λ from a sink, no flux of cells coming from the sink contributes to the final configuration. On the other hand, if the reservoir is a source, no cell can enter it and any flux of cells toward the source must be zero.

From Eq. (8), it follows that if one of the η_i is constant, then the correlations including the site i are zero; consequently, it is easy to express the border conditions in terms of the multi-point correlation functions with one or more of its points belonging to a reservoir. It is important to stress that, despite the presence of reservoirs that invalidates the general forms of Eqs. (9, 11, 12), substituting the values of the border condition into the evolution equations of the density and the correlations explicitly derived from Eqs. (9, 12) produces the same correct results as if they were obtained from the evolution equations with the ulterior constraints due to the presence of reservoirs.

Let us consider the case of the linear model and explicitly compute the equations for the density and the two-point connected function in the hexagonal lattice. Using Eqs. (9, 12), and dropping the explicit dependence of the densities on time and unnecessary indices, the system of equations, in the region of interest far from the reservoirs, is:

$$\partial_t \rho(i) = \frac{1}{6} \sum_{k \in \mathcal{V}(i)} \left\{ \alpha[\rho(k) - \rho(i)] + \beta \sum_{s \in \mathcal{M}(i,k)} [C_2(k,s) + \rho(k)\rho(s) - C_2(i,s) - \rho(i)\rho(s)] + \right.$$

$$\left. + \gamma \sum_{s \in \mathcal{N}(i,k)} [\rho(i,k,s) - \rho(i,s)] - \gamma \sum_{s \in \mathcal{N}(k,i)} [\rho(k,i,s) - \rho(k,s)] \right\},$$

$$\partial_t C_2(i,j) = \frac{1}{6} \sum_{\substack{k \in \mathcal{V}(i) \\ k \neq j}} \left\{ \alpha\Big[C_2(k,j) - C_2(i,j)\Big] + \beta \sum_{s \in \mathcal{M}(i,k)} \Big[\rho(j,k,s) - \rho(j)\rho(k,s) - \rho(j,i,s) + \rho(j)\rho(i,s)\Big] + \right.$$

$$+ \gamma \sum_{s \in \mathcal{N}(i,k)} \Big[\rho(j,i,k,s) - \rho(j)\rho(i,k,s) - \rho(j,i,s) + \rho(j)\rho(i,s)\Big] +$$

$$\left. - \gamma \sum_{s \in \mathcal{N}(k,i)} \Big[\rho(j,k,i,s) - \rho(j)\rho(k,i,s) - \rho(j,k,s) + \rho(j)\rho(k,s)\Big] \right\} + i \leftrightarrow j, \tag{14}$$

$$\rho(i,j,k) - \rho(i)\rho(j,k) = C_2(i,j)\rho(k) + C_2(i,k)\rho(j) + C_3(i,j,k), \tag{15}$$

$$\rho(i,j,k,l) - \rho(i)\rho(j,k,l) = C_2(i,j)\rho(k)\rho(l) + C_2(i,k)\rho(j)\rho(l) + C_2(i,l)\rho(j)\rho(k) +$$
$$+ C_3(i,j,k)\rho(l) + C_3(i,j,l)\rho(k) + C_3(i,k,l)\rho(j) + C_4(i,j,k,l), \tag{16}$$

where the Eqs. (15, 16) show how to rewrite the quantity $\rho(i,j,k)$ and $\rho(i,j,k,l)$ for the generic sites i, j, k and l in terms of connected correlation functions.

The border conditions for the system Eq. (14), at any time t, are:

$$\begin{aligned} \rho(l_i) &= 1, & \text{for } l_i \text{ in the source} \\ \rho(l_i) &= 0, & \text{for } l_i \text{ in the sink} \\ C_n(l_1, \ldots, l_n) &= 0, & \text{if any } l_i \text{ is in a reservoir.} \end{aligned} \tag{17}$$

The same set of border conditions Eq. (17) can be applied in both cylindrical and radial geometrical dispositions of source and sink, and in addition, they hold true for all three models proposed.

For a lattice with z tiles, in the system of Eqs. (9, 12), there are z differential equations for the density and even more for the two-point connected correlation function. If we consider only the correlations between nearest neighbor sites, the process is described by a total of $(1 + d)z$ ordinary differential equations plus the border conditions, where $d \times z$ equations describe the time evolution of the two-point connected correlations in the d main lattice directions (for the hexagonal tiling $d = 3$). In the cylindrical geometry, due to the invariance under rotations and reflections along the axial direction, for a site there are only two independent equations for the correlations between nearest neighbor sites: $C_\parallel$ and $C_\perp$, see

Fig. 2(a); therefore, the number of equations reduces to $4L$, where L is the minimum number of sites one has to travel through to go from one end to the other of the cylinder.

C. Closure approximation

To solve the set of equations Eqs. (14, 17) for the linear model, it is necessary to express $C_3(i,j,k,t)$ in terms of known quantities. In more general cases such as the gap junctional and the adhesion models, finding a solution of the system of equations Eqs. (9, 12, 17) requires to already know all the correlations $C_n(l_1, \ldots, l_n)$ with $n > 2$, or to add other equations to the initial system which allows one to determine these unknown quantities. One possible way is to express the multi-point correlation functions with the highest number of points with an approximate expression involving only multi-point correlation functions with less points. This approach is called closure approximation. In comparison to the approach in articles [43, 60], where mean-field approximation was adopted and all the correlations were completely disregarded, we take into consideration the short range two-point connected correlation functions with the purpose to obtain more information about the dynamical evolution

of the system and to improve the agreement of the analytical results with the stochastic simulations. Our closure approximations is the following. All the connected correlations defined on more than two points are set to zero:

$$\left\langle \prod_{k=1}^{n>2} (\eta_{l_k} - \langle \eta_{l_k} \rangle) \right\rangle = 0, \qquad (18)$$

where, as above, $l_1, l_2, \ldots, l_n$ are distinct lattice sites, and all the two-point connected correlation functions between cells at a distance bigger than one lattice step λ are set to zero:

$$\langle (\eta_i - \langle \eta_i \rangle) (\eta_j - \langle \eta_j \rangle) \rangle = 0, \qquad \text{if } j \notin \mathcal{V}(i), j \neq i. \quad (19)$$

Therefore, only the information relative to the nearest cell couples remains, as if all the clusters and structures with more than two points would appear in a completely random way (conditioned to the values of the local density of cells and of the two-point nearest neighbor correlation functions).

These approximations are suggested by the rules of our exclusion processes, in which the movement of one cell is directly influenced only by the presence or absence of nearest neighbors. As we shall see, stochastic simulations show that the approximations on the system Eqs. (9, 12) produced by Eq. (18) are reasonably good for the gap junctional model, the linear model when γ is small, and the adhesion model when q is small (weak adhesive interactions). For instance, simulations of the gap junctional model show that, except cell couples moving together, no particular structure or big cluster appears. Nevertheless, neglecting correlations between sites at distances of two lattice steps and above is not so unquestionable because of both "repulsive" interactions (exclusion) and "attractive" interactions (adhesion). On one hand, it is known that, in exclusion processes with no other interaction than mere exclusion, there are long range correlations [16, 67, 68]. From stochastic simulations it seems that correlations between cells at two and three lattice steps apart are much smaller, but not completely negligible (results not shown). In fact, the time evolution of these correlations and their behaviors resemble those of the connected correlation functions between nearest neighbors. On the other hand, our models with large values of the adhesion parameter (q or γ) exhibit large-scale structures of cells. For instance, the adhesion model can be mapped onto the Ising model [57], and it can be shown that a spontaneous phase separation with clustering of cells happens at larges values of q. In such situations, the success of the analytical approach in reproducing exactly results from the stochastic simulations is lost. But, at least when aggregation is absent or weak, it is still possible to retrieve, from the analytical results for the connected correlation functions, important information regarding the system evolution which is characteristic of each type of interaction (see Sec. (III D) and Sec. (IV)).

The closure approximations, Eqs. (18, 19) can be shortly expressed together with the set of border conditions Eq. (17):

$$\begin{aligned}
\rho(l_i) &= 1, & &\text{for } l_i \text{ in the source} \\
\rho(l_i) &= 0, & &\text{for } l_i \text{ in the sink} \\
C_n(l_1, \ldots, l_n) &= 0, & &\forall\, l_i \text{ in a reservoir or } \forall\, n > 2 \\
C_2(l_1, l_2) &= 0, & &\forall\, l_i \text{ in a reservoir or } l_1 \notin \mathcal{V}(l_2).
\end{aligned}$$
$$(20)$$

Other kinds of closure approximations can be applied in place of Eq. (18). One can systematically extend this approximation by dealing with correlations at larger distances or with more than two sites [51–53]. One can also choose a different scheme, as the Kirkwood Superposition Approximation used in [52, 53]. However, this results in dealing with unbounded connected correlation functions which diverge when all sites are empty or when all are full, and do not satisfy the geometrical and initial conditions proposed here.

D. Stochastic simulations and discrete equations

The stochastic simulations consist of a cellular automaton where cells evolve through a number of discrete time steps moving on a hexagonal tiling and performing a series of interaction dependent moves described by the rules in Sec. (II), [69]. In the framework of cellular automata, the rules and the change of the state of the cells are intended to be applied in parallel (synchronously) [70]; nevertheless, in the proposed exclusion processes, the parallel update scheme entails the problem of two cells jumping at the same time on the same empty site. To avoid any ambiguities in the stochastic simulations, the positions of the cells are asynchronously updated following a random order time scheme [71]: at each time step all the cells are chosen once in a new random order and updated. At equilibrium, for the gap junctional model and linear model with $\gamma = 0$, the symmetry of $T_{i,j} = T_{j,i}$ results in no correlations, see Sec. (III B). On the other hand, simulations with cyclic update scheme (all cells, at each time step, are updated once in the same random pre-fixed order) produce spurious correlations at equilibrium. Therefore, the random ordered scheme (all cells, at each time step, are updated once in a new random order) [72] and the random independent scheme (at each time step, a cell is chosen randomly and updated) [73][74] are more appropriate for the simulations. Also, the clocked random waiting time scheme (each cell is updated following its own independent random clock) [75] should not introduce spurious correlations, if the waiting time distribution has the first two moments finite.

The update of a generic cell in the site i consists in choosing the new site j at random with equal probability between all the nearest neighbors $V(i)$, and then, if the site j is empty, moving the cell in the new position with a probability $Q = T_{i,j}(\boldsymbol{\eta})\Delta t$. To be well defined, the probability Q that a cell jumps during the time interval

Δt requires to fix the simulation time step consistently. Therefore, we choose:

$$\Delta t = \begin{cases} 1 & \text{adhesion model} \\ 1 & \text{gap junctional model} \\ \dfrac{1}{\max\limits_{\boldsymbol{\eta'}}(T_{i,j}(\boldsymbol{\eta'}))} & \text{linear model} \end{cases} \qquad (21)$$

where, in the linear model on the hexagonal tiling, the normalization factor $\max_{\boldsymbol{\eta'}}(T_{i,j}(\boldsymbol{\eta'})) = \alpha + \max(2\beta, 0)$ and the parameters must satisfy the inequality $\alpha \geq -3\gamma - \min(2\beta, 0)$.

To reduce the stochastic noise, we averaged the outcomes of a series of independent stochastic simulations with the same initial conditions for each model and each parameter relative to the interaction and geometry. The system of ODEs, Eqs. (9, 12, 20) with the initial conditions corresponding to the CA simulations:

$$\begin{aligned} \rho(l_i, t=0) &= 0, & \text{for } l_i \text{ not in the source} \\ C_2(l_1, l_2, t=0) &= 0, & \forall l_i, \end{aligned} \qquad (22)$$

are numerically integrated using a fourth order Runge-Kutta method and compared with the results of the CA.

Let us show and discuss in detail the results for the more interesting cases.

Gap junctional model — density profile. The density profile at different times on a cylindrical geometry for the gap junctional model with interaction parameter $p = 1$ and $p = 0.9$ are shown in Figs. 4(a), 5(a). Cells, which are initially all positioned in the source at $x = 0$, migrate away from the full reservoir; as time advances, some of them move into the empty region on the right resulting in the advancing of the front of the density profile which gets closer to the sink at $x = L$. The effect of the sink is perceived only at large times, when cells arrive at the empty reservoir. Before that time, the slope of the density profile is strictly negative. The solution of the ODEs for the density profile is in good agreement with the simulations; nevertheless, some differences are noticeable on the right part of the density profile where $\rho(x, t) \ll 1$ due to the closure approximations.

Gap junctional model — discontinuity. At large times, in the position at one lattice step away from the sink, the density presents a discontinuity when $p = 1$. This is due to couples approaching the empty reservoirs. When they are in contact with it, the next favorable jump in the direction of the sink annihilates one cell of the couple, leaving the other in stall without functional gap junctional connections. It results in a small accumulation of cells on the sites just before the empty reservoir. This phenomenon is quite evident in the stochastic simulations where the slope of the density profile between the sites $L-2$ and $L-1$ becomes positive. From the solution for the steady state equations of the gap junctional model, the discontinuity also appears when the source and the sinks are few lattice steps apart, for example $L = 36$, but there is no inversion of the slope of the density profile. To obtain such results, it is necessary to increase the distance between the reservoirs. In fact, the density at a specific distance from the sink decreases when the cylinder length L increases and the decrease is slower at position $L-1$ than at other distances; eventually, for L large enough, it will hold true that $\rho(L-2) < \rho(L-1)$, Fig. 6. Hence, even though the strong effect of the closure approximations requires an increase of the distance between source and sink, taking into consideration the nearest neighbors two-point connected correlation function in the equations for the evolution of the system is enough to analytically reproduce the discontinuity when $p = 1$. As p decreases, the discontinuity in the density profile at the last point close to the sink becomes quickly less evident and at $p = 0.99$, it disappears. The reason is that, in these cases, single cells can jump and do not accumulate; therefore, the slope of the density profile remains negative at all positions for all times, Fig. 5(a).

Gap junctional model — correlation. In Figs. 4(b), 5(b), the correlation obtained with the cellular automata is very different from the correlation from the solution of the ODEs. The effect of the approximations makes it impossible to recover the exact correlation values of the stochastic simulations, but one can see that the correlations from the analytical model and simulations share the same properties. The particular shape of the connected correlation function at a given time t, Figs. 4(b), 5(b), divides the figure in three regions. In the *first region* that goes from the source of the cells to the point where the curve crosses the x axis, the correlation has a negative tail, and this is due to the exclusion process which forbids several cells from occupying the same site. This introduces in the system a short range repulsion between cells, especially close to the source where cells are crowded resulting in a negative value of the correlation. The *middle region* goes from where the correlation becomes strictly positive to the point where the stochastically simulated correlation is indistinguishable from zero (meaning that there the error bar of the correlation cross the x axis). It corresponds to the zone where the density starts to become low, and cells begin to feel the lack of neighbor cells. For any $p > 0.5$, cells tend to maintain gap junctions with the neighborhood during their motion due to the binding interaction, but the more the density approaches an approximative value $1/3$, the harder it is to preserve a contact with other cells. As a result, cells feel the crowding effect less, and the repulsive effect of the exclusion process is surpassed by the binding interaction. It is in this less dense region that, at the microscopical scale, it is possible to see couples moving together with a tendency not to separate until another neighbor gets close enough to form a new gap junction. At $p = 1$, the effect becomes very strong, and cells moves only if they maintain at least one functional gap junction and as soon as the density drops down, a population of moving gap junctional communicating couples and single stalled cells appears. The *third region* goes from the end of the previous region to the sink; here, the correlation is almost zero and clearly, the sink is yet too far from the

population of cells to be perceived.

Let us consider the time evolution of the connected correlation. At the beginning, the repulsion between the cells is stronger, but with the advancing of time, it decrease. The peak of the correlation moves toward the sink with a velocity $v(t) = t^{-1/2}$ and its height slowly decreases. Eventually, the correlation front will reach the sink and its peak slowly disappears leaving a very long and small negative correlation tail. At very large time, when the correlation is at the steady state, the connected correlation between two nearest neighbor sites parallel to the bases of the cylinder shows some abrupt changes in sign in the proximity of the empty reservoirs. Indeed, the large negative values of $C_\parallel(L-1)$ and $C_\perp(L-2)$ indicate a small number of cell couples in perfect accordance with the explanation of accumulation of single cells just before the sink.

Although the results above rely only on simulations in the cylindrical and radial geometries with hexagonal tiling, we will argue in Sec. (IV) that they are independent of the specific lattice and spatial dimensionality (in particular, the power that governs the decay of the velocity v of the front is always $-1/2$).

Linear model. The linear model with $\gamma = 0$ does not differ qualitatively from the gap junctional model. The former has a smaller correlation than the latter, and for $\alpha = 0$ at the steady state, the density presents a similar discontinuity near the sink. This model has the best agreement between the simulations and ODEs solutions.

Adhesion model. For the adhesion model, the greatest differences between the values of the correlation obtained from the CA simulations and from the solutions of the ODEs are in the region close to the source (see Fig. 7), and they increase with the increasing of the parameter q. The repulsion produced by the exclusion process is almost inexistent in the case of strong adhesion between cells, and the correlation is peaked at the position where the density is almost $1/2$ showing that, at the front of the density profile, cells are almost free, and they do not form any structures, but diffuse away. In contrast, at higher density, cells gather together, and because in part of the crowding and of the high number of links, they get trapped, which produces strong correlations between cells. The correlation produced by the adhesion has a peaked shape which moves in the direction of the sink with a velocity $v(t) \propto t^{-1/2}$. This is similar to what happens in the gap junctional model. On the other hand, its behavior is very different in the other two models. Very remarkable facts are that the height of the peak is constant in time, that it is much stronger than in the gap junctional case, and that, for large values of q, there is no negative correlation tail close to the source region from the beginning of the simulation on. The same constant peak of the connected correlation can be obtained in the linear model with $\gamma < 0$ and $\beta \leq 0$.

The main differences between CA simulations and analytical models are due to the closure approximations which neglect the long range correlations (we have seen on Fig. 3 how taking short-range correlations into account improve the agreement with CA simulations). Observing the ratio between the values from the ODEs and from the stochastic simulations of the two-point connected correlation function at different times, which is approximately proportional to $\ln(t)$ and $\frac{t}{t+k}$ (depending on the contact interaction), we conclude that such a discrepancy is produced by the exclusion rule, a common factor between different models.
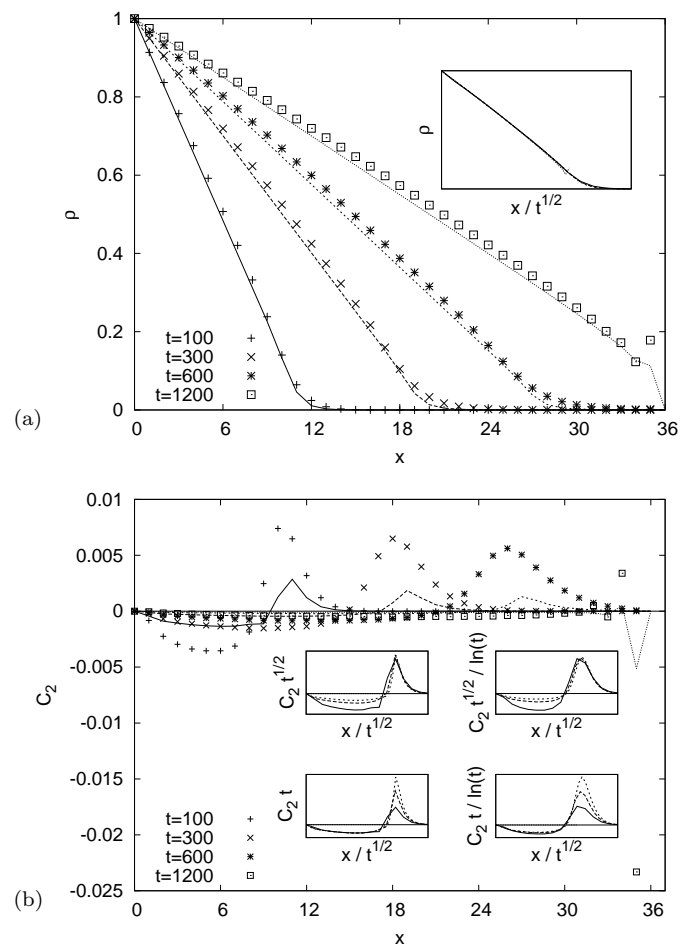

FIG. 4. Profile of (a) the density and (b) the connected correlation at different times for the gap junctional model on a cylindrical geometry with p=1 and L=36. The dots represent the results of the stochastic simulations, and the error bars are smaller than the size of the symbols. The lines refer to the results of Eqs. (9, 12, 20). The insets show the scaling behaviors. In (b), the insets on the left-hand side refer to the analytical results, and those on the right are the respective results obtained from cellular automata simulations. The scaling behaviors of the peak and on the negative tail are shown on the top row and bottom row of insets respectively.
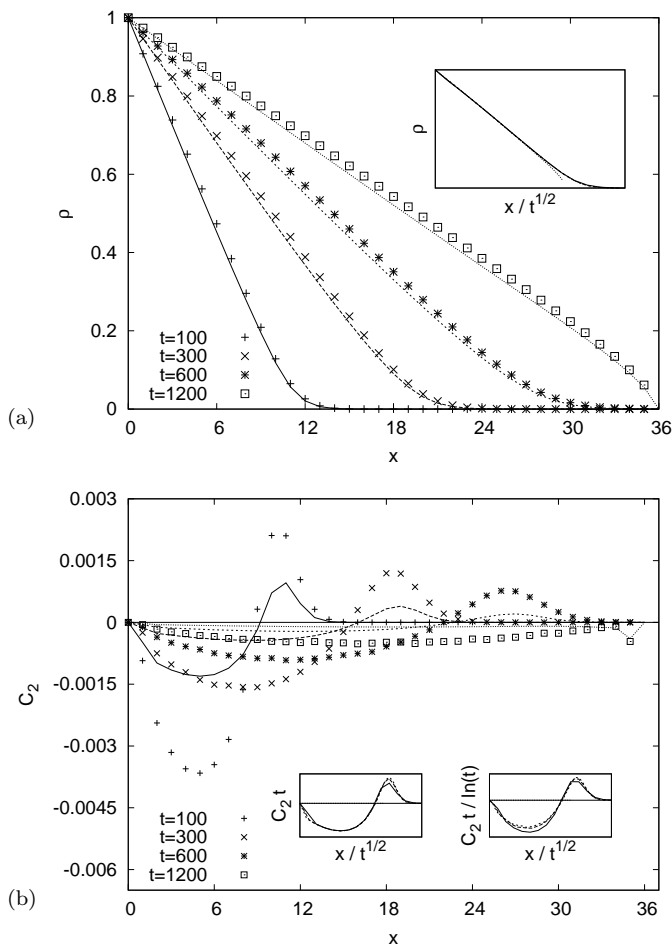
FIG. 6. Gap junctional model on a cylindrical geometry with p=1. The straight line shows $\rho(L - 1)$ and the dashed line represents $\rho(L - 2)$. From a numerical fits $\rho(L - 1)$ goes approximately as $\frac{1}{L^{1/3}}$ and $\rho(L - 2)$ goes approximately as $\frac{1}{L^{1/2}}$.

FIG. 5. Profile of (a) the density and (b) the connected correlation at different times for the gap junctional model on a cylindrical geometry with p=0.9 and L=36. The dots represent the results of the stochastic simulations, and the error bars are smaller than the size of the symbols. The lines refer to the results of Eqs. (9, 12, 20). The insets show the scaling behaviors. In (b), the inset on the left-hand side refers to the analytical results, and that on the right shows the respective results obtained from cellular automata simulations.

PDEs will enable a discussion of the self-similar phenomena that take place in the different models [76].

A.   Principle of the derivation

We are interested in variations of the local cell density and local correlations on a length scale R much larger than the lattice step λ. This length R can be the distance between source and sink in a steady-state regime, or the size of the region of a Petri dish that cells exiting a spheroid have already invaded Fig. 2(b). If $\boldsymbol{r}_i \in \mathbb{R}^n$ is the position (in continuous space) of the center of lattice tile i, we assume that there exists some function ρ such that $\langle \eta_i \rangle = \rho(\boldsymbol{r}_i/R)$ for all i and that ρ is differentiable as many times as necessary w.r.t. its argument $\boldsymbol{r}_i/R$. This implies that the average number of cells in the lattice site i, $\langle \eta_i \rangle$, varies from the average number in the nearest neighbor site $j \in \mathcal{V}(i)$ from no more than a quantity of the order of $1/R$. On the light of the previous results, we know that it is an excellent approximation most of the time. The exceptions are in the set-up with migration out of a spheroid, at short times after the initial condition when ρ has a steep slope (but this lasts less than ≈ 10 time steps), and in the steady state close to a sink of cells, for some particular values of the parameters of the models.

Likewise, we assume that there exists a function C_2 such that the connected correlation function between nearest neighbor sites i and $j \in \mathcal{V}(i)$ reads

$$\langle \eta_i \eta_j \rangle - \langle \eta_i \rangle \langle \eta_j \rangle = R^{-\xi} C_2 \left(\frac{\boldsymbol{r}_i + \boldsymbol{r}_j}{2R}, \boldsymbol{r}_j - \boldsymbol{r}_i \right).$$

Since the correlation between the sites i and j does not depend on the order of i and j, it is natural to use the

IV.   THE HYDRODYNAMIC LIMIT

In this section, we show and discuss how to derive, in the limit of infinitely large lattices, a finite set of coupled PDEs from the infinite system of ODEs of the previous section. Even if the PDEs are approximate, their solutions reproduce qualitatively the most probable (typical) behavior of the stochastic system at large times and large distances, the so-called hydrodynamic limit. In our case, this is not a very tight restriction since we have to disregard only a few time steps after the initial condition, and details on the scale of a single lattice step, such as the discontinuity close to the empty reservoir discussed earlier. On the other hand, the benefit is that only these
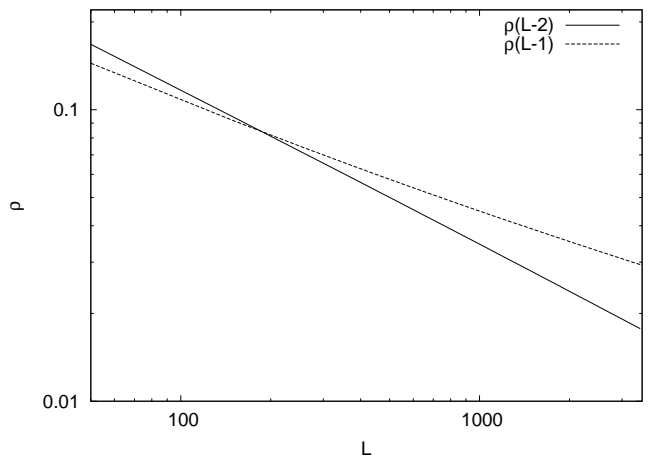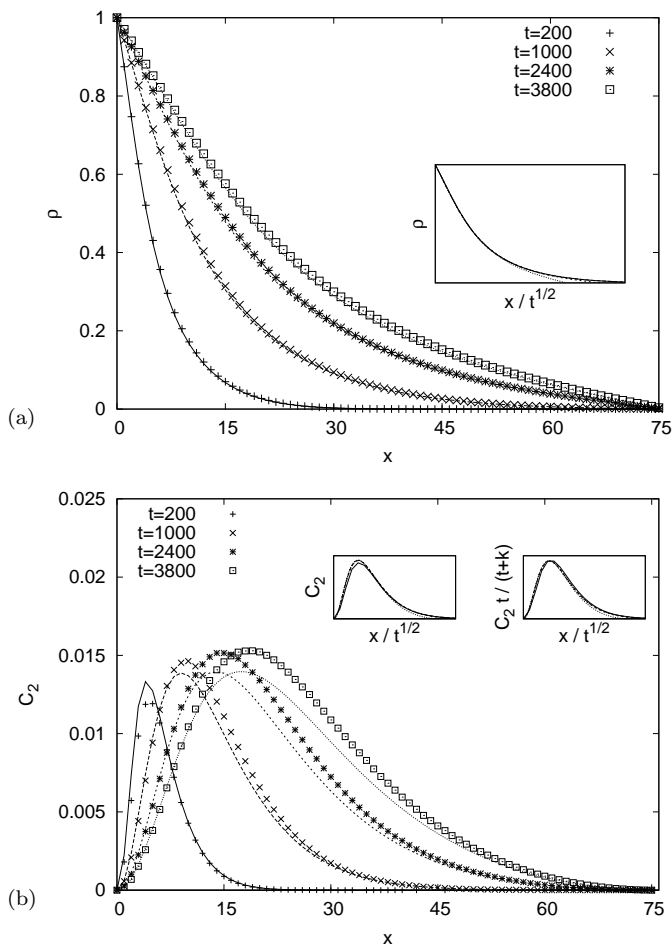
FIG. 7. Profile of (a) the density and (b) the connected correlation at different times for the adhesion model on a cylindrical geometry with q=0.2 and L=75. The dots represent the results of the stochastic simulations and the error bars are smaller than the size of the symbols. The lines refer to the results of Eqs. (9, 12, 20). The insets show the scaling behaviors. In (b), the inset on the left-hand side refers to the analytical results, and that on the right shows the respective results obtained from cellular automata simulations. The parameter $k = 6$ is properly chosen to show the scaling of the stochastic simulations.
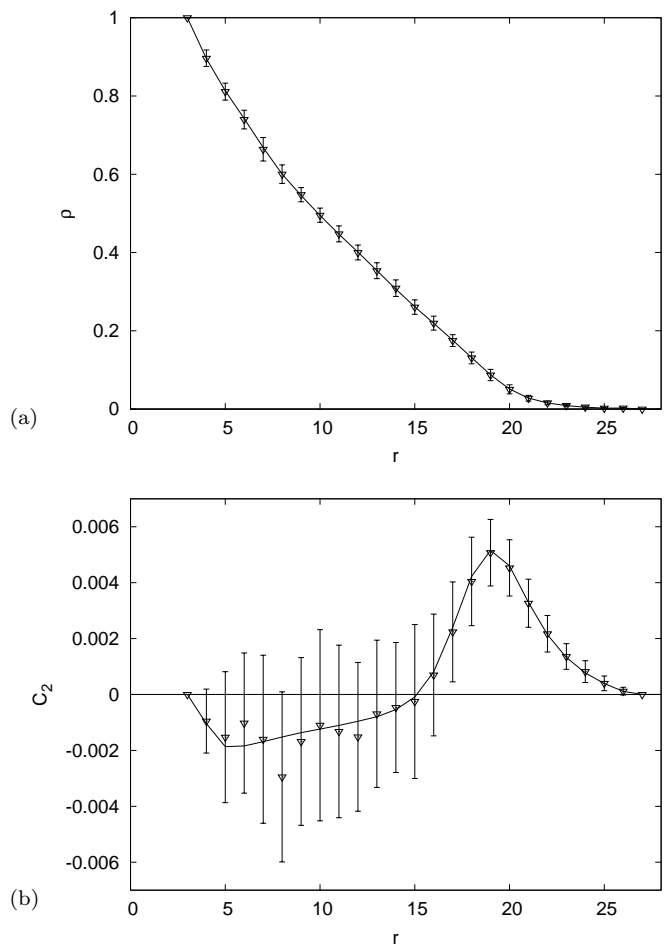

FIG. 8. Profile of 8(a) the density and 8(b) the connected correlation at time $t = 450$ for the gap junctional model on a radial geometry with $r_{src} = 3$ and $r_{snk} = 27$. The lines show the results of $3 \cdot 10^6$ averaged independent stochastic simulations at time t with same initial conditions and parameter p=1. The dots and the bars represent the average and the error for only 10^3 stochastic simulations.

middle point of i and j as the first argument of C_2. The second argument of C_2 can only take a finite number of values since the lattice is regular and it gives the direction of the two-point connected correlation function (sign is unimportant, i and j can be exchanged). To make expressions shorter, in the sequel, we will write $C_2 \left(\boldsymbol{r}, \boldsymbol{r}_j - \boldsymbol{r}_i \right)$ as C^l where l is the number of the direction $\boldsymbol{r}_j - \boldsymbol{r}_i$. Finally, in some of the previous simulations, we have seen that finite regions of the system may tend to homogenize (the density ρ gets uniform). Therefore, we introduce the possibility that C^l vanishes as R goes to infinity, and we assume that it involves a power law with the exponent ξ to be chosen later. The exponent

ξ must be nonnegative since the correlation function is always between -1 and 1.

It is generally not possible to define a regular function of two arguments, say $C_2(\boldsymbol{r}_i/R, \boldsymbol{r}_j/R)$: indeed, $C_2(\boldsymbol{r}_i/R, \boldsymbol{r}_{j'}/R)$ may have a difference of order 1 (not $1/R$), if j' is another nearest neighbor of i than j, but without $\boldsymbol{r}_{j'} - \boldsymbol{r}_i$ being colinear to $\boldsymbol{r}_j - \boldsymbol{r}_i$. This is for instance the case on the cylinder of Fig. 2 where C^1 and C^2 (e.g. correlations along the cylinder axis and perpendicular to the cylinder axis) can be quite different one from another because the boundary conditions break the rotation invariance of the lattice. The same holds if j' is not a nearest neighbor of i.

Then we insert the expressions for ρ and C_2 defined above into the set of discrete equations (14) for the time evolution of $\rho(i)$ and $C_2(i, j)$, and we perform a Taylor

expansion in powers of R^{-1} on the right-hand side of the equations around the points $\frac{r_i}{R}$ and $\frac{r_i+r_j}{2R}$ respectively, neglecting terms of orders three and above in R^{-1}. There are $1+d$ equations for each lattice site, where d is the number of different directions, hence a total of $(1+d)N$ equations for the whole lattice of N sites. But, since the equations are independent of the site up to an index shift, except for the sites on the boundaries of the system, we end up with only $1+d$ PDE, plus boundary conditions — the (approximate) "continuous space model", or "hydrodynamic limit".

Of course, one could systematically generalize this procedure to take into account correlations at distances larger than one lattice step or between three or more points — this may further improve the quality of the approximation of the continuous space model. To do this, one can redefine the multi-points density function:

$$\rho_p(\eta(\boldsymbol{r}_1), \ldots, \eta(\boldsymbol{r}_p)) = \rho_p(\eta(\Delta\boldsymbol{r}_1), \ldots, \eta(\Delta\boldsymbol{r}_p); \boldsymbol{r}_C)$$

where $\boldsymbol{r}_C$ is the centroid $\frac{\sum_k^p \boldsymbol{r}_k}{p}$, and $\Delta\boldsymbol{r}_k = \boldsymbol{r}_k - \boldsymbol{r}_C$ is the position of k-th point of ρ_p with respect to the centroid. Under the assumptions above, ρ_p with fixed $\Delta\boldsymbol{r}_k$ for all k is a regular function of $\boldsymbol{r}_C$. Then we could perform a series expansion, at fixed $\Delta\boldsymbol{r}_k$, in powers of $1/R$, of the multipoint probability density which can be in turn inserted into the system of discrete equations to yield a system of PDE. But here, we restrict ourselves to equations with ρ and C^j only.

B. Explicit expressions

Before explicitly applying the Taylor expansion to the discrete equations (14) for the evolution of the density and the connected correlations, it is important to remark that microscopic symmetries related to the regularity of the lattice reflect on the hydrodynamic limit.

Therefore, the equation of the two-point correlation between neighbor sites in one direction can be changed in the equation for the correlation in another direction by applying rotations of $\pi/3$ rad as consequence of the invariance of the hexagonal lattice under such transformations. This symmetry holds everywhere in the region of interest between the reservoirs, except in the sources and in the sinks.

To easily express the results and the symmetries of the system, we introduce a set D of unitary vectors identifying the three main lattice directions $e_0, e_1, e_2, \in \mathbb{R}^2$ such that:

$$e_0 = \begin{pmatrix} 0 \\ 1 \end{pmatrix}, \quad e_1 = \begin{pmatrix} \frac{\sqrt{3}}{2} \\ \frac{1}{2} \end{pmatrix}, \; e_2 = \begin{pmatrix} -\frac{\sqrt{3}}{2} \\ \frac{1}{2} \end{pmatrix}, \quad (23)$$

and the directional derivative $\nabla_i = e_i.\nabla$, where the index i means that the derivative is taken along the direction e_i.

In the case of the linear model, after performing the expansion in series in respect to the lattice distance λ to each term of the right-hand side of Eq. (14), the results are:

$$\partial_t \rho(\boldsymbol{r}, t) = \sum_{k \in \{\alpha, \beta, \gamma\}} k \left[\frac{B_{k,\rho}}{R^2} + \sum_{j=1}^3 \frac{B_{k,C^j} + B_{k,\rho,C^j}}{R^{2+\xi}} \right], \quad (24)$$

$$\frac{\partial_t C^j(\boldsymbol{r}, t)}{R^\xi} = \sum_{k \in \{\alpha, \beta, \gamma\}} k \left[\frac{A_k^j}{R^\xi} + \frac{B_{k,\rho}^j}{R^2} + \sum_{i=1}^3 \frac{B_{k,C^i}^j + B_{k,\rho,C^i}^j}{R^{2+\xi}} \right],$$
$$(25)$$

where the terms named A are the zero order and the terms named B are the second order of the Taylor expansion. In both equations, for the density and the correlations, there are no terms with odd order due to reflection symmetries of the lattice with respect to the lattice main directions. The equation for the density does not have any zero order term because the number of cells is locally conserved. The terms $B_{k,\rho}$, and B_{k,C^i} both contain terms proportional to the parameter $k \in \{\alpha, \beta, \gamma\}$ and each term only depends on the density and the correlation in direction i, respectively. B_{k,ρ,C^i} is the interaction part that takes into account the coupling between the density and the correlation in direction i. The same definitions hold for the terms in Eq. (25) where the upper index i (resp. j) refers to one of the main directions of the two-point connected correlation. From here on, all upper indices must be considered modulus three, even though it is not explicitly written. The details and expressions of all terms of the Taylor expansion A and B are in the Appendix.

Defining S and s the spatial subsets of $\mathbb{R}^2$ where there are sources and sinks respectively, the border and the initial conditions used to determine the solution of the system of partial differential equations are:

$$\begin{aligned}
\rho(\boldsymbol{r}, t) &= 1 & \forall \boldsymbol{r} \in S, \\
\rho(\boldsymbol{r}, t) &= 0 & \forall \boldsymbol{r} \in s, \\
C^i(\boldsymbol{r}, t) &= 0 & \forall \boldsymbol{r} \in s \cup S, \\
\rho(\boldsymbol{r}, t=0) &= C^i(\boldsymbol{r}, t=0) = 0 \;\; \forall \boldsymbol{r}.
\end{aligned}$$

The computation of the hydrodynamic limit for the adhesion model and the gap junctional model do not present ulterior problems, but a subtle more analytical complexity, and it can be performed in the same way as for the linear model. The hydrodynamic limit introduces some approximations in comparison to the discrete equations because one disregards terms of order bigger than $1/R^2$ in the series expansion. The numerical solutions of the PDEs are obtained by first re-discretizing the space derivatives using a second order finite difference method and then following the same numerical procedures as for the ODEs. They are accurately in agreement with the solutions of the ODEs for all the models proposed except when close to the sink. The analytical solutions are $C^{(\infty)}$; therefore, at the last step and two lattice steps away

from the empty reservoirs for the density and the correlations respectively, the numerical solutions smoothly go to zero in all cases making it impossible to retrieve the discontinuity observed in the discrete representation for some values of the parameters of the models at the steady state.

C. Self-similar behaviors

When non-interacting agents randomly jump to the nearest neighbor site with a constant transition rate, in the continuous limit, the concentration ρ of agents at the position $\boldsymbol{r}$ obeys: $\partial_t \rho(\boldsymbol{r}, t) = D\Delta\rho(\boldsymbol{r}, t)$. A generalization of the previous equation is the diffusion equation in porous media [77]: $\partial_t \rho(\boldsymbol{r}, t) = \boldsymbol{\nabla}[D(\rho)\boldsymbol{\nabla}\rho(\boldsymbol{r}, t)]$ which is useful to describe the concentration of agents when the agents interact [78]. Both equations with the initial condition $\rho(\boldsymbol{r}, t = 0) = \delta(\boldsymbol{r})$ show a self-similar behavior. Indeed, the solution of the diffusion equation at a given time t_1 is self-similar to the solution at time t_2: $\rho(\boldsymbol{r}, t_1) = \rho(\boldsymbol{r}\sqrt{\frac{t_1}{t_2}}, t_2)$. In this specific case, the self-similarity of the density function is exact, but in other cases, when correlations are taken into consideration, such behavior holds true only approximately and at large times far from the initial transient regime [79].

This form of self-similarity does not depend on the specific lattice, or space dimension in which the lattice is embedded. Indeed, if we choose a different type of tiling or a different space dimension, the diffusion equation above will keep the same structure, with a first-order time derivative and a (sum of) second-order space derivatives. Only numerical coefficients in D will change.

For the geometrical dispositions of the reservoirs proposed in this work, the presence of a sink destroys the self-similar behaviors of the density of cells; therefore, we consider the situation where the sink is sufficiently far ($T \leq L$, where T is the maximum time we run the stochastic simulation) so that the cells can migrate for long enough without perceiving the sink.

To investigate the self-similarity of both the density and the two-point connected correlation, we exploit the PDEs derived from the hydrodynamic limit for the linear model to retrieve the scaling of the solutions. When there are no correlations, the Eq. (24) becomes a particular type of diffusion equation in porous media; therefore, after eliminating the factor R^{-2} thanks to the change of variable $t = R^2 \tilde{t}$, we can write the solution of Eq. (24) as:

$$\rho(\boldsymbol{r}, \tilde{t}) = f(\frac{\boldsymbol{r}}{\sqrt{\tilde{t}}}) + \epsilon g(\boldsymbol{r}, \tilde{t}) \qquad (26)$$

where the function $g(\boldsymbol{r}, \tilde{t})$ on the right-hand side tells us how the exact solution of the density profile differs from the self-similar result $f(\frac{\boldsymbol{r}}{\sqrt{\tilde{t}}})$ obtained in case of no correlations. The parameter ϵ is the magnitude of such discrepancy and for negligible correlations in compari-

son with the density, it holds true that $\epsilon \ll 1$. Inserting the solution $f(\frac{\boldsymbol{r}}{\sqrt{\tilde{t}}})$ into Eq. (25), one can find the approximate solutions for the correlations and use them again in the equations for the density. Repeating the process iteratively, more accurate solutions for the density and the correlations can be found, and at each repetition, it is possible to check the validity of the hypothesis $\epsilon \ll 1$. In the insets of Figs. 4(a), 5(a), 7(a), the stochastic simulations of each model at different times are scaled using the scaling behaviors of the function $f(\frac{\boldsymbol{r}}{\sqrt{t}})$. We have observed that the same self-similar behavior holds for stochastic simulations of the linear model with various values of parameters which reproduce the adhesion or gap junctional model behaviors. The perfect overlap between the curves shows that analyzing the scaling of $\rho(\boldsymbol{r}, t)$ under the conditions of negligible correlations is very good. As we stated above, this self-similar behavior will hold also in different space dimensions and on other lattices because the structure of Eq. (24) will be the same (first-order time derivative, second-order space derivative), even though numerical coefficients may be different from those provided in the Appendix.

In the sequel, we change the notation $\tilde{t}$ to t for simplicity, and we consider that $\rho(\boldsymbol{r}, t) = f(\frac{\boldsymbol{r}}{\sqrt{t}})$. Eq. (25) now reads:

$$\frac{\partial_t C^j(\boldsymbol{r}, t)}{R^{\xi+2}} = \sum_{k \in \{\alpha, \beta, \gamma\}} k \left[\frac{A_k^j}{R^\xi} + \frac{B_{k,\rho}^j}{R^2} + \sum_{i=1}^3 \frac{B_{k,C^i}^j + B_{k,\rho,C^i}^j}{R^{2+\xi}} \right]. \qquad (27)$$

Linear model — only exclusion interactions. To begin, let us consider the simplest case $\alpha \neq 0$ and $\beta = \gamma = 0$. In Eq. (27), choosing $\xi \neq 2$ leads either to a trivial solution $C = 0$ or to an inconsistent solution $\nabla\rho = 0$. Therefore, the only possible choice is $\xi = 2$ which results in $C^j(\boldsymbol{r}, t) \sim \Delta\rho(\boldsymbol{r}, t) \sim \frac{h(\frac{\boldsymbol{r}}{\sqrt{t}})}{t}$, while $\partial_t C^j(\boldsymbol{r}, t) \sim \nabla_j^2 C^j(\boldsymbol{r}, t)$ are asymptotically negligible.

Linear model — mimicking gap junctions p < 1. The same discussion is valid for $\gamma = 0$ and both α and β different from zero. This brings us to the same self-similar behavior $C^j(\boldsymbol{r}, t) \sim \Delta\rho(\boldsymbol{r}, t) \sim \frac{h(\frac{\boldsymbol{r}}{\sqrt{t}})}{t}$. In the left insets of Fig. 5(b), the numerical solutions of $C_\parallel$ for the gap junctional model are rescaled using $C^j(\boldsymbol{r}, t) = \frac{h(\frac{\boldsymbol{r}}{\sqrt{t}})}{t}$ as self-similar behaviors. The good agreement given by the overlapping of the curves also shows that the gap junctional model for $p < 1$ shares the same self-similar behaviors as the linear model.

The right inset of Fig. 5(b) shows that the results of the stochastic simulations scale as $C^j(\boldsymbol{r}, t) = \ln(t)\frac{h(\frac{\boldsymbol{r}}{\sqrt{t}})}{t}$. The logarithmic correction in the scaling of the correlations is due to the long range correlation produced by the exclusion process, and they cannot be retrieved using the hydrodynamic limit equations.

Linear model — mimicking gap junctions p = 1. The case $\alpha = \gamma = 0$ and $\beta \neq 0$ is particularly interesting because the solution of Eq. (24) with no correlations is no longer a smooth function of the position, but it is a

piecewise function with two regions: the part closer to the source, where the density decreases with a finite negative slope as its distance from the full reservoir increases, and the part that goes from the positions where the density becomes zero onwards, in which the density is uniformly null [80].

The discontinuity in the first derivative of the density profile for $\alpha = 0$ is true only when there are no correlations, while, when correlations are introduced, such discontinuity disappears. Hence, always under the condition $\rho(\boldsymbol{r},t) = f(\frac{r}{\sqrt{t}})$, in the region where $\rho(\boldsymbol{r},t) \neq 0$, the correlation rescale as $C^j(\boldsymbol{r},t) = \frac{h(\frac{r}{\sqrt{t}})}{t}$.

On the other hand, in the region where the density is zero, all terms containing ρ or its derivatives are zero, and Eq. (27) becomes:

$$\frac{\partial_t C^j(\boldsymbol{r},t)}{R^{\xi+2}} = \beta \left[\frac{A_\beta^j}{R^\xi} + \sum_{i=1}^3 \frac{B_{\beta,C^i}^j}{R^{2+\xi}} \right]. \qquad (28)$$

The leading order term in $1/R$ must be zero. Hence, $A_\beta^0 = A_\beta^1 = A_\beta^2 = 0$ and Eq. (28) becomes a system of three standard diffusion equations without sinks or sources. From the expression of A_β^j, we conclude that $C^0 = C^1 = C^2$. Therefore, Eq. (28) simplifies to a standard diffusion equation for any of the quantities C^j. If we look for a nontrivial result for C^j, the three diffusion equations become identical to: $\partial_t C^0 = \frac{1}{24} \Delta C^0$.

From the conservation of the correlation C^j and the solution of the diffusion equation follow that the correlations in all directions scale as: $C^j(\boldsymbol{r},t) = \frac{q(\frac{r}{\sqrt{t}})}{\sqrt{t}}$. If the complete solution Eq. (24) is used instead of the approximate solution of the density with no correlations, then the region with $\rho = 0$ changes in $\rho \ll 1$. This correction, even though it is very small ($\epsilon \ll 1$), solves the inconsistency of having a correlation different from zero in a region with no cells, and leaves unchanged the self-similar behavior.

In the top and bottom left insets of Fig. 4(b), we see that the numerical solutions of the ODEs at different times for the gap junctional model with $p = 1$ reproduce the self-similarity of the case $\alpha = \gamma = 0$, and $\beta \neq 0$, where the negative tail of the correlations scales as $\frac{h(\frac{r}{\sqrt{t}})}{t}$ and the positive peak of the correlations as $\frac{q(\frac{r}{\sqrt{t}})}{\sqrt{t}}$. The top and bottom right insets of Fig. 4(b) show the scaling for the stochastic results of the gap junctional model with $p = 1$, and also in this case the logarithmic correction is due to the long range correlations.

Linear model — mimicking adhesion. In the case with the parameter $\gamma \neq 0$, the scaling behavior is given by A_γ^j which contains terms proportional to positive powers of $\rho(\boldsymbol{r},t)$ and R^ξ. One has to choose $\xi = 0$ to get consistent results. Consequently, the self-similar form of the correlation is $C^j(\boldsymbol{r},t) \sim \rho(\boldsymbol{r},t) \sim f(\frac{r}{\sqrt{t}})$, meaning that the height of the correlation peak remains constant during the migration process. When both β and γ are negative, the linear model mimics the behaviors of the adhesion model. Indeed, scaling the numerical solutions at different times of the ODEs for the adhesion model with the same scaling behavior of the corresponding linear model, we have a perfect overlapping of the curves, left inset of Fig. 7(b). In the right inset, the results of the stochastic simulations show an increasing of the height of the peak due to long range correlations. In the adhesion model such effects are much smaller and asymptotically negligible than in the other cases where correlations produce logarithmic correction. In all the insets in Figs. 4, 5, 7, the curves that deviate the most from the self-similar behaviors are those where the sink has been reached and the constraint $t \leq L$ does not hold any more.

V. DISCUSSION

A. Measuring the correlations in experiments

Let us try to estimate for what parameters (population size, number of repetitions, duration of experiment) one could get an exploitable measure of the connected correlation function in migration assays of cells on Petri dishes. To be closer to this experimental situation, we performed stochastic simulations and got numeric solutions of Eqs. (14, 17) on the radial geometry of Fig. 2(b). In this geometry, the density profile is computed by averaging the densities over the sites belonging to the same annulus centered on the source center, O_{src}, and one lattice step thick, so that the density depends only on the distance from the source center. The profile of the nearest neighbor two-point connected correlation is obtained by averaging the correlation over the six nearest neighbors of one of its sites, then the result is averaged over all the sites belonging to the respective annulus as for the density profile. At large times, the radius of the outer region invaded by cells is large, so that one can expect that values of the observables in that region will be similar to what is found with the cylindrical geometry of Fig. 2(a). It may be different both at short times and close to the center of the disk. However, it turns out that the "correlation wave" observed and discussed in Sec. (III) and Sec. (IV) is still present in this setup and even amplified. In addition, in the case of the gap junctional model where correlations are the most important far from the center of the disk, the relative error on the measure of the correlations tends to decrease with time, since the measure is done by averaging over a region which gets larger and larger. To show these results, in Fig. 8, we have plotted the average of 1000 repeated simulations at a given time and the respective error bars in comparison with the results obtained from the average of a much higher number of simulations for the gap junctional model. The figure, on one side, shows that it is not necessary to have a prohibitively high statistic to observe the two-point connected correlations, while, on the other side, it shows that simulations on the radial ge-

ometry are very useful to compare the proposed models with experimental data on Petri dishes.

B. Perspectives

Let us comment on the connection of the present approach to other techniques. The structure of Eq. (26) can be seen as an iterative computation of the local density ρ, where the first step would yield the mean-field approximation to the evolution equation of ρ (all correlations C_2 being replaced with 0) and the second step involves an improvement to this equation that is expressed by means of C_2. If one formally integrates the evolution equation for C_2, replacing in it the value of ρ that solves the mean-field equation for ρ, $C_2(t)$ gets expressed as an integral of $\rho(t)$ over the time range $0 \leq t' \leq t$. Then, replacing $C_2(t)$ in the equation for ρ yields an integro-differential equation for $\rho(t)$ — a non-markovian model with memory kernels [81]. This is much like what is obtained in the Mori-Zwanzig formalism [82], the non-markovian character being due to a tentative of description of a complex situation (many spatially inhomogeneous configurations) with a simple quantity (the local density ρ) by a kind of "projection". However, there are many ways to do this systematic reduction of numbers of freedom [21, 81–83], and, when truncated to the first or second step of iteration, not all schemes yield analytical approximation with the same degree of agreement to simulations.

It would also be interesting to have a deeper understanding of which situations (and possibly a criterion to discern them) allow one to get analytical results of good quality while taking into account only short-range correlations. Our situation might be related to the case of the Smoluchowski theory of aggregation discussed in [21]. The success of this method in spite of the simple approximation it relies on may be explained, at least for some of the models it was applied to, by the observation that there is no propagator renormalization in the corresponding field theories (in an RG approach), and hence no anomalous dimension for the diffusion constant or the fields.

There are natural and biologically relevant extensions of the models considered in this work that can be studied with the same techniques, and we plan to deal with some of them in our future work. First, one can incorporate the effect of cell proliferation. According to previous works [51–53], cell proliferation leads to short range correlations (because daughter cells are close to the position of the mother cell) which have the same order of magnitude as the ones produced by our contact interactions. In a realistic model, one should probably take care of the migration-proliferation dichotomy in the behavior of cells [84, 85], which may enhance correlations. Then, the techniques can be useful for other models of exclusion processes with interactions like the ones of [86] and [87]. It can also straightforwardly be extended to 3D migration, for which experimental data is more diffi-

cult to obtain, but still accessible thanks e.g. to confocal microscopy. The techniques of our work can be used in spaces with dimension other than two, or on other lattices than the hexagonal tiling. In these cases, as explained in Sec. (IV C), we would find an unchanged self-similar behavior for each respective model.

An extension to disordered systems would also be useful. In the simple situation of one or a few obstacles or heterogeneities in the lattice with a known position, or a boundary with an irregular, but known shape, our techniques can be straightforwardly extended by writing equations specific to this geometry. The invariance and symmetry properties of the density and correlation functions are lost, but we expect the same large-time self-similar behaviors. A more complicated situation with many obstacles or inhomogeneous space can be modeled for instance by deformed lattices [38, 55]. These cases may yield more realistic models of biological processes and make it possible to avoid some artifacts caused by regular lattices. As a first step to address this problem, one can use the so-called annealed approximation, where the computation of the density of cells is performed as if the obstacles would be redistributed randomly at each time step. The result is an approximate value for the diffusivity D that depends on the density of obstacles, valid for not too large densities of obstacles [60]. In this case, we expect that the self-similar behaviors that we found in this work still hold.

C. Conclusion

We have shown how to extend simple mean-field analytical approximations of spatially inhomogeneous exclusion processes with local contact-like interactions as can be found in biological situations of interacting migrating cells. Supplementing the local density of agents (or cells) with short-range correlations in the analytical, deterministic, macroscopic approximations of the stochastic processes yields not only results in better agreement with stochastic simulations. It can also be used to infer more precise estimates of the interaction parameters from experimental data about density and correlations of migrating cells than from density values alone, when the type of cell-cell contact interactions is known. If the latter is unknown, it can be discovered by comparison with the classification scheme that we provide, which was made possible thanks to the PDEs approximation of the exclusion processes: the self-similar behavior of the two-point connected correlation function is highly dependent on the type of contact interactions, adhesive or contact-maintaining for instance, at the microscopic level.

of this work was done. We acknowledge financial support from the Comité de financement des théoriciens of the IN2P3.

Appendix: Hydrodynamic limit explicit expressions

Using the techniques developed in Sec. (IV A), we explicitly derived the A and B terms introduced in Eqs. (24, 25) for the hexagonal tiling in the cases of gap junctional, adhesion, and linear models. For the sake of shortness, we write only the expressions of the linear model, which are enough to study all the self similar behaviors of the three models. In each term, the first lower index represents the corresponding interaction parameter and, when present, the second lower index shows the variable on which the term depends. To distinguish the correlations, we use an upper index in the terms A and B, the value of which identifies the direction given by the vectors in Eq. (23).

$$B_{\alpha,\rho} = \frac{1}{4}\Delta\rho$$
$$B_{\alpha,C^j} = B_{\alpha,C^{j+1}} = B_{\alpha,C^{j+2}} = B_{\alpha,\rho,C^j} = B_{\alpha,\rho,C^{j+1}} = B_{\alpha,\rho,C^{j+2}} = 0$$

$$B_{\beta,\rho} = \frac{2}{3}\left[(\nabla_1\rho)^2 + \nabla_2\rho\nabla_1\rho + (\nabla_2\rho)^2 + \frac{3}{4}\rho\Delta\rho\right]$$
$$B_{\beta,C^0} = \frac{1}{12}(3\Delta - 4\nabla_0^2)C^0$$
$$B_{\beta,C^1} = \frac{1}{12}(3\Delta - 4\nabla_1^2)C^1$$
$$B_{\beta,C^2} = \frac{1}{12}(3\Delta - 4\nabla_2^2)C^2$$
$$B_{\beta,\rho,C^0} = B_{\beta,\rho,C^1} = B_{\beta,\rho,C^2} = 0$$

$$B_{\gamma,\rho} = \frac{1}{3}\left\{2(5-7\rho)[(\nabla_1\rho)^2 + \nabla_2\rho\nabla_1\rho + (\nabla_2\rho)^2] + \frac{3}{4}(10-7\rho)\rho\Delta\rho\right\}$$
$$B_{\gamma,C^0} = B_{\gamma,C^1} = B_{\gamma,C^2} = 0$$
$$B_{\gamma,\rho,C^0} = \frac{1}{12}\left[(1-\rho)(4\nabla_0^2 + 3\Delta)C^0 + C^0(3\Delta - 14\nabla_0^2)\rho - 18\nabla_0\rho\nabla_0 C^0\right]$$
$$B_{\gamma,\rho,C^1} = \frac{1}{12}\left[(1-\rho)(4\nabla_1^2 + 3\Delta)C^1 + C^1(3\Delta - 14\nabla_1^2)\rho - 18\nabla_1\rho\nabla_1 C^1\right]$$
$$B_{\gamma,\rho,C^2} = \frac{1}{12}\left[(1-\rho)(4\nabla_2^2 + 3\Delta)C^2 + C^2(3\Delta - 14\nabla_2^2)\rho - 18\nabla_2\rho\nabla_2 C^2\right]$$

$$A_\alpha^j = \frac{1}{3}(C^{j+1} + C^{j+2} - 5C^j)$$
$$B_{\alpha,\rho}^j = -\frac{1}{6}\left[(\nabla_{j+1} + \nabla_{j+2})\rho\right]^2$$
$$B_{\alpha,C^j}^j = B_{\alpha,\rho,C^j}^j = B_{\alpha,\rho,C^{j+2}}^j = B_{\alpha,\rho,C^{j+2}}^j = 0$$
$$B_{\alpha,C^{j+1}}^j = \frac{1}{24}\nabla_{j+2}^2 C^{j+1}$$
$$B_{\alpha,C^{j+2}}^j = \frac{1}{24}\nabla_{j+1}^2 C^{j+2}$$

$$A_\beta^j = \frac{1}{3}[C^{j+1} + C^{j+2} - 2(1+3\rho)C^j]$$

$$B_{\beta,\rho}^j = \frac{1}{6}\left[(1-4\rho)(\nabla_j\rho)^2 - 2\rho(1-\rho)\nabla_{j+1}\nabla_{j+2}\rho\right]$$

$$B_{\beta,C^j}^j = 0$$

$$B_{\beta,C^{j+1}}^j = \frac{1}{24}\nabla_{j+2}^2 C^{j+1}$$

$$B_{\beta,C^{j+2}}^j = \frac{1}{24}\nabla_{j+1}^2 C^{j+2}$$

$$B_{\beta,\rho,C^j}^j = -\frac{1}{48}C^j(28\nabla_{j+1}\nabla_{j+2} + 57\Delta)\rho$$

$$B_{\beta,\rho,C^{j+1}}^j = \frac{1}{3}[C^{j+1}\nabla_{j+1} - (\nabla_{j+2}C^{j+1})](\nabla_j + \nabla_{j+1})\rho$$

$$B_{\beta,\rho,C^{j+2}}^j = \frac{1}{3}[C^{j+2}\nabla_{j+2} - (\nabla_{j+1}C^{j+2})](\nabla_j + \nabla_{j+2})\rho$$

$$A_\gamma^j = \frac{1}{3}\left\{\rho\left[-3\rho(1-\rho)^2 R^{-\xi} + (\rho+2)(C^{j+1} + C^{j+2}) - (7\rho+5)C^j]\right]\right\} - C^j$$

$$B_{\gamma,\rho}^j = \frac{1}{12}\left\{\left(24\rho^2 - 26\rho + 3\right)(\nabla_j\rho)^2 + \rho(\rho-1)\left[(3-14\rho)(\nabla_{j+1}^2 + \nabla_{j+2}^2) + 6(1-4\rho)(\nabla_{j+1}\nabla_{j+2})\right]\rho\right\}$$

$$B_{\gamma,C^j}^j = B_{\gamma,C^{j+1}}^j = B_{\gamma,C^{j+2}}^j = 0$$

$$B_{\gamma,\rho,C^j}^j = \frac{1}{24}\left\{4(1-\rho)\rho\nabla_j^2 C^j + 4(4\rho-3)\nabla_j C^j\nabla_j\rho\right.$$
$$\left. -C^j\left[146(\nabla_j\rho)^2 - 224\nabla_{j+1}\rho\nabla_{j+2}\rho + (21\nabla_j^2 - 32\nabla_{j+1}\nabla_{j+2})\rho + \rho(158\nabla_j^2 - 128\nabla_{j+1}\nabla_{j+2})\rho\right]\right\}$$

$$B_{\gamma,\rho,C^{j+1}}^j = \frac{1}{24}\left\{\rho(\rho-1)(\nabla_{j+2}^2 - 8\nabla_{j+1}\nabla_{j+2} - 8\nabla_{j+1}^2)C^{j+1} + 3\nabla_{j+2}^2 C^{j+1} + 2C^{j+1}\nabla_j\rho(\nabla_j + 16\nabla_{j+2})\rho\right.$$
$$+ 2(1+\rho)C^{j+1}(26\nabla_{j+1}\nabla_{j+2} + 5\nabla_{j+1}^2 + 9\nabla_{j+2}^2)\rho - 64C^{j+1}\nabla_{j+1}\nabla_{j+2}\rho$$
$$\left. + 8\nabla_{j+1}C^{j+1}[(2\rho-1)\nabla_j + (\rho-1)\nabla_{j+2}]\rho - 4\nabla_{j+2}C^{j+1}[(\rho+1)\nabla_j + 4(2\rho-1)\nabla_{j+2}]\rho\right\}$$

$$B_{\gamma,\rho,C^{j+2}}^j = \frac{1}{24}\left\{\rho(\rho-1)(\nabla_{j+1}^2 - 8\nabla_{j+2}\nabla_{j+1} - 8\nabla_{j+2}^2)C^{j+2} + 3\nabla_{j+1}^2 C^{j+2} + 2C^{j+2}\nabla_j\rho(\nabla_j + 16\nabla_{j+1})\rho\right.$$
$$+ 2(1+\rho)C^{j+2}(26\nabla_{j+2}\nabla_{j+1} + 5\nabla_{j+2}^2 + 9\nabla_{j+1}^2)\rho - 64C^{j+2}\nabla_{j+2}\nabla_{j+1}\rho$$
$$\left. + 8\nabla_{j+2}C^{j+2}[(2\rho-1)\nabla_j + (\rho-1)\nabla_{j+1}]\rho - 4\nabla_{j+1}C^{j+2}[(\rho+1)\nabla_j + 4(2\rho-1)\nabla_{j+1}]\rho\right\}$$

---

[1] J. von Neumann, *Theory of Self-Reproducing Automata* (University of Illinois Press, Urbana and London, 1966).

[2] U. Frisch, B. Hasslacher, and Y. Pomeau, Phys. Rev. Lett. **56**, 1505 (1986).

[3] D. d'Humières, P. Lallemand, and U. Frisch, Europhys. Lett. **2**, 291 (1986).

[4] D. H. Rothman and S. Zaleski, Rev. Mod. Phys. **66**, 1417 (1994).

[5] J.-P. Boon, D. Dab, R. Kapral, and A. Lawniczak, Phys. Rep. **273**, 55 (1996).

[6] F. Ritort and P. Sollich, Adv. Phys. **52**, 219 (2003), arXiv:cond-mat/0210382.

[7] D. L. Smith, B. Lucey, L. A. Waller, J. E. Childs, and L. A. Real, Proc. Natl. Acad. Sci. USA **99**, 3668 (2002).

[8] D. Chowdhury, L. Santen, and A. Schadschneider, Phys. Rep. **329**, 199 (2000).

[9] D. Helbing, Rev. Mod. Phys. **73**, 1067 (2001).

[10] H. J. Hilhorst and C. Appert-Rolland, J. Stat. Mech. **2012**, P06009 (2012).

[11] A. Deutsch and S. Dormann, *Cellular Automaton Modeling of Biological Pattern Formation: Characterization, Applications, and Analysis* (Birkhäuser, Boston, 2005).

[12] A. R. A. Anderson, M. A. J. Chaplain, and K. A. Rejniak, eds., *Single-cell-based Models in Biology and Medicine* (Birkhäuser, Boston, 2007).

[13] M. S. Alber, M. A. Kiskowski, J. A. Glazier, and Y. Jiang, in *Mathematical Systems Theory in Biology, Communication, and Finance*, The IMA Volumes in mathematics & its applications, Vol. 134 (Springer-Verlag, Berlin, 2003) pp. 1–39.

[14] T. M. Liggett, *Interacting Particle Systems*, Classics in Mathematics (Springer, Berlin/Heidelberg, 1985).

[15] T. M. Liggett, *Stochastic Interacting Systems: Contact, Voter and Exclusion Processes*, Grundlehren Der Mathematischen Wissenschaften (Springer, Berlin/Heidelberg, 1999).

[16] H. Spohn, *Large Scale Dynamics of Interacting Particles* (Springer-Verlag, Berlin, 1991).

[17] C. Appert-Rolland, B. Derrida, V. Lecomte, and F. van Wijland, Phys. Rev. E **78**, 021122 (2008).

[18] J. P. Garrahan, P. Sollich, and C. Toninelli, "Kinetically constrained models," in *Dynamical heterogeneities in Glasses, colloids and granular media and jamming transitions*, International series of monographs in physics (Oxford University Press, Oxford, 2011) Chap. 10, pp. 341–369, arXiv:1009.6113.

[19] A. E. Fernando, K. A. Landman, and M. J. Simpson, Phys. Rev. E **81**, 011903 (2010).

[20] K. Anguige and C. Schmeiser, J. Math. Biol. **58**, 395 (2009).

[21] U. C. Täuber, M. Howard, and B. P. Vollmayr-Lee, J. Phys. A – Math. Gen. **38**, R79 (2005).

[22] P. J. Murray, C. M. Edwards, M. J. Tindall, and P. K. Maini, Phys. Rev. E **85**, 021921 (2012).

[23] A. De Masi and E. Presutti, *Mathematical Methods for Hydrodynamic Limits* (Springer-Verlag, Berlin, 1991).

[24] C. Kipnis and C. Landim, *Scaling Limits of Interacting Particle Systems* (Springer-Verlag, Berlin, 1999).

[25] J. Fritz, J. Stat. Phys. **47**, 551 (1987).

[26] M. Z. Guo, G. C. Papanicolaou, and S. R. S. Varadhan, Commun. Math. Phys. **118**, 31 (1988).

[27] S. R. S. Varadhan, Commun. Math. Phys. **135**, 313 (1991).

[28] S. R. S. Varadhan and H.-T. Yau, Asian J. Math. **1**, 623 (1997).

[29] P. Gonçalves, C. Landim, and C. Toninelli, Annales de l'I.H.P. (B) Probabilités et statistiques **45**, 887 (2009), arXiv:0704.2242.

[30] S. Chapman and T. G. Cowling, *The Mathematical Theory of Non-Uniform Gases – 3rd edition* (Cambridge University Press, Cambridge, 1970).

[31] N. G. van Kampen, *Stochastic Processes in Physics and Chemistry – 2nd Edition* (Elsevier, Amsterdam, 1997).

[32] S. Turner, J. A. Sherratt, K. J. Painter, and N. J. Savill, Phys. Rev. E **69**, 021910 (2004).

[33] M. Alber, N. Chen, T. Glimm, and P. M. Lushnikov, Phys. Rev. E **73**, 051901 (2006).

[34] P. M. Lushnikov, N. Chen, and M. Alber, Phys. Rev. E **78**, 061904 (2008), arXiv:0809.2260.

[35] D. Drasdo, Advances in Complex Systems **8**, 319 (2005).

[36] H. Hatzikirou, L. Brusch, and A. Deutsch, Acta Physica Polonica B Proc. Suppl. **3**, 399 (2010).

[37] P. J. Murray, C. M. Edwards, M. J. Tindall, and P. K. Maini, Phys. Rev. E **80**, 031912 (2009).

[38] M. Block, E. Schöll, and D. Drasdo, Phys. Rev. Lett. **99**, 248101 (2007).

[39] J.-P. Boon and J. F. Lutsko, Europhys. Lett. **80**, 60006 (2007).

[40] J. F. Lutsko and J.-P. Boon, Phys. Rev. E **77**, 051103 (2008).

[41] J.-P. Boon, J. F. Lutsko, and C. Lutsko, Phys. Rev. E **85**, 021126 (2012).

[42] C. J. Penington, B. D. Hughes, and K. A. Landman, Phys. Rev. E **84**, 041120 (2011).

[43] C. Deroulers, M. Aubert, M. Badoual, and B. Grammaticos, Phys. Rev. E **79**, 031917 (2009).

[44] M. J. Simpson, C. Towne, D. L. S. McElwain, and Z. Upton, Phys. Rev. E **82**, 041901 (2010).

[45] P. N. Butcher and S. Summerfield, J. Phys. C – Solid State Phys. **14**, L1099 (1981).

[46] R. A. Tahir-Kheli and R. J. Elliott, Phys. Rev. B **27**, 844 (1983).

[47] M. G. Rudavets, J. Phys. A – Math. Gen. **26**, 5313 (1993).

[48] J. Joo and J. L. Lebowitz, Phys. Rev. E **70**, 036114 (2004).

[49] H. Matsuda, N. Ogita, A. Sasaki, and K. Satō, Prog. Theor. Phys. **88**, 1035 (1992).

[50] S. P. Ellner, J. Theor. Biol. **210**, 435 (2001).

[51] R. E. Baker and M. J. Simpson, Phys. Rev. E **82**, 041905 (2010).

[52] S. T. Johnston, M. J. Simpson, and R. E. Baker, Phys. Rev. E **85**, 051922 (2012).

[53] M. J. Simpson and R. E. Baker, Phys. Rev. E **83**, 051922 (2011).

[54] E. Khain, M. Katakowski, S. Hopkins, A. Szalad, X. Zheng, F. Jiang, and M. Chopp, Phys. Rev. E **83**, 031920 (2011).

[55] M. Aubert, M. Badoual, S. Féreol, C. Christov, and B. Grammaticos, Phys. Biol. **3**, 93 (2006).

[56] E. Khain, L. M. Sander, and C. M. Schneider-Mizell, J. Stat. Phys. **128**, 209 (2007).

[57] E. Khain, C. M. Schneider-Mizell, M. O. Nowicki, E. A. Chiocca, S. E. Lawler, and L. M. Sander, Europhys. Lett. **88**, 28006 (2009).

[58] M. Aubert, M. Badoual, C. Christov, and B. Grammaticos, J. R. Soc. Interface **5**, 75 (2008).

[59] L. Cronier, S. Crespin, P.-O. Strale, N. Defamie, and M. Mesnil, Antioxidants & Redox Signaling **11**, 323 (2009).

[60] M. Badoual, C. Deroulers, M. Aubert, and B. Grammaticos, Phys. Biol. **7**, 046013 (2010).

[61] R. Baker, C. Yates, and R. Erban, Bulletin of Mathematical Biology **72**, 719 (2010).

[62] J. Jäckle and A. Krönig, J. Phys. Cond. Mat. **6**, 7633 (1994).

[63] A. Krönig and J. Jäckle, J. Phys. Cond. Mat. **6**, 7655 (1994).

[64] A. C. Pan, J. P. Garrahan, and D. Chandler, Phys. Rev. E **72**, 041106 (2005), arXiv:cond-mat/0410525.

[65] L. O. Hedges and J. P. Garrahan, J. Phys. Cond. Mat. **19**, 205124 (2007), arXiv:cond-mat/0610635.

[66] R. Erban, J. Chapman, and P. Maini, "A practical guide to stochastic simulations of reaction-diffusion processes," (2007), arXiv:0704.1908v2.

[67] H. Spohn, J. Phys. A – Math. Gen. **16**, 4275 (1983).

[68] P. L. Garrido, J. L. Lebowitz, C. Maes, and H. Spohn, Phys. Rev. A **42**, 1954 (1990).

[69] S. Asmussen and P. Glynn, *Stochastic Simulation: Algorithms and Analysis*, Stochastic Modelling and Applied Probability (Springer, 2007).

[70] S. Wolfram, *Cellular Automata and Complexity: Collected Papers* (Addison-Wesley, Reading, MA, 1994).

[71] D. Cornforth, D. G. Green, and D. Newth, Physica D Nonlinear Phenomena **204**, 70 (2005).

[72] T. Kitagawa, Mathematical Biosciences **19**, 27 (1974).

[73] T. E. Ingerson and R. L. Buvel, Physica D Nonlinear Phenomena **10**, 59 (1984).

[74] N. Rajewsky, L. Santen, A. Schadschneider, and M. Schreckenberg, Journal of Statistical Physics **92**, 151 (1998).

[75] P. Biane, R. Durrett, and R. Durrett, in *Lectures on Probability Theory*, Lecture Notes in Mathematics, Vol. 1608 (Springer, Berlin / Heidelberg, 1995) pp. 97–201.

[76] G. Barenblatt, *Scaling, Self-similarity, and Intermediate Asymptotics: Dimensional Analysis and Intermediate Asymptotics*, Cambridge Texts in Applied Mathematics (Cambridge University Press, 1996).

[77] M. J. Simpson, R. E. Baker, and S. W. McCue, Phys. Rev. E **83**, 021901 (2011).

[78] T. P. Witelski, Journal of Mathematical Biology **35**, 695 (1997).

[79] P. Sachdev, *Self-Similarity and Beyond: Exact Solutions of Nonlinear Problems*, Chapman & Hall/CRC mono-

graphs and surveys in pure and applied mathematics (Taylor & Francis, UK, 2000).

[80] J. L. Vazquez, *The Porous Medium Equation: Mathematical Theory (Oxford Mathematical Monographs)*, 1st ed. (Oxford University Press, Oxford, 2006).

[81] C. Deroulers and R. Monasson, Phys. Rev. E **69**, 016126 (2004), arXiv:cond-mat/0309637.

[82] D. J. Evans and G. P. Morris, *Statistical Mechanics of Nonequilibrium Liquids* (Cambridge University Press, Cambridge, 1990).

[83] M. Doi, J. Phys. A – Math. Gen. **9**, 1465 (1976).

[84] A. Giese, R. Bjerkvig, M. E. Berens, and M. Westphal, J. Clin. Oncol. **21**, 1624 (2003).

[85] S. Fedotov, A. Iomin, and L. Ryashko, Phys. Rev. E **84**, 061131 (2011).

[86] M. J. Simpson, K. A. Landman, B. D. Hughes, and A. E. Fernando, Physica A **389**, 1412 (2010).

[87] K. A. Landman and A. E. Fernando, Physica A **390**, 3742 (2011).