Confluent and nonconfluent phases in a model of cell tissue

Eial Teomy and David A. Kessler

Department of Physics, Bar-Ilan University, Ramat-Gan 52900, Israel

Herbert Levine

Center for Theoretical Biological Physics, Rice University, Houston, Texas 77005, USA

The Voronoi-based cellular model is highly successful in describing the motion of two-dimensional confluent cell tissues. In the homogeneous version of this model, the energy of each cell is determined solely by its geometric shape and size, and the interaction between adjacent cells is a by-product of this additive energy. We generalize this model so as to allow zero or partial contact between cells. We identify several phases, two of which (solid confluent and liquid confluent) were found in previous studies that imposed confluency as well as others that were not. Transitions in this model may be relevant for understanding both normal development as well as cancer metastasis.

I. INTRODUCTION

In many biological processes, epithelial cells adopt collective organization, stabilized by cell-cell adhesions. The simplest possibility for such a system is a confluent phase where there is no free space between any of the cells, but other types of organization are also possible. For example, individual cells or small cell clusters can, under some conditions, undergo phenotypic transitions and break free of the bulk confluent tissue; this could be relevant for processes such as cancer metastasis [1]. It is clearly of interest to develop a model which can identify under what conditions the system exhibits nonconfluent behavior and what different nonconfluent phases are possible.

In this paper, we consider Voronoi-based models [2–9] (which are closely related to vertex models [3,10–15]), which have often been used to simulate cell tissue. In these models, the shapes of cells in a confluent aggregate is approximated as a convex polygon. Although such an approximation cannot capture all morphologies, it is sufficient in many cases. Most previous studies considered confluent tissues in either periodic [12] or open boundary conditions [4,9,10]. Although periodic boundary conditions may well describe the behavior of cells inside a large cluster, they cannot address the behavior of cells at the boundary of the cluster. Furthermore, in order to account for separation of cells from the main cluster, one needs to go beyond models which focus only on confluent tissues. There have in fact been several works which have allowed for nonconfluent cell arrangements [5–7] in Voronoi-based models. The basic idea is to introduce a length scale, requiring that each cell lie entirely within a distance ℓ of the reference point that describes its location. As a result, the cell boundaries may consist not only of polygonal segments but also circular arcs, and there can be intercellular regions between cells. These works, while introducing this class of model, did not address the possible phases that can be exhibited by such a system.

There are several ways to account for the forces acting by and on the cells. Here we consider an energy functional which depends on the shape of the cells, specifically the quadratic energy functional used in Ref. [16]. However, as we argue later, the specific form of the energy functional is not key. Besides these forces, the cells also perform active motion, which we approximate here as "thermal" fluctuations induced by an effective temperature [17,18]. This, of course, rules out the possibility of organized cell motion, which requires an extension to the model; this will be addressed elsewhere. We first present the model and discuss some basic structures which can serve as building blocks for the organization of the tissue. We then present both simulations and analytic estimates of the zero-temperature phase diagram and close with a discussion of finite-temperature effects.

II. THE MODEL

In Voronoi-based models, each of the N cells is defined by its reference point, $\vec{r}_i$, and it contains all the points which are closest to its reference point. This is equivalent to a Voronoi tessellation of the plane. We consider a two-dimensional model and use the quadratic energy functional used by Refs. [2,16,19], according to which the total energy of the system is given by

$$E = \sum_i K_i (A_i - A_{0,i})^2 + \sum_i \Gamma_i P_i^2 + \sum_{\langle i,j \rangle} \Lambda_{i,j} P_{i,j}, \quad (1)$$

where A_i and P_i are the area and circumference respectively of cell i, $P_{i,j}$ is the length of the interface between cells i and j, $A_{0,i}$ is the target area of cell i, K_i is the elastic modulus of cell i, Γ_i is the contractility coefficient of cell i, and $\Lambda_{i,j}$ is the tension between cells i and j. The first two sums are over all cells, and the third sum is over all pairs of neighboring cells. As the cells are not necessarily confluent, we also need to consider in the third sum the contribution to the energy of interfaces between cells and unoccupied regions. We consider herein the situation where all the cells have the same K_i, $A_{0,i}$, Γ_i, and $\Lambda_{i,j}$. If, in addition, the tensions for cell-cell

interfaces and cell-unoccupied region interfaces are identical, then up to an unimportant additive constant the energy takes the simple form

$$E = K_A \sum_i (A_i - A_0)^2 + K_P \sum_i (P_i - P_0)^2. \quad (2)$$

Physically, it is more appropriate to consider different tensions for cell-cell interfaces and cell-boundary interfaces. However, this does not affect the phase diagram qualitatively but merely moves the location of the phase boundaries, as we show in Appendix E. Therefore, for simplicity, we assume in the following that the cell-cell tension is equal to the cell-unoccupied region tension. On top of these four parameters and the $2N$ degrees of freedom, there is, as noted above, another length scale ℓ representing the maximal radius of each cell. ℓ can be either fixed as a control parameter, or it can be another degree of freedom, subject to optimization to minimize the global energy. This latter case will be referred to as the dynamical ℓ model.

We consider here overdamped dynamics and approximate the active motion of the cells as "thermal" fluctuations induced by an effective temperature T, such that each reference point advances in time according to an overdamped Langevin equation,

$$\frac{d\vec{r}_i}{dt} = -\frac{1}{\xi} \vec{\nabla}_i E + \sqrt{\frac{2k_B T}{\xi}} \vec{\eta}_i(t), \quad (3)$$

where ξ is the effective damping strength, $\vec{\nabla}_i$ stands for the derivative of the energy with respect to $\vec{r}_i$, k_B is the Boltzmann constant, T is the effective temperature, and $\vec{\eta}_i$ is a white noise with zero mean, and

$$\langle \vec{\eta}_i(t) \vec{\eta}_j(t') \rangle = \delta_{i,j} \delta(t - t'). \quad (4)$$

If in addition ℓ is taken to be a free variable which changes so as to minimize the energy, its dynamics are described by $\frac{\partial \ell}{\partial t} = -\zeta \frac{\partial E}{\partial \ell}$. Hence, at equilibrium we require $\partial E/\partial \ell = 0$, or equivalently,

$$0 = 2(\langle \tilde{A}^2 \rangle - \tilde{A}_0 \langle \tilde{A} \rangle) + \tilde{K}(\langle \tilde{P}^2 \rangle - \tilde{P}_0 \langle \tilde{P} \rangle), \quad (5)$$

which follows from rescaling the reference points by ℓ. The quantities with a tilde are dimensionless, rescaled by an appropriate power of the density.

A. No density-independent phase transition in confluent tissues

In previous works on the vertex model and Voronoi-based model under periodic boundary conditions, it was claimed that the system undergoes a density-independent phase transition [8,15]. Here we prove that there is no physical density-independent phase transition in confluent tissues. In order to have a density-independent phase transition, the density should be a control parameter. If the number of cells N is fixed and the tissue is confluent, the density is trivially related to the average area $\langle A \rangle$. We therefore parametrize the density by $\ell = \sqrt{\langle A \rangle}$, such that the rescaled area $\langle \tilde{A} \rangle = 1$.

First, note that under these conditions all the terms in the energy which depend on A_0 do not depend on the free variables, but rather only on the control parameters. Thus, these terms may be considered to be unimportant additive constants.

Hence, the parameter A_0 cannot affect any kind of transition. This means that the phase transition investigated in [8,15] is not governed by the density-independent dimensionless parameter $P_0/\sqrt{A_0}$, but rather by $P_0/\sqrt{\langle A \rangle}$, which is clearly density dependent.

In fact, there are no physical density-independent phase transitions under these conditions. Using Landau's theory of phase transitions, we now show that there are no density-independent phase transitions as long as both K_A and K_P are positive.

First, note that the only dimensionless density-independent parameter is

$$\hat{K} = \frac{K_P}{K_A P_0^2}. \quad (6)$$

We rewrite the energy as

$$\hat{E} = \hat{\ell}^4 \langle \tilde{A}^2 \rangle + \hat{K}(\hat{\ell}^2 \langle \tilde{P}^2 \rangle - 2\hat{\ell} \langle \tilde{P} \rangle), \quad (7)$$

with

$$\hat{E} = \frac{E}{N K_A P_0^4}, \quad \hat{\ell} = \frac{\ell}{P_0}. \quad (8)$$

According to to the Landau theory of phase transitions, near the phase transition the energy density may be approximated by

$$\hat{E} = a + r\Psi^2 + s\Psi^4 + H\Psi, \quad (9)$$

where Ψ is the order parameter controlling the transition. Because of the form of the energy, Eq. (7), the ℓ dependence of the parameter r is of the form

$$r = \hat{\ell}^4 A_r^{(2)} + \hat{K}(\hat{\ell}^2 P_r^{(2)} - 2\hat{\ell} P_r^{(1)}), \quad (10)$$

where

$$A_r^{(2)} = \frac{d^2 \langle \tilde{A}^2 \rangle}{d\Psi^2}\bigg|_{\Psi=0}, \quad P_r^{(2)} = \frac{d^2 \langle \tilde{P}^2 \rangle}{d\Psi^2}\bigg|_{\Psi=0},$$

$$P_r^{(1)} = \frac{d^2 \langle \tilde{P} \rangle}{d\Psi^2}\bigg|_{\Psi=0} \quad (11)$$

with similar expressions for a, s, and H. Note that the coefficients $A_r^{(2)}$, $P_r^{(1)}$, and $P_r^{(2)}$ do not depend on $\hat{K}$ and $\hat{\ell}$. In particular, r is in general a fourth-order polynomial of ℓ. If the phase transition depends only on $\hat{K}$, then r has the form

$$r = r_0(\hat{K} - \hat{K}_c). \quad (12)$$

Combining Eqs. (10) and (12), we find that

$$r_0 = \hat{\ell}^2 P_r^{(2)} - 2\hat{\ell} P_r^{(1)},$$

$$\hat{K}_c = -\frac{\hat{\ell}^4 A_r^{(2)}}{\hat{\ell}^2 P_r^{(2)} - 2\hat{\ell} P_r^{(1)}}. \quad (13)$$

Since $\hat{\ell}$ scales linearly with ℓ, while $A_r^{(2)}$, $P_r^{(2)}$, and $P_r^{(1)}$ are independent of ℓ, the only way for the phase transition to be density independent is to set $\hat{K}_c = 0$. However, since $\hat{K} > 0$ always, there is no such phase transition.

If we relax the physical constraint that $\hat{K} > 0$, a density-independent phase transition could occur. The effects of such an unphysical phase transition might be seen as $\hat{K}$ decreases and the system nears the unobtainable critical point. However, we do not know what order parameter undergoes such a transition.

III. SECOND-ORDER PHASE TRANSITIONS

For a given set of dimensionless parameters $\tilde{P}_0 = P_0/\ell$, $\tilde{A}_0 = A_0/\ell^2$, and $\tilde{K} = K_p/K_A\ell^2$, the system reaches a local minimum of the energy in some particular configuration. We are interested in second-order phase transitions, at which this configuration changes to a local maximum or a saddle point of the energy. These transitions can readily occur even in simple configurations. We consider here two configurations: a honeycomb lattice in a periodic system and a rosette configuration in a system with open boundary conditions.

A. Honeycomb lattice

We first investigate the stability of a honeycomb lattice in an infinite or periodic boundary conditions. Because of the symmetry of the configuration, we find that it satisfies

$$\vec{\nabla}_i \langle \mathcal{O} \rangle = 0, \tag{14}$$

for $\mathcal{O} = A$, A^2, P, and P^2, where $\vec{\nabla}_i$ is the derivative with respect to the location of cell i. Thus, the honeycomb lattice is an extremum of the energy.

In order to check that it is in fact a minimum of the energy, and so a stable configuration, we perturb it such that each particle is located at

$$\vec{r}_i = \vec{r}_i^s + \vec{\epsilon}_i, \tag{15}$$

where $\vec{r}_i^s$ is the location of cell i and all the vectors are normalized by ℓ. Under the harmonic approximation, the force acting on cell i in direction α is

$$F_{i,\alpha} = -\frac{\partial \tilde{E}}{\partial r_{i,\alpha}} = -\sum_{j,\beta} \frac{\partial^2 \tilde{E}}{\partial r_{i,\alpha} \partial r_{j,\beta}}\bigg|_{\vec{r}_i^s, \vec{r}_j^s} \epsilon_{j,\beta}. \tag{16}$$

Using the Fourier transform of $\vec{\epsilon}_i$

$$\vec{\epsilon}_i = \int d\vec{q}\, \vec{C}(\vec{q}) e^{i\vec{q}\cdot\vec{r}_i - \omega(\vec{q})t}, \tag{17}$$

Eq. (16) transforms into

$$F_{i,\alpha} = -\int d\vec{q} \sum_{j,\beta} C_\beta(\vec{q}) \frac{\partial^2 \tilde{E}}{\partial r_{i,\alpha} \partial r_{j,\beta}}\bigg|_{\vec{r}_i^s, \vec{r}_j^s} e^{i\vec{q}\cdot\vec{r}_j - \omega(\vec{q})t}. \tag{18}$$

Using the relation $\vec{F} = \xi\vec{v}$, we find that

$$-\int d\vec{q} \sum_{j,\beta} C_\beta(\vec{q}) \frac{\partial^2 \tilde{E}}{\partial r_{i,\alpha} \partial r_{j,\beta}}\bigg|_{\vec{r}_i^s, \vec{r}_j^s} e^{i\vec{q}\cdot\vec{r}_j}$$
$$= -\xi \int d\vec{q}\, \omega(\vec{q}) C_\alpha(\vec{q}) e^{i\vec{q}\cdot\vec{r}_i}. \tag{19}$$

The system is stable only if $\omega(\vec{q})$ is positive for all $\vec{q}$, or equivalently if the 2×2 matrix $\Omega(\vec{q})$ has only positive eigenvalues for all $\vec{q}$, where

$$\Omega_{\alpha,\beta} = \sum_j \frac{\partial^2 \tilde{E}}{\partial r_{i,\alpha} \partial r_{j,\beta}}\bigg|_{\vec{r}_i^s, \vec{r}_j^s} e^{i\vec{q}\cdot(\vec{r}_j - \vec{r}_i)}. \tag{20}$$

The full form of the matrix Ω for a honeycomb structure appears in Appendix D.

In order for this configuration to be stable, we require that both eigenvalues of the matrix Ω are positive for all values of $\vec{q}$. The two eigenvalues of Ω are

$$\omega_\pm = \frac{\Omega_{x,x} + \Omega_{y,y}}{2} \pm \sqrt{\left(\frac{\Omega_{x,x} - \Omega_{y,y}}{2}\right)^2 + \Omega_{x,y}^2}. \tag{21}$$

Hence, the transition occurs when the minimum of ω_- is zero for some $\vec{q}$. Note that this occurs when the minimum of the determinant of Ω,

$$|\Omega| = \Omega_{x,x}\Omega_{y,y} - \Omega_{x,y}^2, \tag{22}$$

is zero for some $\vec{q}$. The determinant of Ω may be written as

$$|\Omega| = \frac{\tilde{K}(\tilde{P} - \tilde{P}_0)}{18\tilde{P}^2}\{\tilde{P}^3 g_1(\vec{q}) + 23328\tilde{K}[\tilde{P}g_2(\vec{q}) - \tilde{P}_0 g_3(\vec{q})]\}, \tag{23}$$

with the full expression of the functions $g_i(\vec{q})$ appearing in Appendix D.

Straightforward analysis yields $g_1(\vec{q}) \geqslant 0$ and $0 \leqslant g_3(\vec{q}) \leqslant g_2(\vec{q}) \leqslant 1$ for all $\vec{q}$. Therefore, for $\tilde{P} > \tilde{P}_0$ we find that $|\Omega| > 0$ for all $\vec{q}$. Hence, for $\tilde{P} > \tilde{P}_0$ both eigenvalues $\omega_\pm$ have the same sign for all $\vec{q}$. Choosing some arbitrary $\vec{q}$, we find that the sign is positive, and thus the structure is stable for $\tilde{P} > \tilde{P}_0$.

In order to show that the structure is unstable for $\tilde{P} < \tilde{P}_0$, we need to show that in this range either (1) $|\Omega| < 0$ for some $\vec{q}$ or (2) $|\Omega| > 0$ and $\omega_+ < 0$ for all $\vec{q}$. We now analyze the cubic polynomial in $\tilde{P}$

$$G(\tilde{P}) = \tilde{P}^3 g_1(\vec{q}) + 23328\tilde{K}[\tilde{P}g_2(\vec{q}) - \tilde{P}_0 g_3(\vec{q})]. \tag{24}$$

Since $g_1(\vec{q}) \geqslant 0$ and $g_2(\vec{q}) \geqslant 0$ for all $\vec{q}$, the polynomial G is an increasing function of $\tilde{P}$. Since at $\tilde{P} = 0$ the polynomial G is negative, and at $\tilde{P} \to \infty$ the polynomial G goes to infinity, it must have a single root at some positive $\tilde{P}^*(\vec{q})$. We now define $\tilde{P}^\dagger$ by

$$\tilde{P}^\dagger = \min_{\vec{q}} \tilde{P}^*(\vec{q}), \tag{25}$$

such that for $\tilde{P} > \tilde{P}^\dagger$ there is some $\vec{q}$ for which $G > 0$ and thus $|\Omega| < 0$ (as we consider now only the range $\tilde{P} < \tilde{P}_0$). In the range $\tilde{P} < \tilde{P}^\dagger$, the sign of ω_+ is the same for all values of $\tilde{P}$ and $\vec{q}$. Choosing an arbitrary value of $\tilde{P}$ and $\vec{q}$, we find that $\omega_+ < 0$, and thus the structure is unstable for $\tilde{P} < \tilde{P}_0$.

In summary, in a honeycomb lattice in an infinite system, ℓ plays no role, and the calculation recapitulates that of Refs. [16] and [20]. The structure becomes unstable at $\tilde{P}_0 = q_6\sqrt{\tilde{A}_0}$, independently of $\tilde{K}$, where $q_6 = \sqrt{24/\tan(\pi/6)} \approx 3.7$. For the dynamical ℓ model, we additionally impose Eq. (5), and find that at the critical point the mean area satisfies $\langle A \rangle = A_0$.

B. Rosette configuration

A second example is provided by a rosette configuration with open boundary conditions. Specifically, we consider a symmetric rosette configuration in which an isolated set of cells meet at a single point. It was shown in Ref. [21] that if

all the cells have the same energy parameters and $\Gamma = 0$ [see Eq. (1)], then a vertex connecting four cells is unstable. Here, we show that if $\Gamma > 0$ this type of configuration can be stable and there is again a transition.

For simplicity, we consider a configuration containing four cells, all connected at the same point. The four cells are located at $\vec{r}_1 = R\hat{x}, \vec{r}_2 = R\hat{y}, \vec{r}_3 = -R\hat{y}$, and $\vec{r}_4 = -R\hat{x}$, where $R \leqslant \ell$. As a result of symmetry, the configuration is stable if

$$\frac{\partial E}{\partial x} = 0, \quad \frac{\partial^2 E}{\partial x^2} > 0, \quad \frac{\partial^2 E}{\partial y^2} > 0, \quad (26)$$

where x and y are the coordinates of cell 1, without loss of generality. Equation (26) is equivalent to

$$(\langle\tilde{A}\rangle - \tilde{A}_0)\frac{\partial\langle\tilde{A}\rangle}{\partial\tilde{x}} + \tilde{K}(\langle\tilde{P}\rangle - \tilde{P}_0)\frac{\partial\langle\tilde{P}\rangle}{\partial\tilde{x}} = 0,$$

$$(\langle\tilde{A}\rangle - \tilde{A}_0)\frac{\partial^2\langle\tilde{A}\rangle}{\partial\tilde{z}^2} + \tilde{K}(\langle\tilde{P}\rangle - \tilde{P}_0)\frac{\partial^2\langle\tilde{P}\rangle}{\partial\tilde{z}^2}$$

$$+ \left(\frac{\partial\langle\tilde{A}\rangle}{\partial\tilde{z}}\right)^2 + \tilde{K}\left(\frac{\partial\langle\tilde{P}\rangle}{\partial\tilde{z}}\right)^2 > 0, \quad (27)$$

where z in the second equation stands for either x or y.

Explicitly, we find that the mean average and circumference are

$$\langle\tilde{A}\rangle = \frac{\tilde{R}(\tilde{R} + \sqrt{2 - \tilde{R}^2}) + \cos^{-1}(-\tilde{R}\sqrt{2 - \tilde{R}^2})}{2},$$

$$\langle\tilde{P}\rangle = \sqrt{2}(\tilde{R} + \sqrt{2 - \tilde{R}^2}) + \cos^{-1}(-\tilde{R}\sqrt{2 - \tilde{R}^2}). \quad (28)$$

The first derivatives with respect to x are

$$\frac{\partial\langle\tilde{A}\rangle}{\partial\tilde{x}} = \frac{\tilde{R} + \sqrt{2 - \tilde{R}^2}}{4},$$

$$\frac{\partial\langle\tilde{P}\rangle}{\partial\tilde{x}} = \frac{2 - \sqrt{2}}{4} + \frac{\sqrt{2} - \tilde{R}}{2^{3/2}\sqrt{2 - \tilde{R}^2}}. \quad (29)$$

The second derivatives with respect to y are

$$\frac{\partial^2\langle\tilde{A}\rangle}{\partial\tilde{y}^2} = \frac{1 - \tilde{R}^2}{4\tilde{R}\sqrt{2 - \tilde{R}^2}},$$

$$\frac{\partial^2\langle\tilde{P}\rangle}{\partial\tilde{y}^2} = \frac{4 - 3\sqrt{2}}{8\tilde{R}} + \frac{4 - 4\sqrt{2}\tilde{R} + \sqrt{2}\tilde{R}^3}{8\tilde{R}(2 - \tilde{R}^2)^{3/2}}. \quad (30)$$

The second derivatives with respect to x are

$$\frac{\partial^2\langle\tilde{A}\rangle}{\partial\tilde{x}^2} = \frac{1 - \tilde{R}^2}{4\tilde{R}\sqrt{2 - \tilde{R}^2}},$$

$$\frac{\partial^2\langle\tilde{P}\rangle}{\partial\tilde{x}^2} = \frac{1}{8\tilde{R}^2(2 - \tilde{R})^{3/2}(1 + \tilde{R}\sqrt{2 - \tilde{R}^2})^2}\{(1 + \tilde{R}\sqrt{2 - \tilde{R}^2})[2\tilde{R}(13 - 14\tilde{R}^2 + 3\tilde{R}^4) + \sqrt{2}(1 - \tilde{R})(23 + 11\tilde{R} - 27\tilde{R}^2$$

$$- 17\tilde{R}^3 + 9\tilde{R}^4 + 7\tilde{R}^5 - \tilde{R}^6 - \tilde{R}^7)] - (1 - \tilde{R}^2)^2[2\tilde{R}(3 - 2\tilde{R}^3) + \sqrt{2}(1 - \tilde{R})(11 + 7\tilde{R} - 8\tilde{R}^2 - 6\tilde{R}^3 + \tilde{R}^4 + \tilde{R}^5)]\}.$$

$$(31)$$

We solve $\frac{\partial E}{\partial x} = 0$ numerically for $\tilde{R}$ and set the result in the two inequalities for the second derivatives. The configuration is stable if $0 < \tilde{R} < 1$ and the two second derivatives are positive. Figure 1 shows in blue the region in the $\tilde{A}_0 - \tilde{P}_0$ plane for which the configuration is stable for two values of $\tilde{K}$.

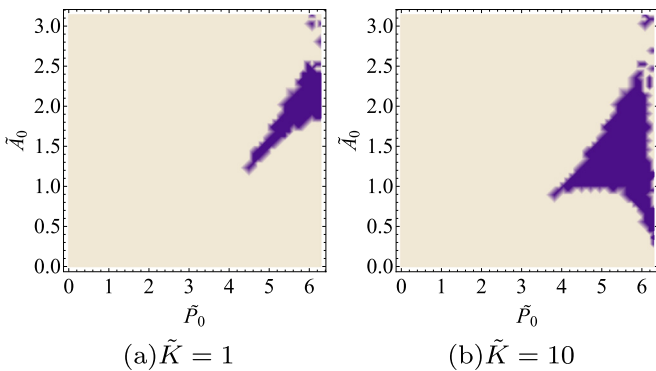(a) $\tilde{K} = 1$ (b) $\tilde{K} = 10$

FIG. 1. A phase diagram in the $\tilde{A}_0 - \tilde{P}_0$ plane showing the stability of the four-cell vertex configuration, for two values of $\tilde{K}$. The configuration is stable in the dark blue regions and unstable in the white regions.

C. Differences between the honeycomb lattice and the rosette pattern

There are several qualitative differences of the rosette pattern from the honeycomb lattice. First is the dependence of the transition line on the value of $K_A/(K_P A_0)$. Second, increasing the value of A_0 causes the rosette configuration to eventually become unstable, while it causes the honeycomb lattice to be more stable.

IV. PHASE DIAGRAM

In our model, with open boundary conditions, the cells do not necessarily compose one connected cluster. We next investigate the different phases exhibited for different $\tilde{P}_0$ and $\tilde{A}_0$ by these connected or unconnected clusters. To begin this study, sweeping through the $\tilde{P}_0$ and $\tilde{A}_0$ plane, with $\tilde{K} = 1$, we ran simulations (for fixed $\ell = 1$) at zero temperature starting with 100 cells (i.e., reference points) placed at random with a density much larger than $1/\ell^2$. We then ran the dynamics until the system reached equilibrium. We found that in different regions of the $\tilde{P}_0$, $\tilde{A}_0$ plane, the system adopted distinctive configurations, which we identify as different phases. This set of typical configurations is shown in Fig. 2. A phase diagram detailing the locations of the different configurations is shown

(a)Gas (**G**) (b)Cluster (**Cl**) (c)Hexagonal (**H**)

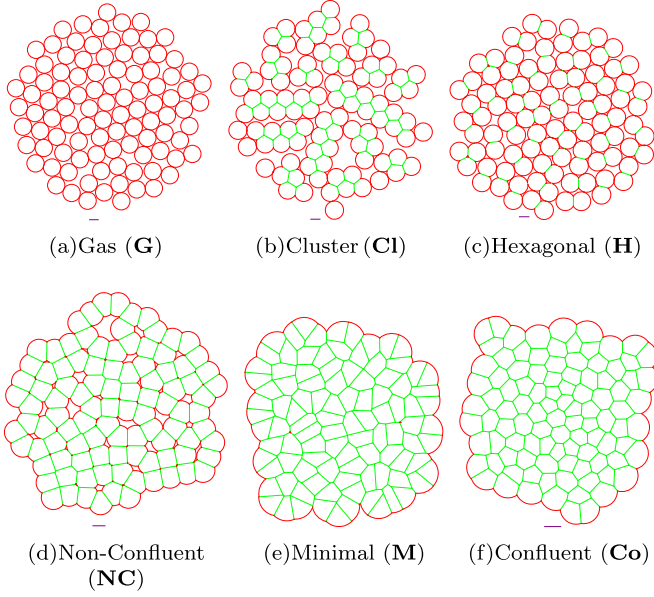(d)Non-Confluent (**NC**) (e)Minimal (**M**) (f)Confluent (**Co**)

FIG. 2. Simulations results for the state of the system after a long time exemplifying each of the phases, starting from the same initial condition with $N = 100$ cells and zero temperature $T = 0$. Green lines are boundaries between adjacent cells, and red lines are the outer boundaries of the cells. The purple line below each configuration is of length $\ell = 1$ to see the different scales. The parameters used are $\tilde{K} = 1$, and (a) $\tilde{A}_0 = 6$ and $\tilde{P}_0 = 5$; (b) $\tilde{A}_0 = 4.5$ and $\tilde{P}_0 = 4.3$; (c) $\tilde{A}_0 = 3.1$ and $\tilde{P}_0 = 6.1$; (d) $\tilde{A}_0 = 1$ and $\tilde{P}_0 = 8$; (e) $\tilde{A}_0 = 1$ and $\tilde{P}_0= 4$; and (f) $\tilde{A}_0 = 4$ and $\tilde{P}_0 = 2.5$.

in Fig. 3. The properties of the configurations obtained from the numerical simulations agree perfectly with the analytically derived phase diagram shown in Fig. 3. We also ran simulations for different values of $\tilde{K}$ and found qualitatively similar results.

We can understand the structure of the phase diagram analytically by considering two-cell interactions, three-cell interactions, and geometrical constraints.
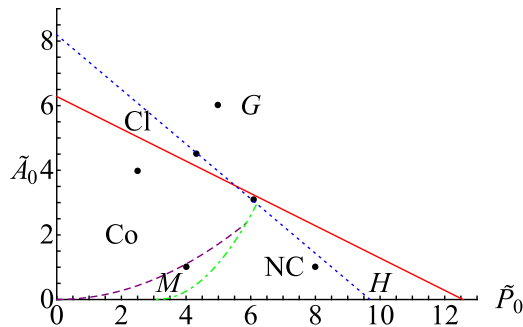

FIG. 3. The phase diagram of fixed open boundary conditions for $\tilde{K} = 1$. It is qualitatively the same for all values of $\tilde{K}$. The letters correspond to the different phases: gas, **G**; cluster, **Cl**; hexagonal, **H**; nonconfluent, **NC**; minimal, **M**; and confluent, **Co**. The lines are boundaries between the phases: Eq. (37) in continuous red, Eq. (40) in dotted blue, $\tilde{P}_0 = q_5\sqrt{\tilde{A}_0}$ in dashed purple, and Eq. (44) in dot-dashed green. The black circles in each phase corresponds to the parameters of the configurations shown in Fig. 2.

A. Two-cell and three-cell interactions

As a rough approximation for two-cell contacts, we consider two cells located at $\vec{r}_\pm = \pm(\ell - \epsilon)\hat{x}$, and for three-cell contacts, we consider three cells located on the vertices of an equilateral triangle with side length $\sqrt{3}(\ell - \epsilon)$, such that when $\epsilon = 0$ the three cells touch at one point. Note that in both cases $\epsilon \geqslant 0$. The three cells are located at

$$\vec{r}_0 = (\ell - \epsilon)\hat{y},$$

$$\vec{r}_\pm = (\ell - \epsilon)\left(-\frac{\hat{y}}{2} \pm \frac{\sqrt{3}\hat{x}}{2}\right). \tag{32}$$

The contacts are attractive if

$$\left.\frac{\partial E}{\partial \epsilon}\right|_{\epsilon=0} < 0. \tag{33}$$

In the two-cell configuration, the area and circumference of each cell are

$$A = (\ell - \epsilon)\sqrt{\ell^2 - (\ell - \epsilon)^2} + \frac{\ell^2\theta}{2},$$

$$P = 2\sqrt{\ell^2 - (\ell - \epsilon)^2} + \ell\theta. \tag{34}$$

The force acting on the cells is

$$\tilde{f} = -2\sqrt{\epsilon(2 - \tilde{\epsilon})}\left[(\tilde{A} - \tilde{A}_0) + \tilde{K}\frac{(\tilde{P} - \tilde{P}_0)}{2 - \tilde{\epsilon}}\right]. \tag{35}$$

At $\epsilon \approx 0$, the force may be approximated by

$$\tilde{f}(\tilde{\epsilon} \approx 0) \approx \sqrt{2\epsilon}[2\tilde{A}_0 - 2\pi + \tilde{K}(\tilde{P}_0 - 2\pi)]. \tag{36}$$

Therefore, the two-cell interaction is attractive if

$$2(\pi - \tilde{A}_0) + \tilde{K}(2\pi - \tilde{P}_0) > 0. \tag{37}$$

In the three-cell configuration, the area and circumference of the cells are

$$A = \frac{\sqrt{3}(\ell - \epsilon)[\ell - \epsilon + \sqrt{4\ell^2 - 3(\ell - \epsilon)^2}]}{4} + \ell^2\theta_3,$$

$$P = \ell - \epsilon + \sqrt{4\ell^2 - 3(\ell - \epsilon)^2} + \ell\theta_3, \tag{38}$$

with

$$\theta_3 = 2\pi - \cos^{-1}\left[\frac{3}{4}\left(1 - \frac{\epsilon}{\ell}\right)^2 - \frac{1}{2}\right.$$
$$\left. - \frac{3}{4}\left(1 - \frac{\epsilon}{\ell}\right)\sqrt{4 - 3\left(1 - \frac{\epsilon}{\ell}\right)^2}\right]. \tag{39}$$

Using Eqs. (33) and (39), we find that the three-cell interaction is attractive if

$$\left(\frac{3\sqrt{3} + 4\pi}{6} - \tilde{A}_0\right)$$
$$+ \frac{2}{3}(3 - \sqrt{3})\tilde{K}\left(\frac{2(3 + 2\pi)}{3} - \tilde{P}_0\right) > 0. \tag{40}$$

Each of the two interactions can be independently either attractive or repulsive. If both of them are repulsive, all the cells repel each other and the system is in the gas phase, marked as **G** in Fig. 3 and exemplified in Fig. 2(a). If the three-cell interaction is attractive and the two-cell interaction is repulsive, then at low temperatures clusters in which all the cells are connected by three-cell contacts are stable. These clusters repel each other, since at low temperatures the only contact between them is via two-cell contacts. There might also be clusters with fingers that do not touch each other. We call this the cluster phase, marked as **Cl** in Fig. 3 and a representative sample of which is displayed in Fig. 2(b). If the three-cell interaction is repulsive and the two-cell interaction is attractive, the cells are connected in a single cluster. Whenever three or more cells meet at a single point, it is always at a distance ℓ from each other. The reason is that if they do get closer, the interactions between them turns from three two-cell interactions to one three-cell interaction, which is repulsive. Because of this constraint, they tend to be arranged in a hexagonal lattice at low temperature, and we thus name this phase the hexagonal phase (**H**), shown in Fig 2(c). If both types of interactions are attractive, the cells aggregate in a single cluster, which may appear in several different phases.

We note in passing that inside the gas phase and near the gas-cluster transition, there are rare stable structures that include more than three cells. However, as these types of stable structures are rare, their net effect on the global parameters is negligible, so we do not consider this a different phase. See Appendix C for an example of such a structure.

B. Geometric constraints

In addition to the key role of the two-cell and three-cell interactions, we identify two other critical $\tilde{K}$-independent curves in the $\tilde{A}_0 - \tilde{P}_0$ space. The first one is connected to the aforementioned transition in the periodic honeycomb lattice. A closely related transition curve was found in confluent models and it lies at $\frac{\tilde{P}_0}{\sqrt{\tilde{A}_0}} = q_5$, where

$$q_n = \sqrt{4n \tan\left(\frac{\pi}{n}\right)} \tag{41}$$

is the shape parameter of a regular polygon with n sides [15,16] and $q_5 \approx 3.8$. The fact that the transition depends on q_5 instead of q_6 is due to the prevalence of pentagonal cell shapes, as seen for example in Fig. 2(f). This has been previously noted (under periodic boundary conditions) both in a Voronoi-based model [8] and in a vertex model [15,16].

To the right of this curve, we find the minimal (**M**) phase. In the minimal phase, we have $\langle \tilde{P} \rangle = \tilde{P}_0$ and $\langle \tilde{A} \rangle = \tilde{A}_0$, and hence the total energy of the system is equal to 0. Note also that in the minimal phase there is an abundance of vertices connecting four or more cells. As discussed above, these rosette-type configurations are stable, and this is correlated with the lack of a $T1$ transition barrier. To the left of this curve is the phase labeled as confluent (**Co**). Here, we find from the simulation that the cells do not satisfy this zero energy condition. This is due to the geometrical constraint determining the minimal area for a given perimeter [20] and
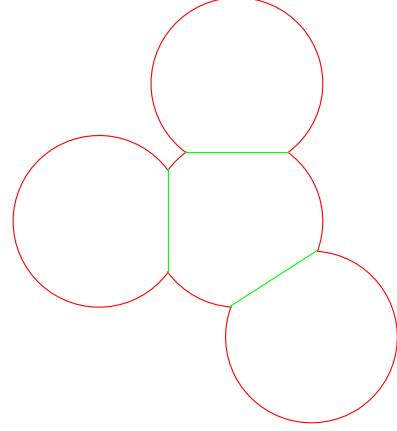

FIG. 4. An illustration of a cell surrounded by three nonoverlapping cells.

here there are many local minima, indicative of frustration. The cells instead aggregate into confluent clusters. We also observe that in this region

$$\langle \tilde{P} \rangle_{in} / \sqrt{\langle \tilde{A} \rangle_{in}} = q_5, \tag{42}$$

where $\langle \cdot \rangle_{in}$ is the average over all cells which are not on the boundary. This is shown in Fig. 6(b). It is straightforward to check that what we have labeled as the cluster phase is always within the overall confluent region.

In a standard vertex model, the **M** phase can extend to large $\tilde{P}_0$. Here, however, a second phase-boundary curve occurs when due to the finite radius of the cells ℓ, they cannot adjust their shape to reach the global energy minimum in which $P_i = P_0$ and $A_i = A_0$. We can describe this curve approximately as the locus of perimeter and area values for a cell around which we place n nonoverlapping other cells (i.e., cells separated by at least 2ℓ) which do overlap with the original cell; see Fig. 4 for an illustration.

The area and circumference of the central cell are

$$\tilde{A} = \pi - n\left[\frac{\cos^{-1}\left(\frac{\tilde{R}^2}{2} - 1\right)}{2} - \frac{\tilde{R}\sqrt{4 - \tilde{R}^2}}{4}\right],$$

$$\tilde{P} = 2\pi - n\left[\cos^{-1}\left(\frac{\tilde{R}^2}{2} - 1\right) - \sqrt{4 - \tilde{R}^2}\right]. \tag{43}$$

Figure 5 is a parametric plot of $\tilde{P}$ versus $\sqrt{\tilde{A}}$ as $\tilde{R}$ is varied from 0 to 2, for values of n ranging from 2 to 6. In the region of interest, we find that this relation is very well approximated by

$$\tilde{P} \approx 3.09 + 1.80\sqrt{\tilde{A}}. \tag{44}$$

Beyond this second curve lies the region we refer to as the nonconfluent (**NC**) phase, shown in Fig. 2(d). Inside the region, we find that $\langle \tilde{P} \rangle$ and $\langle \tilde{A} \rangle$ satisfy Eq. (5). This is shown in Fig. 6(a).

C. Order parameters

The transitions between the different phases can also be quantified by looking at the discontinuities of the derivatives
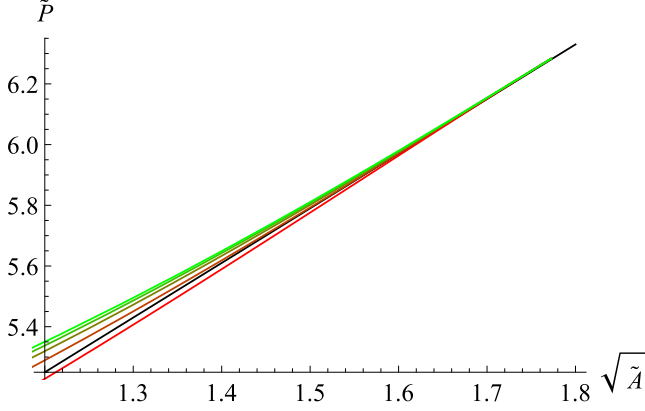
FIG. 5. The relation between the circumference $\tilde{P}$ and the square root of the area, $\tilde{A}$, for a cell in the nonconfluent phase under the mean field approximation. The colored lines correspond to different values of n from $n = 2$ (red at the bottom) to $n = 6$ (green at the top). The black line (second from the bottom) is $\tilde{P} = 3.09 + 1.80\sqrt{\tilde{A}}$.

of various order parameters with respect to the control parameters, $\tilde{A}_0$, $\tilde{P}_0$, and $\tilde{K}$. The minimal-confluent and minimal-nonconfluent transitions involve a discontinuity in the derivative of the energy, since in the minimal phase the energy is equal to zero, while in the other two phases it depends nontrivially on the control parameters.

In order to approximate the energy in the confluent and nonconfluent phases, we note that we numerically found that in those phases the variances of the area and circumference are very small, and we thus assume that $\langle \tilde{P}^2 \rangle \approx \langle \tilde{P} \rangle^2$ and $\langle \tilde{A}^2 \rangle \approx \langle \tilde{A} \rangle^2$. Since we know the relation between the mean circumference and area, we may write the total energy in terms of $\langle \tilde{A} \rangle$ and minimize it. In the confluent phase, the area-perimeter relationship is taken to be that of a perfect pentagon, while in the nonconfluent phase we use Eq. (44). We find that the mean circumference is the solution of the
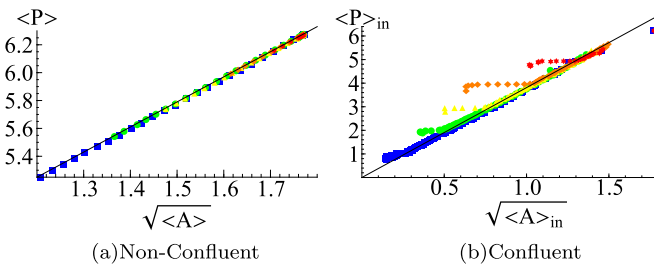

FIG. 6. Simulation results showing the behavior of the (a) non-confluent and (b) confluent phases, as a parametric plot of the mean circumference of (a) all cells $\langle \tilde{P} \rangle$ or (b) the inner cells $\langle \tilde{P} \rangle_{in}$, vs the square root of the mean area of (a) all cells $\langle \tilde{A} \rangle^{1/2}$ or (b) the inner cells $\langle \tilde{A} \rangle_{in}^{1/2}$. In both plots $\tilde{K} = 1$. The black lines are (a) $\langle \tilde{P} \rangle = 3.09 + 1.80 \langle \tilde{A} \rangle^{1/2}$ and (b) $\langle \tilde{P} \rangle_{in} = q_5 \langle \tilde{A} \rangle_{in}^{1/2}$. The symbols correspond to different values of $\tilde{P}_0$: (a) $\tilde{P} = 6$ or (b) $\tilde{P} = 1$ (blue squares), (a) $\tilde{P} = 7$ or (b) $\tilde{P} = 2$ (green circles), (a) $\tilde{P} = 8$ or (b) $\tilde{P} = 3$ (yellow triangles), (a) $\tilde{P} = 9$ or (b) $\tilde{P} = 4$ (orange diamonds), and (a) $\tilde{P} = 10$ or (b) $\tilde{P} = 5$ (red stars).
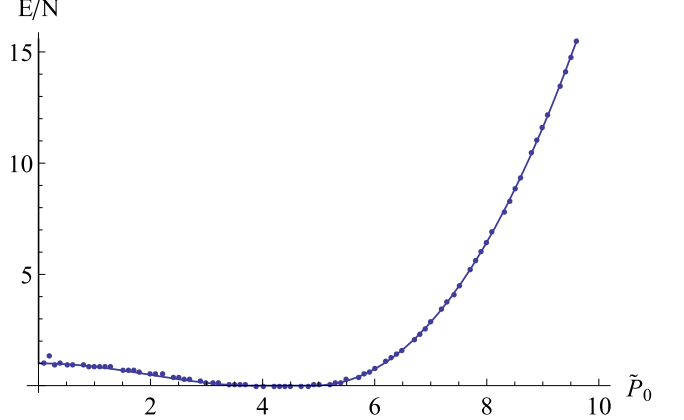

FIG. 7. The mean energy per cell vs $\tilde{P}_0$ for $\tilde{K} = \tilde{A}_0 = 1$. The symbols are the numerical results, and the continuous line is the analytical expression obtained from Eqs. (2), (45), and (46).

cubic equation

$$\frac{2}{c_2^4}(\langle \tilde{P} \rangle - c_1)^3 + \left(\tilde{K} - \frac{2\tilde{A}_0}{c_2^2} \right)(\langle \tilde{P} \rangle - c_1) - \tilde{K}(\tilde{P}_0 - c_1) = 0,$$
(45)

and the mean area is the solution of the equation

$$2\langle \tilde{A} \rangle - 2\tilde{A}_0 + \tilde{K}c_2^2 - \frac{c_2 \tilde{K}(\tilde{P}_0 - c_1)}{\sqrt{\langle \tilde{A} \rangle}} = 0,$$
(46)

where in the confluent phase $c_1 = 0$ and $c_2 = q_5$, while in the nonconfluent phase $c_1 = 3.09$ and $c_2 = 1.80$. As shown in Fig. 7, we find a nearly perfect agreement with the numerical results in the nonconfluent phase and a lower bound on the energy in the confluent phase. The latter is to be expected, as the approximate treatment as perfect pentagons invariably breaks down near the boundaries, where the cells have higher energies.

The confluent-cluster, cluster-gas, and hexagonal-gas transitions can be quantified by looking at the number of clusters which grows from 1 in the confluent and hexagonal phases, to N in the gas phase, and is large and parameter dependent in the cluster phase. Figure 8 shows how the number of clusters grows in the transition from the confluent phase to the gas phase. The transition between the nonconfluent and the hexagonal phases can be quantified via examining the number of vertices connecting three or more cells (finite in the nonconfluent phase and zero in the hexagonal phases), as shown in Fig. 9.

D. The phase diagram in the dynamical ℓ model

In the dynamical ℓ model, in which the maximum radius of the cells can change in order to minimize the energy, the phase diagram is much simpler. For the two-cell and three-cell interactions, we combine Eqs. (5), (37), and (40) and find that the two-cell interaction is attractive if $P_0/\sqrt{A_0} > 2\sqrt{\pi} \approx 3.54$ and that the three-cell interaction is attractive if

$$\frac{P_0}{\sqrt{A_0}} > \frac{2(3 + 2\pi)}{\sqrt{\frac{9\sqrt{3}}{2} + 6\pi}} \approx 3.6.$$
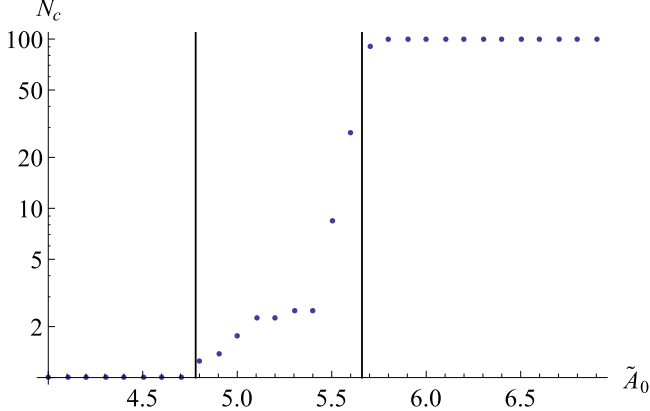(47)

FIG. 8. The number of clusters, N_c, in a system with $N = 100$ cells for $\tilde{K} = 1$ and $\tilde{P}_0 = 3$. The two black lines designate the theoretical transition points between the confluent and cluster phases ($\tilde{A}_0 = 4.78$) and between the cluster and gas phases ($\tilde{A}_0 = 5.66$). Symbols are averages over eight simulations.

In the region $P_0/\sqrt{A_0} < 3.54$, both interactions are repulsive, and thus the system is in the gas phase. For $3.54 < P_0/\sqrt{A_0} < 3.6$, the two-cell interaction is attractive, while the three-cell interaction is repulsive; hence, the system is in the hexagonal phase. In the region $3.6 < P_0/\sqrt{A_0} < q_5 \approx 3.8$ the system is in the confluent phase, due to the same geometric constraints that occur in all related models [20]. Finally, in the region $P_0/\sqrt{A_0} > 3.8$ the system is in the minimal phase. The phase diagram of the dynamical ℓ model is illustrated in Fig. 10.

E. Generalization to other models

We emphasize here that the boundaries between the different phases remain qualitatively the same irrespective of the specific energy functional chosen, since the phases and the boundaries between them can be described in general terms. To summarize our overall picture: The gas phase occurs when both two-cell and three-cell interactions are repulsive. The
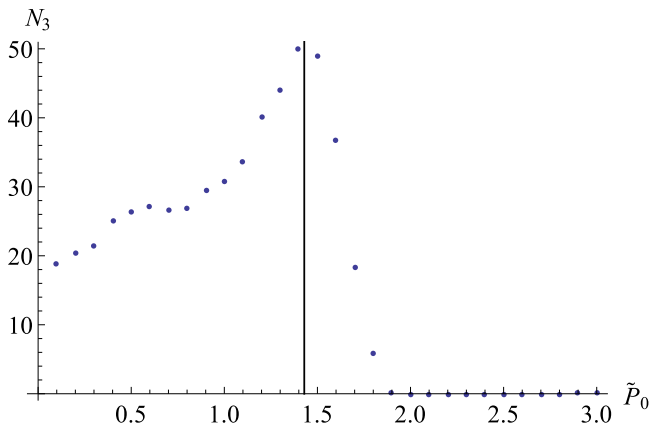

FIG. 9. The number of vertices connecting three cells, N_3, as a function of $\tilde{P}_0$ for $\tilde{K} = 1$ and $\tilde{A}_0 = 8$. The black line shows the theoretical transition point between the nonconfluent phase ($\tilde{P}_0 \lesssim 1.43$) and the hexagonal phase ($\tilde{P}_0 \gtrsim 1.43$). Symbols are averages over 60 simulations.
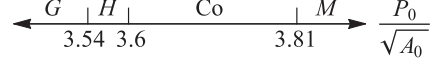

FIG. 10. The phase diagram of the dynamical ℓ model, which depends only on the value of the shape parameter $P_0/\sqrt{A_0}$. The letters correspond to the gas, hexagonal, confluent, and minimal phases.

cluster phase occurs when the two-cell interaction is repulsive while the three-cell interaction is attractive. The hexagonal phase occurs when the two-cell interaction is attractive while the three-cell interaction is repulsive. The confluent and nonconfluent phases occur when the local minimum is for each cell to have identical neighboring cells connected in a confluent or nonconfluent tissue respectively. The minimal phase arises whenever there is enough flexibility for the system to reach a perfect state with zero energy.

V. FINITE EFFECTIVE TEMPERATURE

The effect of the temperature on the phase diagram is rather weak. However, the temperature does give information about the different phases. By expanding the energy to second order in the deviation of the reference points from the global energy minimum obtained at $T = 0$, we expect that at low temperatures the energy per cell, ϵ, may be approximated by

$$\epsilon(T) = \epsilon(T = 0) + N_f T, \tag{48}$$

where N_f is the number of nonzero eigenvalues per particle of the Hessian matrix. See Appendix F for the next order expansion of the energy. In all phases except the confluent phase, we find that at $T = 0$ the system indeed relaxes to the global energy minimum and Eq. (48) is valid. Simulation results for the mean energy per cell as a function of the temperature in the different phases is shown in Fig. 11. In all
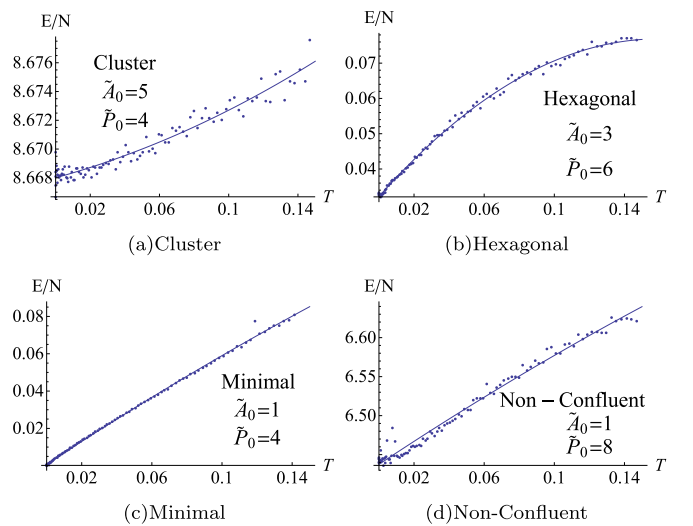

FIG. 11. The mean energy per cell E/N as a function of the temperature T for a representative choice of parameters $\tilde{A}_0$ and $\tilde{P}_0$ for the cluster, hexagonal, minimal, and nonconfluent phases. In all cases, $\tilde{K} = 1$. Each point in the plot represents the results of a single realization. The continuous lines are fits to a quadratic form.
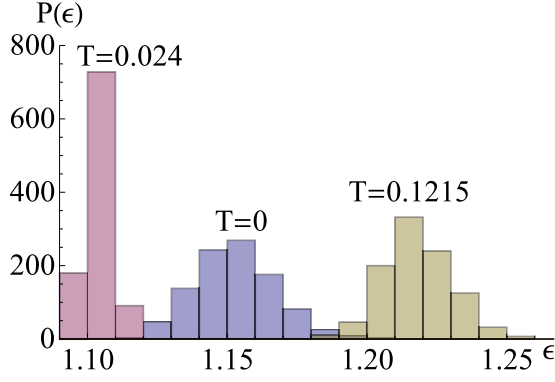
FIG. 12. Histogram of the energy minima the system relaxes to after a very long time in the confluent phase for three different temperatures and a specific set of parameters ($\tilde{K} = 1$, $\tilde{A}_0 = 2$, $\tilde{P}_0 = 3.5$). The data include 1000 realizations. At $T = 0$, the wide distribution shows the many energy minima the system can relax to. At $T \approx 0.024$, the mean energy is minimal for this set of parameters, and the distribution is very narrow. At $T = 0.1215$, the wide distribution is due to the thermal fluctuations around the global minimum.

FIG. 13. The value of the exponent α describing the relaxation of the energy per cell with time, $\epsilon - \epsilon_{\min} \propto t^\alpha$, as a function of the temperature for $\tilde{K} = 1$, $\tilde{A}_0 = 2$, and $\tilde{P}_0 = 3.5$.

phases except the confluent phase, the energy increases with the temperature and agrees at low temperature with Eq. (48).

In the gas and cluster phases, we find that $N_f \approx 0.1$. We discuss later why the cluster phase behaves similarly to the gas phase. Note that $N_f > 0$ because the cells still interact occasionally. In the minimal phase, $N_f \approx 0.5$–0.8, which shows the large freedom of movement the cells have within the aggregate. In the hexagonal, confluent and nonconfluent phases, we find that $N_f \approx 1.3$–1.8, which shows that, in order for a cell to move, it must overcome energy barriers. From these results, we conclude that the gas and cluster phases behave as a gas, the minimal phase behaves as a liquid, while the hexagonal and nonconfluent phases behave as a solid. As will be shown below, the confluent phase behaves as a glass.

In the confluent phase, the system relaxes to only a local minimum of the energy, whose value depends on the initial conditions, as shown in Fig. 12. As the temperature is increased, the mean energy actually decreases because the thermal fluctuations allow the system to overcome some of the energy barriers between those local minima and relax to different, lower minima. At a sufficiently high temperature, the system behaves effectively as a thermal system fluctuating around the global minimum. Strictly speaking, this is not a phase transition because there is no discontinuity. Note that although at low temperatures each realization relaxes to a certain configuration, the mean energy per cell continues to relax very slowly to the global minimum as $\epsilon - \epsilon_{\min} \propto t^\alpha$, with the dependence of α on the temperature shown in Fig. 13. The value of the exponent α reaches a minimum of about 0.23 at around $T \approx 0.25$, which is the same temperature at which the mean energy is minimal.

In the cluster phase, the temperature has a critical effect. At a sufficiently high temperature, the three-cell contacts are broken by the thermal fluctuations. Because the two-cell interaction is repulsive, the cells cannot reconnect, and this causes the phase transition to the gas phase. Although the thermal fluctuations also break the contacts in the other phases, the
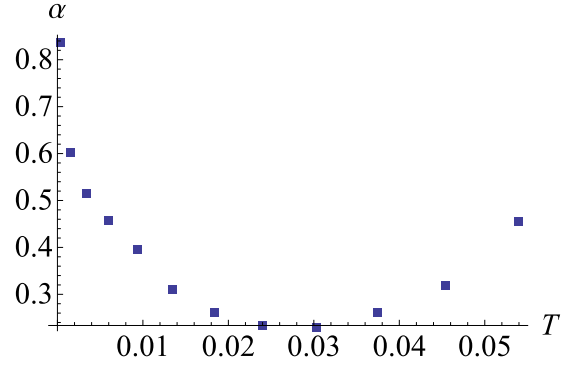
transition there should be continuous because the two-cell interaction is attractive in all the other phases (except for the gas phase) and thus the cells can reconnect. We present in the Supplemental Material [22] two videos of a system in the cluster phase ($\tilde{K} = 1$, $\tilde{A}_0 = 4.5$, and $\tilde{P}_0 = 4.3$) with the same initial condition at different temperatures ($T = 6 \times 10^{-7}$ and $T = 9.375 \times 10^{-7}$). At the lower temperature, the cells mostly vibrate in place, while at the higher temperature, the clusters are broken and do not reconnect. Note that even the higher temperature is extremely small compared to the temperature range we consider elsewhere.

In each of the phases, there are different properties which are robust to the thermal fluctuations. In order to see them, we divide the energy per cell, Eq. (2), into four parts:

$$\tilde{\epsilon} = \tilde{A}_{\text{var}} + \tilde{A}_{\text{mean}} + \tilde{P}_{\text{var}} + \tilde{P}_{\text{mean}}, \tag{49}$$

with

$$\tilde{A}_{\text{var}} = \langle \tilde{A}^2 \rangle - \langle \tilde{A} \rangle^2, \quad \tilde{P}_{\text{var}} = \tilde{K}(\langle \tilde{P}^2 \rangle - \langle \tilde{P} \rangle^2),$$
$$\tilde{A}_{\text{mean}} = (\langle \tilde{A} \rangle - \tilde{A}_0)^2, \quad \tilde{P}_{\text{mean}} = \tilde{K}(\langle \tilde{P} \rangle - \tilde{P}_0)^2. \tag{50}$$

Figure 14 shows how each part depends on the temperature in the different phases. We find that far from the minimal phase, the variance terms, $\tilde{A}_{\text{var}}$ and $\tilde{P}_{\text{var}}$, are not affected by the temperature. Only in the minimal phase can these two terms be the dominant terms in the energy. We also find that only in the confluent phase $\tilde{A}_{\text{mean}}$ decreases with temperature and $\tilde{P}_{\text{mean}}$ increases with temperature, while the opposite behavior is observed at the cluster, nonconfluent, and hexagonal phases, namely that $\tilde{A}_{\text{mean}}$ increases with temperature and $\tilde{P}_{\text{mean}}$ decreases. Geometrically, this means that in the confluent phase increasing the temperature causes the cells to elongate, while in the cluster, nonconfluent, and hexagonal phases, increasing the temperature evidently causes the cells to become more compact.

VI. SUMMARY

In this paper, we investigated a Voronoi-based cellular model at thermal equilibrium, in which the total energy of the system is given by the area and circumference of each cell and their distance from a global preferred area and circumference. By requiring that each cell be completely within a distance
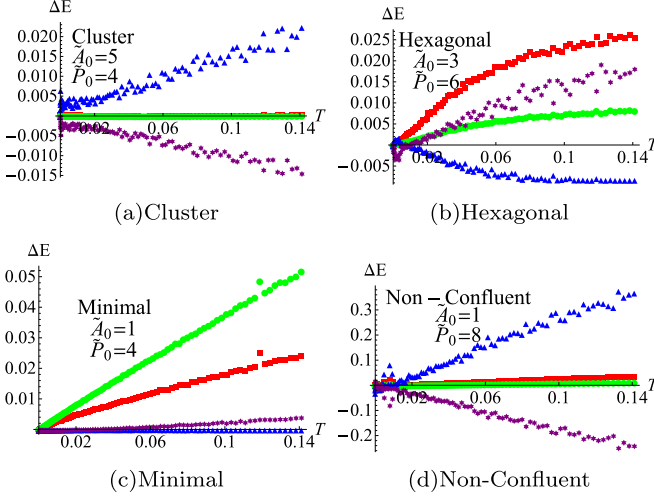
FIG. 14. The deviation of each of the four parts of the energy, Eq. (50), from its value at $T = 0$ as a function of the temperature for a representative choice of parameters $\tilde{A}_0$ and $\tilde{P}_0$ for the cluster, hexagonal, minimal, and nonconfluent phases. In all cases, $\tilde{K} = 1$. Each point in the plot represents the results of a single realization. Each color represents a different part: $\tilde{A}_{\mathrm{var}}$ in red squares, $\tilde{P}_{\mathrm{var}}$ in green circles, $\tilde{A}_{\mathrm{mean}}$ in blue triangles, and $\tilde{P}_{\mathrm{mean}}$ in purple stars.

ℓ of its reference point, the model allows for nonconfluent as well as confluent configurations. In open boundary conditions, we found six different phases, which occur due to competition between several factors: two-cell interactions, three-cell interactions, and geometric constraints. Previous studies, which constrained the cells to form a confluent tissue, identified two of these phases. All the other transitions cannot occur when the tissue is forced to be confluent. Furthermore, we argue that these phases are generic to open boundary models, regardless of the specific choice of energy functional. We expect that these transitions will help us understand the possible behaviors of cells in wound healing and in cancer metastasis, both cases where the tissue necessarily has an outer interface.

We approximated the active motion of cells as fluctuations induced by an effective temperature. Although this is a very crude approximation which cannot account for organized cell motion, amongst other things, our main focus in the paper is on the phase diagram of the system without activity. This approximation, however, allows us to probe two aspects of the system. First, we analyzed the Hessian and found that the minimal phase is liquid-like, since the Hessian has many zero eigenvalues, while the hexagonal and nonconfluent phases are solidlike. Additionally, we investigated the energy landscape of the confluent phase and showed that it corresponds to glassy systems with many metastable states.

In a way, our work bridges the gap between several related works. The confluent and minimal phases are the two phases identified in works in which the confluency of the tissue is imposed, especially in Ref. [9], which considers a confluent tissue under open boundary conditions. A more recent work considers deformable repulsive cells under compression [23], which complements the gas and confluent phases and exhibits another way (compression) to achieve a phase transition.

The transition between the confluent and minimal phases has been seen in many experimental systems. The gas phase appears when the cells are fully repulsive and detach from each other; see, for example, Ref. [24]. To our knowledge, the other three phases (cluster, nonconfluent, and hexagonal) have not been identified in experiment, perhaps in part since they had not been characterized as being distinct from the confluent, minimal, and gas phases. We note that in the nonconfluent phase there are many almost degenerate pairs of three-cell vertices which effectively appear to be four-cell vertices, similar to what is seen in Ref. [25]. The cluster phase, as we have shown, is extremely sensitive to fluctuations, and thus dynamical measurements are needed to identify it. It general it may be simpler to identify the new phases from their dynamical behavior rather than their structure, especially in the presence of noise. We also note that the difference between interior cells, with their polygonal shape, and the outer cells, with their rounded outer-facing boundaries, is readily apparent in typical pictures of tissues (see, e.g., Fig. 4(h) from Ref. [26]).

In this study, we considered only two-dimensional systems. It would be interesting to expand our work to three-dimensional systems. We expect the phase diagram to be richer, since in three dimensions the basic vertex in a confluent aggregate connects four cells, which adds another line to the phase diagram and may divide the existing phases further. Other ways to extend this study includes adding self-propulsion to the cells, considering cells of different phenotypes with different interactions, and considering elliptically bounded cells.

APPENDIX A: SIMULATION DETAILS

In the numerical simulations, we used the Euler method to advance the evolution equation, Eq. (3), for $T = 0$, and the Euler-Maruyama method for $T > 0$. We set the time step to $\Delta t = 0.001$. After each time step, we first produced the Voronoi diagram using the plane-sweep algorithm [27] and then adjusted it to account for the finite radius of the cells. At $t = 0$, the cells are randomly placed in a square of size $\ell \times \ell$.

The time until the system equilibrates varies, depending on the phase. In the minimal, cluster, hexagonal, and nonconfluent phases, the equilibration is fast at about 1–10 time units (i.e., 10^3–10^4 time steps). In the gas phase, the system nearly equilibrates after 10^3 time steps, but there are small two-cell repulsive contacts that can prolong the last steps of the equilibration by two or more orders of magnitude. In most cases, we did not wait for final equilibration in the gas phase. In the confluent phase, the equilibration is very slow, since the system is in a glassy state. Depending on how deep inside the confluent phase the system is, the equilibration can take from 10^4 to 10^8 time steps, or even more.

APPENDIX B: CALCULATION OF THE FORCES

In the simulations, we randomly generate the initial locations of the reference points of the N cells and let them relax to a stable configuration. In order to do this, we calculate the forces acting on each cell according to

$$\vec{f}_i = -\frac{\partial \tilde{E}}{\partial \vec{r}_i}, \tag{B1}$$

where $\vec{r}_i$ is the location of the reference point of the cell. We assume overdamped dynamics such that the velocity of each cell is proportional to the force acting on it.

Under open boundary conditions, the shape of the cells is not determined solely by a Voronoi tessellation of the plane, as this would generate cells with infinite area. The cells have a maximum radius ℓ, such that each cell contains all the points which are closest to its reference point, $\vec{r}_i$ and are at most a distance of ℓ from it [28].

The location of a vertex connecting three cells i, j, and k, $\vec{h}^{in}$, is given by

$$\vec{h}^{in} = \alpha_i^{jk}\vec{r}_i + \alpha_j^{ki}\vec{r}_j + \alpha_k^{ij}\vec{r}_k, \tag{B2}$$

where

$$\alpha_i^{jk} = \frac{|\vec{r}_j - \vec{r}_k|^2(\vec{r}_i - \vec{r}_j)\cdot(\vec{r}_i - \vec{r}_k)}{2|(\vec{r}_i - \vec{r}_j)\times(\vec{r}_j - \vec{r}_k)|^2}. \tag{B3}$$

An outer vertex connecting two cells i and j is located at one of the two points which are at a distance ℓ from both $\vec{r}_i$ and $\vec{r}_j$. It is thus located at one of the two points

$$\vec{h}^{out} = \frac{\vec{r}_i + \vec{r}_j}{2} \pm \frac{\sqrt{4\ell^2 - |\vec{r}_i - \vec{r}_j|^2}}{2|\vec{r}_i - \vec{r}_j|}(\vec{r}_i - \vec{r}_j)\times\hat{z}. \tag{B4}$$

In order to calculate the perimeter and area of cell i, we enumerate the vertices surrounding it clockwise. The area and perimeter are given by

$$P_i = \sum_m P_{m,i}, \quad A_i = \sum_m A_{m,i}, \tag{B5}$$

where the summation variable indicates the vertices surrounding cell i. If at least one of the two vertices m and $m+1$ is an inner vertex, then $P_{m,i} = |\vec{h}_m - \vec{h}_{m+1}|$, and

$$A_{m,i} = \frac{1}{2}|\vec{h}_m\times\vec{h}_{m+1} + \vec{r}_i\times(\vec{h}_m - \vec{h}_{m+1})|. \tag{B6}$$

If both vertices m and $m+1$ are outer vertices, than $P_{m,i} = \ell\theta_m$ and $A_{m,i} = \frac{\ell^2\theta_m}{2}$, where

$$\theta_m = \begin{cases} \theta_m^{(0)} & \theta_m^{(0)} < \pi \\ 2\pi - \theta_m^{(0)} & \theta_m^{(0)} > \pi \end{cases} \tag{B7}$$

with

$$\theta_m^{(0)} = \cos^{-1}\left[\frac{(\vec{h}_m - \vec{r}_i)\cdot(\vec{h}_{m+1} - \vec{r}_i)}{\ell^2}\right]. \tag{B8}$$

The force acting on cell i in direction x may therefore be written as

$$f_{i,x} = -\sum_m \frac{\partial E}{\partial \vec{h}_m}\cdot\frac{\partial \vec{h}_m}{\partial x_i} - \frac{\partial E}{\partial x_i}, \tag{B9}$$

where the sum is over all vertices connected to cell i, and x_i is the x coordinate of cell i. We obtain an analogous expression for the force acting in the y direction.

The derivatives $\frac{\partial \vec{h}_m}{\partial x_i}$ depend on whether the vertex is an inner vertex or an outer vertex. For inner vertices, we use Eq. (B2) and find that

$$\frac{\partial \vec{h}^{in}}{\partial x_i} = \alpha_i^{j,k}\hat{x} + \frac{\partial\alpha_j^{i,k}}{\partial x_i}\vec{r}_j + \frac{\partial\alpha_k^{i,j}}{\partial x_i}\vec{r}_k, \tag{B10}$$

where j and k are the other two cells connected to the vertex. Using Eq. (B3), we find that

$$\frac{\partial\alpha_j^{i,k}}{\partial\vec{r}_i} = \alpha_j^{i,k}\left[\frac{\vec{r}_k - \vec{r}_j}{(\vec{r}_j - \vec{r}_i)\cdot(\vec{r}_j - \vec{r}_k)} + 2\frac{\vec{r}_i - \vec{r}_k}{|\vec{r}_i - \vec{r}_k|^2}\right. $$
$$\left. - 2\frac{(\vec{r}_k - \vec{r}_j)\times\hat{z}}{|(\vec{r}_i - \vec{r}_j)\times(\vec{r}_j - \vec{r}_k)|}\right]. \tag{B11}$$

For the outer vertices, we use Eq. (B4) and find that

$$\frac{\partial\vec{h}^{out}}{\partial x_i} = \frac{\hat{x}}{2} \mp \frac{2\ell^2(x_i - x_j)[(y_i - y_j)\hat{x} - (x_i - x_j)\hat{y}]}{\sqrt{4\ell^2 - |\vec{r}_i - \vec{r}_j|^2}|\vec{r}_i - \vec{r}_j|^3}$$
$$\mp \sqrt{\frac{\ell^2}{|\vec{r}_i - \vec{r}_j|^2} - \frac{1}{4}}\hat{y}. \tag{B12}$$

The derivative of the energy with respect to the location of the vertices is

$$\frac{\partial E}{\partial \vec{h}} = 2\sum_i K_A(A_i - A_0)\frac{\partial A_i}{\partial \vec{h}} + K_P(P_i - P_0)\frac{\partial P_i}{\partial \vec{h}}, \tag{B13}$$

where the sum is over the cells connected to the vertex. Using Eq. (B5), the derivative of the area with respect to $\vec{h}_m$ is

$$\frac{\partial A_i}{\partial\vec{h}_m} = \frac{\partial A_{m,i}}{\partial\vec{h}_m} + \frac{\partial A_{m-1,i}}{\partial\vec{h}_m}, \tag{B14}$$

and similarly for the circumference. For an edge connecting two cells these derivatives are

$$\frac{\partial A_{m,i}}{\partial\vec{h}_m} = \frac{(\vec{h}_{m+1} - \vec{r}_i)\times\hat{z}}{2}$$
$$\times \text{sgn}\{\hat{z}\cdot[\vec{h}_m\times\vec{h}_{m+1} + \vec{r}_i\cdot(\vec{h}_m - \vec{h}_{m+1})]\},$$

$$\frac{\partial A_{m-1,i}}{\partial\vec{h}_m} = \frac{(\vec{h}_{m-1} - \vec{r}_i)\times\hat{z}}{2}$$
$$\times \text{sgn}\{\hat{z}\cdot[\vec{h}_m\times\vec{h}_{m-1} + \vec{r}_i\cdot(\vec{h}_m - \vec{h}_{m-1})]\},$$

$$\frac{\partial P_{m,i}}{\partial\vec{h}_m} = \frac{\vec{h}_m - \vec{h}_{m+1}}{|\vec{h}_m - \vec{h}_{m+1}|},$$

$$\frac{\partial P_{m-1,i}}{\partial\vec{h}_m} = \frac{\vec{h}_m - \vec{h}_{m-1}}{|\vec{h}_m - \vec{h}_{m-1}|}. \tag{B15}$$

For an edge connecting a cell to the unoccupied region, we use Eq. (B8) and find that

$$\frac{\partial\theta_m^{(0)}}{\partial\vec{h}_m} = -\frac{\vec{h}_{m+1} - \vec{r}_i}{\sin\theta_m^{(0)}}. \tag{B16}$$

The only place in which the explicit derivative of E with respect to $\vec{r}_i$ in Eq. (B9) does not vanish is at the boundaries
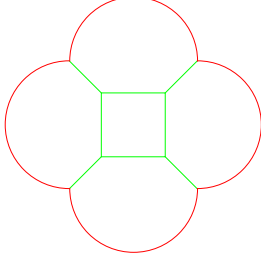
FIG. 15. An example of a rare five-cell structure which is stable even when three-cell interactions are repulsive.

between a cell and the unoccupied region, the contribution from which is straightforward to evaluate by using

$$
\frac{\partial \theta_m^{(0)}}{\partial \vec{r}_i} = \frac{\vec{h}_m + \vec{h}_{m+1} - 2\vec{r}_i}{\sin \theta_m^{(0)}}. \tag{B17}
$$

APPENDIX C: STABILITY OF FIVE CELL CONFIGURATION INSIDE THE GAS PHASE

Within the gas phase, where both the two-cell and three-cell interaction are repulsive, there are several stable configurations. Here we analyze one of them.

Consider the five-cell structure shown in Fig. 15. The central cell is located at the origin, and the four outer cells

are located at $\vec{r}_1 = R\hat{x}$, $\vec{r}_2 = R\hat{y}$, $\vec{r}_3 = -R\hat{x}$, and $\vec{r}_4 = -R\hat{y}$. The area and circumference of the central cell are

$$
A_{\text{cen}} = R^2, \qquad P_{\text{cen}} = 4R, \tag{C1}
$$

and the area and circumference of the four outer cells are

$$
\begin{aligned}
A_{\text{out}} &= \tfrac{1}{4}[R(R + 2\sqrt{2\ell^2 - R^2}) + 2\ell^2 \theta_{\text{out}} \\
&\quad + 4\ell^2(\pi - \theta_{\text{out}})H(R - \ell)], \\
P_{\text{out}} &= R + \sqrt{4\ell^2 - 2R^2} + \ell \theta_{\text{out}} + 2\ell(\pi - \theta_{\text{out}})H(R - \ell),
\end{aligned} \tag{C2}
$$

where $H(z)$ is the Heaviside step function, and

$$
\theta_{\text{out}} = \cos^{-1}\left(-\frac{R\sqrt{2\ell^2 - R^2}}{\ell^2}\right). \tag{C3}
$$

The structure is stable when $\partial E / \partial R > 0$ at $R = \sqrt{2}\ell$. Solving the inequality yields

$$
\sqrt{2}(10 + 3\pi - 8\tilde{A}_0) + 4\tilde{K}(10\sqrt{2} + 3\pi - 4\tilde{P}_0) > 0. \tag{C4}
$$

This critical line lies inside the gas phase.

APPENDIX D: FULL EXPRESSIONS FOR ANALYZING THE STABILITY OF A HONEYCOMB LATTICE

The full expression of the matrix Ω appearing in Eq. (20) is

$$
\begin{aligned}
\Omega_{x,x} &= \frac{\tilde{P}^2 \cos^2(\tilde{q}_x)\cos^2(\tilde{q}_y)}{6} - 4\tilde{K}[3 - \cos(2\tilde{q}_x)\cos^2(\tilde{q}_y) - 7\sin^2(\tilde{q}_y) - 2\cos(\tilde{q}_x)\cos(\tilde{q}_y)] \\
&\quad - \frac{8\tilde{K}\tilde{P}_0}{\tilde{P}}[1 + 4\sin^2(\tilde{q}_y) - \cos(\tilde{q}_x)\cos(\tilde{q}_y)], \\
\Omega_{y,y} &= \frac{\tilde{P}^2 \sin^2(\tilde{q}_y)[\cos(\tilde{q}_x) + 2\cos(\tilde{q}_y)]^2}{18} + \frac{4\tilde{K}}{3}\{20 - 2\cos^2(\tilde{q}_x)\cos^2(\tilde{q}_y) - \cos(4\tilde{q}_y) \\
&\quad - 2\cos(\tilde{q}_x)[8\cos(\tilde{q}_y) + \cos(3\tilde{q}_y)]\} - \frac{24\tilde{K}\tilde{P}_0}{\tilde{P}}[1 - \cos(\tilde{q}_x)\cos(\tilde{q}_y)], \\
\Omega_{x,y} = \Omega_{y,x} &= \frac{\tilde{P}^2}{3^{1/4}24}\sin(\tilde{q}_x)\sin(2\tilde{q}_y)[\cos(\tilde{q}_x) + 2\cos(\tilde{q}_y)] \\
&\quad - \frac{8\tilde{K}}{\sqrt{3}}\sin(\tilde{q}_x)\sin(\tilde{q}_y)[2 - \cos(2\tilde{q}_y) - \cos(\tilde{q}_x)\cos(\tilde{q}_y)] + \frac{4\sqrt{3}\tilde{K}\tilde{P}_0}{\tilde{P}}\sin(\tilde{q}_x)\sin(\tilde{q}_y),
\end{aligned} \tag{D1}
$$

where we used for brevity

$$
\tilde{q}_x = \frac{q_x P}{4}, \quad \tilde{q}_y = \frac{q_y P}{4\sqrt{3}}, \tag{D2}
$$

with P being the circumference of each cell at equilibrium.

The full expression of the functions $g_i(\vec{q})$ appearing in Eq. (23) are

$$
\begin{aligned}
g_1(\vec{q}) &= 42 + 4\cos(2\tilde{q}_y) - 14\cos(4\tilde{q}_y) + 4\cos(6\tilde{q}_y) + 2\cos(3\tilde{q}_x)[5\cos(\tilde{q}_y) + 4\cos(3\tilde{q}_y)] \\
&\quad + 2\cos(2\tilde{q}_x)[-9 - 14\cos(2\tilde{q}_y) + 5\cos(4\tilde{q}_y)] + 2\cos(\tilde{q}_x)[4\cos(\tilde{q}_y) - 18\cos(3\tilde{q}_y) + 5\cos(5\tilde{q}_y)], \\
g_2(\vec{q}) &= \tfrac{1}{243}\{129 - 52\cos(2\tilde{q}_y) - 7\cos(4\tilde{q}_y) + 2\cos(2\tilde{q}_y) + \cos(3\tilde{q}_x)[5\cos(\tilde{q}_y) + 4\cos(3\tilde{q}_y)] \\
&\quad + 4\cos(2\tilde{q}_x)[2 - 5\cos(2\tilde{q}_y)] - \cos(\tilde{q}_x)[104\cos(\tilde{q}_y) - 18\cos(3\tilde{q}_y) - 5\cos(5\tilde{q}_y)]\}, \\
g_3(\vec{q}) &= \tfrac{2}{27}\{\cos(2\tilde{q}_x) + [3 - 4\cos(\tilde{q}_x)\cos(\tilde{q}_y)][2 - \cos(2\tilde{q}_y)]\}.
\end{aligned} \tag{D3}
$$

APPENDIX E: DIFFERENT CELL-CELL AND CELL-BOUNDARY INTERFACE ENERGY

In this section, we show how letting the cell-cell and cell-boundary interface energies differ affects the phase diagram. Specifically, we start from Eq. (1) and set $A_{0,i} = A_0$ for all cells, $\Gamma_i = K_P$ for all cells, $\Lambda_{i,j} = -K_P P_0$ for cell-cell interfaces, and $\Lambda_{i,j} = -K_P P_0 + \frac{\Delta}{2}$ for cell-boundary interfaces, such that up to a constant the energy may be written as

$$E = K_A \sum_i (A_i - A_0)^2 + K_P \sum_i (P_i - P_0)^2 - \Delta \sum_i P_i^{\text{out}},$$

(E1)

where P_i is the total circumference of the cells and P_i^{out} is the part which is not in contact with any other cell.

Because the only difference a nonzero value of Δ can have is on the cells at the boundaries, it does not affect the transition between the confluent and minimal phases and between the minimal and nonconfluent phases. The only effect it has is on the other transitions, which are determined by whether the two- and three-cell interactions are attractive or repulsive.

Now the two-cell interaction is attractive if

$$2(\pi - \tilde{A}_0) + \tilde{K}(2\pi - \tilde{P}_0) - 2\pi\tilde{\Delta} > 0,$$

(E2)

and the three-cell interaction is attractive if

$$\left(\frac{3\sqrt{3} + 4\pi}{6} - \tilde{A}_0 \right) + \frac{2}{3}(3 - \sqrt{3})\tilde{K}\left(\frac{2(3 + 2\pi)}{3} - \tilde{P}_0 \right)$$
$$- \frac{8\pi}{9}(3 - \sqrt{3})\tilde{\Delta} > 0,$$

(E3)

where

$$\tilde{\Delta} = \frac{\Delta}{K_A \ell^4}.$$

(E4)

Hence, in the $\tilde{P}_0 - \tilde{A}_0$ plane, the transitions are still two nonparallel straight lines for any fixed value of $\tilde{\Delta}$. The only qualitative difference would be if the two lines intersect at a nonphysical range of parameters, such that one of the phases will not be accessible. The only physical restriction on the parameters is $\tilde{K} > 0$ and $\tilde{A}_0 > 0$, while $\tilde{P}_0$ and $\tilde{\Delta}$ can have any real value. The value of $\tilde{A}_0$ at the intersection between the two lines is

$$\tilde{A}_{\text{int}} = \frac{1}{66}[-9(4 + 3\sqrt{3}) + 4\pi(21 + 2\sqrt{3})$$
$$+ 8(\pi - 3)(5 + \sqrt{3})\tilde{K} - 8\pi(5 + \sqrt{3})\tilde{\Delta}].$$

(E5)

This is negative for $\tilde{\Delta} > \tilde{\Delta}_c(\tilde{K})$ where

$$\tilde{\Delta}_c(\tilde{K}) = \frac{4\pi(9 - \sqrt{3}) - 9(1 + \sqrt{3})}{16\pi} + \left(1 - \frac{3}{\pi}\right)\tilde{K}$$
$$\approx 1.3 + 0.045\tilde{K}.$$

(E6)

Therefore, the hexagonal phase is inaccessible for large enough values of $\tilde{\Delta}$ as long as $\tilde{K}$ is small enough.

APPENDIX F: EXPANSION OF THE ENERGY TO SECOND ORDER IN THE TEMPERATURE

In this section, we investigate the effects of low temperatures on the energy. Assuming the system has achieved thermal equilibrium, the probability to find it in a certain configuration C is given by the Boltzmann distribution,

$$\mathcal{P}(C) = \frac{e^{-E(C)/T}}{\int e^{-E(C)/T} dC},$$

(F1)

where the configuration C encodes the location of all the cells. Assuming that at $T = 0$ the system reaches its lowest possible energy, then for small T we may expand the energy around that minimum,

$$E \approx E_0 + \frac{1}{2} \sum_{i,j=1}^{2N} x_i x_j U_{i,j} \equiv E_0 + \frac{1}{2} \vec{x}^T U \vec{x},$$

(F2)

where x_i are the deviations of the $2N$ degrees of freedom of the N cells from their locations in the minimum energy configuration. The mean energy per cell is thus

$$\frac{\langle E \rangle}{N} = \frac{1}{N} \frac{\int \left(E_0 + \frac{1}{2}\vec{x}^T U \vec{x} \right) e^{-(E_0 + \frac{1}{2}\vec{x}^T U \vec{x})/T} d\vec{x}}{\int e^{-(E_0 + \frac{1}{2}\vec{x}^T U \vec{x})/T} d\vec{x}}$$
$$= \frac{1}{N} \left(E_0 + \frac{\int \frac{1}{2}\vec{x}^T U \vec{x} e^{-\frac{1}{2}\vec{x}^T U \vec{x}/T} d\vec{x}}{\int e^{-\frac{1}{2}\vec{x}^T U \vec{x}/T} d\vec{x}} \right).$$

(F3)

Using a well-known trick, we can simplify the nominator

$$\frac{\langle E \rangle}{N} = \frac{1}{N} \left(E_0 + \frac{-T\frac{\partial}{\partial \alpha} \int e^{-\frac{1}{2}\alpha \vec{x}^T U \vec{x}/T} d\vec{x}}{\int e^{-\frac{1}{2}\vec{x}^T U \vec{x}/T} d\vec{x}} \right)\Bigg|_{\alpha=1}.$$

(F4)

Evaluating the integrals yields

$$\frac{\langle E \rangle}{N} = \frac{1}{N} \left(E_0 + \frac{-T\frac{\partial}{\partial \alpha} \sqrt{\frac{(2\pi)^{2n}}{|\alpha U'/T|}}}{\sqrt{\frac{(2\pi)^{2n}}{|U'/T|}}} \right)\Bigg|_{\alpha=1},$$

(F5)

where U' is the matrix U reduced to its nonzero eigenvalues, $2n$ is the size of U', and $|U'|$ is the determinant of U'. Simplifying yields

$$\frac{\langle E \rangle}{N} = \frac{1}{N} \left(E_0 - T\frac{\partial}{\partial \alpha} \alpha^{-n} \right)\Bigg|_{\alpha=1}.$$

(F6)

Evaluating the derivative and setting $\alpha = 1$ yields

$$\frac{\langle E \rangle}{N} = \frac{1}{N}(E_0 + nT).$$

(F7)

In order to find the next order correction, we expand the energy to fourth order in $\vec{x}$ and assume that it can be written as

$$E = E_0 + \frac{1}{2} \sum_i \sigma_i y_i^4 - \sum_i \mu_i y_i^2,$$

(F8)

where y_i are linear combinations of x_i that diagonalize E. The mean energy per cell is therefore approximated by

$$\frac{\langle E \rangle}{N} = \frac{E_0}{N} - \frac{T}{N} \frac{\frac{\partial}{\partial \alpha} \int e^{-\frac{1}{2}\alpha(\sum_i \sigma_i y_i^4 - \sum_i \mu_i y_i^2)/T} d\vec{y}}{\int e^{-\frac{1}{2}(\sum_i \sigma_i y_i^4 - \sum_i \mu_i y_i^2)/T} d\vec{y}}\Bigg|_{\alpha=1}.$$

(F9)

Changing the integration variables to $z_i = y_i^2$ yields

$$\frac{\langle E \rangle}{N} = \frac{E_0}{N} - \frac{T}{N} \frac{\frac{\partial}{\partial \alpha} \int e^{-\frac{1}{2}\alpha(\sum_i \sigma_i z_i^2 - \sum_i \mu_i z_i)/T} \frac{d\vec{z}}{\prod_i 2\sqrt{z_i}}}{\int e^{-\frac{1}{2}(\sum_i \sigma_i z_i^2 - \sum_i \mu_i z_i)/T} \frac{d\vec{z}}{\prod_i 2\sqrt{z_i}}} \bigg|_{\alpha=1}.$$

(F10)

Evaluating the integrals yields

$$\frac{\langle E \rangle}{N} = \frac{E_0}{N} - \frac{T}{N} \frac{\frac{\partial}{\partial \alpha} \prod_i e^{\alpha w_i} K_{\frac{1}{4}}(w_i)}{\prod_i e^{w_i} K_{\frac{1}{4}}(w_i)} \bigg|_{\alpha=1},$$

(F11)

where K is the modified Bessel function of the second kind and we defined for brevity

$$w_i = \frac{\mu_i^2}{4 T \sigma_i}.$$

(F12)

Performing the derivative and setting $\alpha = 1$ yields

$$\frac{\langle E \rangle}{N} = \frac{E_0}{N} - \frac{T}{N} \sum_i \left[1 - \frac{K_{3/4}(w_i) + K_{5/4}(w_i)}{2 K_{1/4}(w_i)} \right] w_i.$$

(F13)

For small T (and thus large w_i), we use the asymptotic expansion of the modified Bessel function such that

$$\frac{\langle E \rangle}{N} = \frac{E_0}{N} - \frac{T}{N} \sum_i^{2n} \left(-\frac{1}{2} + \frac{3T}{8} \frac{\sigma_i}{\mu_i^2} \right)$$

$$= \frac{E_0}{N} + \frac{n}{N} T - \frac{3T^2}{8} \left\langle \frac{\sigma}{\mu^2} \right\rangle.$$

(F14)

[1] L. T. Tam and R. A. Weinberg, Nat. Med. **19**, 1438 (2013).

[2] G. Brodland, J. Biomech. Eng. **124**, 188 (2002).

[3] G. Brodland, Appl. Mech. Rev. **57**, 47 (2004).

[4] P. Pathmanathan, J. Cooper, A. Fletcher, G. Mirams, P. Murray, J. Osborne, J. Pitt-Francis, A. Walter, and S. Chapman, Phys. Biol. **6**, 036001 (2009).

[5] F. Graner, J. Theor. Biol. **164**, 455 (1993).

[6] F. Graner and Y. Sawada, J. Theor. Biol. **164**, 477 (1993).

[7] G. Schaller and M. Meyer-Hermann, Phys. Rev. E **71**, 051910 (2005).

[8] D. Bi, X. Yang, M. C. Marchetti, and M. L. Manning, Phys. Rev. X **6**, 021011 (2016).

[9] D. Barton, S. Henkes, C. Weijer, and R. Sknepnek, PLoS Comput. Biol. **13**, e1005569 (2017).

[10] H. Honda and G. Eguchi, J. Theor. Biol. **84**, 575 (1980).

[11] M. Weliky and G. Oster, Development **109**, 373 (1990).

[12] T. Nagai and H. Honda, Philos. Mag. B **81**, 699 (2001).

[13] A. Fletcher, M. Osterfield, R. Baker, and S. Shvartsman, Biophys. J. **106**, 2291 (2014).

[14] A. Fletcher, J. Osborne, P. Maini, and D. Gavaghan, Prog. Biophys. Mol. Biol. **113**, 299 (2013).

[15] D. Bi, J. Lopez, J. Schwarz, and M. Manning, Nat. Phys. **11**, 1074 (2015).

[16] R. Farhadifar, J.-C. Röper, B. Aigouy, S. Eaton, and F. Jülicher, Curr. Biol. **17**, 2095 (2007).

[17] G. Szamel, E. Flenner, and L. Berthier, Phys. Rev. E **91**, 062304 (2015).

[18] S. Steffenoni, K. Kroy, and G. Falasco, Phys. Rev. E **94**, 062139 (2016).

[19] N. B. Ouchi, J. A. Glazier, J. P. Rieu, A. Upadhyaya, and Y. Sawada, Physica A **329**, 451 (2003).

[20] D. Staple, R. Farhadifar, J.-C. Röper, B. Aigouy, S. Eaton, and F. Jülicher, Eur. Phys. J. E **33**, 117 (2010).

[21] M. Spencer, Z. Jabeen, and D. Lubensky, Eur. Phys. J. E **40**, 2 (2017).

[22] See Supplemental Material at http://link.aps.org/supplemental/10.1103/PhysRevE.98.042418 for two animations showing the critical dependance of the cluster phase on the temperature.

[23] A. Boromand, A. Signoreiello, F. Ye, C. O'Hern, and M. Shattuck, arXiv:1801.06150.

[24] L. Zeng, X. Yang, H. Li, Y. Li, C. Yang, W. Gu, Y. Zhou, J. Du, H. Wang, J. Sun, D. Wen, and J. Jiang, Respir. Res. **17**, 164 (2016).

[25] S. Kaliman, C. Jayachandran, F. Refeldht, and A. Smith, Frontiers Physiol. **7**, 551 (2016).

[26] E. Marieb and K. Hoehn, *Human Anatomy and Physiology* (Pearson, Harlow, United Kingdom, 2013).

[27] S. Fortune, Algorithmica **2**, 153 (1987).

[28] M. Bock, A. Tyagi, J.-U. Kreft, and W. Alt, Bull. Math. Biol. **72**, 1696 (2010).