Topology of dividing planar tilings: Mitosis and order in epithelial tissues

M. Kokalj Ladan,[1,*] P. Ziherl,[2,3] and A. Šiber[3,4]

[1]*Faculty of Pharmacy, University of Ljubljana, Aškerčeva 7, SI-1000 Ljubljana, Slovenia*

[2]*Faculty of Mathematics and Physics, University of Ljubljana, Jadranska 19, SI-1000 Ljubljana, Slovenia*

[3]*Jožef Stefan Institute, Jamova 39, SI-1000 Ljubljana, Slovenia*

[4]*Institute of Physics, Bijenička 46, HR-10000 Zagreb, Croatia*

(Received 27 February 2019; revised manuscript received 28 May 2019; published 29 July 2019)

We investigate a range of rule-based models of the in-plane structure of growing single-cell-thick epithelia represented by the distribution of frequencies of polygon classes. Within the Markovian framework introduced by Gibson *et al.* [Nature (London) **442**, 1038 (2006)], we discuss various topologically allowed cell division schemes assumed to control the structure of the tissue as well as a phenomenological Gaussian scheme, and we compute the stationary distributions for all of them. Some of the distributions reproduce those seen in tissues characterized by unbiased mitotic events but also in certain tissues with a preferred orientation of the mitotic plane or a cell-rearrangement process such as neighbor exchange. In addition, we propose the asynchronous-division variant of the model, which builds on the Lewis law and on the Aboav-Weaire law as well as on the fact that the dividing cells are larger than the resting cells. This generalization *a posteriori* validates the original model.

DOI: 10.1103/PhysRevE.100.012410

I. INTRODUCTION

Epithelia, thin sheets of tissues made of one or a few layers of cells, perform important and versatile functions during development as well as in fully developed organisms. They cover other tissues and organs, protecting them from physical stimuli and forming a barrier which facilitates a selective transport of substances. The physical properties of epithelia depend on the interactions between the densely packed cells in them. The strong adhesion between neighboring cells ensures that the cell-cell contacts are mostly flat so that the structure of a simple, singe-cell-thick epithelium can be accurately represented by its en face view, which can be regarded as a tiling of plane by polygons. Most vertices in these tilings are three way; four-way vertices are rare [1] and vertices of valence of more than four occur only in transient structures such as rosettes [2].

The most accessible observables of cells in a tissue are their morphometric features. These quantities are dynamical and change as the tissue develops, and they generally vary across the tissue. In single-cell-thick epithelia, the most easily measured morphometric quantities are the projected (i.e., apical) cell area and the number of neighbors of cells, and their distributions in a given tissue are of particular interest. One of the first statistical descriptions of the in-plane structure of epithelia was reported by Lewis in 1926 [3]. He observed that the average projected cell area is a linear function of the number of cell sides, i.e., the number of its neighbors. This finding correlates a geometrical feature of cells with the local topology of the tiling [4,5]. Also well established are the topological correlations between the number of sides of a given cell and the average number of sides of its nearest neighbors. This effect is described by the Aboav-Weaire law [6,7], which states that the average number of sides of neighbors of a cell with i neighbors is smaller than 6 if i is larger than 6 and vice versa. In other words, cells with few sides are surrounded by cells with many sides whereas many-sided cells are surrounded by few-sided cells.

The topological structure of the tissue is conveniently represented by the distribution of the frequencies of the polygon classes (triangles, quadrilaterals, pentagons, etc.) seen in the en face view. In typical epithelia the most abundant polygon class are hexagons followed by pentagons and heptagons [8,9]; quadrilaterals are usually rare and so are octagons, nonagons, and other polygons with many sides. If all vertices are three way, the frequencies of polygon classes p_i satisfy the sum rule following from the Euler formula

$$\sum_i (i-6)p_i = 0, \tag{1}$$

where the sum is over all polygon classes; this rule implies that the average number of polygon sides is 6. The distribution of p_i's is usually skewed since it is bounded at the lower end (because no polygon can have less than three sides) but not at the upper end.

Many approaches were explored to interpret the different epithelial topologies, and each of them involves a certain approximation. Some of them study the role of genes and macromolecules in cell membranes and cell-cell interactions, which play an important role in tissue organization [9]. Others are more physical and attempt to describe the observed patterns in terms of mechanics and thermodynamics, building on concepts such as the energy or the entropy of the tissue viewed as a tiling [10]. A comprehensive model which considers energy-based cell mechanics and includes cell division as a

*meta.kokalj.ladan@ffa.uni-lj.si

generator of local topological transformations was elaborated in Ref. [11]. Depending on parameters, the nondividing variant of such a model tissue adopts either a disordered state with a distribution of polygon classes or an ordered state containing only hexagons [11]. Once the cells are allowed to divide, the tissue is disordered for all combinations of parameters, suggesting that its structure should generally be viewed as a stationary rather than an equilibrium feature. [Here we switch to the two-dimensional (2D) terminology pertaining to the tiling representation of the tissue, which is used in the rest of the paper unless stated otherwise.]

Yet another view of an epithelium, which focuses exclusively on cell division, was proposed in Ref. [12]. In this mean-field model, where cells divide from generation to generation and where all spatial and mechanical aspects of the tissue are disregarded, the state of the tissue is not represented by an actual tiling but solely by the vector

$$\mathbf{p} = [p_3, p_4, p_5, p_6, \ldots] \tag{2}$$

giving the frequencies of polygon classes p_i. The essence of this model is a concise set of rules which specify how an average cell divides depending on the number of its sides. The rules can be formulated within a discrete Markov-chain model yielding a stationary state with a distribution of polygon classes which compares favorably to some experimentally observed distributions [12].

The predictions of this approach are quite interesting as it sees the tissue in a very different light than, e.g., the model of Ref. [11]. Of course, one may worry that the representation of the tissue based on $\mathbf{p}$ that is not backed up by an actual tiling may be too stripped down [13]. This is a valid concern, but similar although lesser reservations apply to the commonly used 2D vertex-model representation of epithelia [11,14,15] which pertain to the projected (apical) geometry of the tissue and disregard the actual three-dimensional (3D) shape of cells. Given that in most epithelia the apical side is easily observed, this is a natural choice yet still a simplification even if the cells are prismatic, and much more so if they are not [16]. At the same time, in proliferating epithelia one does not need to worry about this very much because, as mentioned above, if cells divide then tissue structure changes dramatically [11] so that the mechanical aspects of cells such as their energy cannot play a decisive role. More importantly, any no-gap and no-overlap tiling is subject to strong constraints embodied in the sum rule [Eq. (1)] which too restrict the impact of cell geometry on the structure of the tissue. These considerations suggest that the Markov-chain model of Ref. [12] is well worth analyzing in more detail.

Before spelling out the objectives of this paper, we summarize the Markov-chain model [12] where the vector of polygon frequencies at a time t is given by

$$\mathbf{p}_t = \mathbf{p}_{t-1}\mathsf{PS}; \tag{3}$$

here time is measured in units of cell cycle and thus gives the generation number. The P matrix embodies the division rules, the entry P_{ij} giving the probability that an i-sided mother cell divides and produces a j-sided daughter cell. If no triangles are allowed, each daughter cell must inherit at least two vertices from the mother cell such that upon formation of the new side that divides the mother into two, each daughter will have at least four vertices. This leaves the remaining $i-4$ vertices to be distributed between the daughters. In Ref. [12], these $i-4$ vertices are divided randomly between the daughters. In addition to the vertices inherited from the mother, each daughter also acquires two additional vertices due to the newly created side separating the daughters. Thus, the probability for a j-sided cell to emerge from an i-sided mother cell is

$$P_{ij} = \frac{C_{\mathrm{comb}}(i-4, j-4)}{2^{i-4}}, \tag{4}$$

where $C_{\mathrm{comb}}(a, b)$ is the combinatorial coefficient.

The S matrix referred to as the shift matrix represents the effect of division of the neighbors of a reference cell, which too increases the number of sides of the cell if the joint side is split into two, S_{ij} being the probability that an i-sided reference cell becomes j sided due to the division of a neighbor. In an epithelium containing N cells, $2N$ new sides are added because of this effect in each round of division and if they are distributed evenly among the $2N$ cells present after division, the number of sides in a cell increases by 1 on average. Thus, $S_{ij} = 1$ if $j = i + 1$ and 0 otherwise. This is a mean-field approximation; if the tissue were represented as an actual tiling, some cells would gain no sides and others would gain more than 1 side [12].

One of the key assumptions of this model is that apart from the two vertices that each daughter must receive from the mother cell, the mother's vertices are distributed like in a binomial distribution. This assumption disregards the fact that the vertices are geometrical rather than combinatorial entities and are not independent because those that belong to a given daughter must be adjacent to each other. As such, vertices cannot be divided in the same way as the balls can be thrown in two bins. The above topological limitation should be taken into account, and here we propose a variant of the cell division scheme from Ref. [12] that includes it as well as a range of generalizations including the size-dependent asynchronous-division version of the model which all complement earlier reconsiderations of the Markov-chain approach [13,17].

The disposition of the paper is as follows: In Sec. II we introduce and analyze five types of rule-based division schemes and the phenomenological Gaussian scheme, in Sec. III we propose the asynchronous-division model where the probability that a cell divides depends on the number of its sides consistent with experimental observations; this model relies on the Lewis law and the Aboav-Weaire law. In Sec. IV we compare the results to experimental data and Sec. V concludes the paper.

II. RULE-BASED SYNCHRONOUS-DIVISION SCHEMES

The modification of the model of Ref. [12] that we propose implements the limitations to the combinatorics of cell division that are required by geometry and topology. As suggested in Ref. [13], there may exist various kinds of cell division rules and we explore five such schemes, all of them departing from that in Ref. [12]. Each of these rules is encoded in a different P matrix and thus leads to a different stationary distribution of polygon classes.

The dominant topology of cell division is that where the mitotic plane cuts from one of the sides toward another, splitting each of the two sides into two. This leads to three-way vertices in the tiling. The other scenario, where the mitotic plane begins or ends in a vertex of the mother cell such that one of the sides does not split and a four-way vertex is created instead, is very rare; even more rare is the case where the mitotic plane connects two vertices of the mother. Like Ref. [12], we consider the dominant scenario. The other condition imposed in Ref. [12] is that the frequency of triangles is zero, which is based on the observation that such cells are rarely found in experiments. Here, we analyze cell division schemes which allow triangles as well as those where not only triangles but also quadrilaterals and pentagons are absent, and we compare the effects of such restrictions.

Like in Ref. [12], we concentrate on the distribution of vertices among the daughter cells and we formulate the cell division rules by focusing on the vertices as the geometrical entities defining the mother cell and two daughter cells. The division rules may equivalently be formulated by choosing the two edges which are to be split into two in order to insert the dividing edge, thereby generating the daughter cells.

A. Random division (R_m)

In the random division scheme, each of the daughters inherits at least one, two, and three vertices from the mother if the smallest-index polygon class allowed are triangles, quadrilaterals, and pentagons, respectively; as noted above, additional two vertices are created at points where the division line intersects the sides of the mother cell. Figure 1 shows the step-by-step branching algorithm used to determine the two disjoint sets of vertices, one belonging to the first daughter [red (dark gray) circles in Fig. 1] and the other to the second daughter [blue (light gray) circles in Fig. 1] separated by the mitotic planes (dashed lines in Fig. 1); the example pertains to the division of a six-sided cell into daughters which can be triangles, quadrilaterals, pentagons, hexagons, and heptagons. In this algorithm, an arbitrary vertex of the mother is assigned to the first daughter. Then, the adjacent vertex in the clockwise direction is assigned either to the first or to the second daughter cell with equal probabilities of $\frac{1}{2}$. If this vertex is assigned to the first daughter, the procedure of vertex assignment continues in the same manner in the clockwise direction, whereas otherwise the procedure samples the vertices in counterclockwise sense starting from the first assigned vertex. This procedure is terminated as soon as the vertices belonging to one daughter are enclosed by those belonging to the other, or when all vertices are assigned to the two daughters. In Fig. 1, the final partitions are indicated by a bold outline.

We note that the branching algorithm illustrated in Fig. 1 is an auxiliary device which allows us to identify the possible final states rather rather than a physical model of cell division, where the vertices belonging to each daughter are not selected in a sequential manner. Thus, the probability associated with a given final partition is given by the number of times it occurs in the scheme divided by the total number of the final partitions and is not weighted by the probability of arriving at each final partition in the auxiliary scheme.

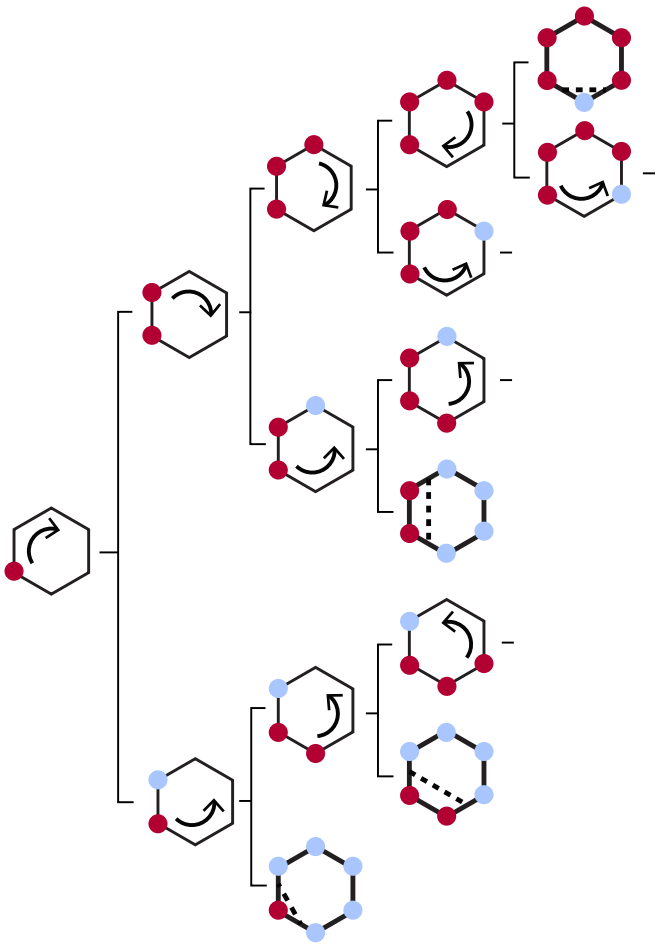

FIG. 1. Schematic of the random division scheme applied to a hexagon, with triangles as the smallest polygon class. The algorithm starts with a random vertex assigned to the first daughter (left), and in each step proceeding from left to right one of the neighbors of the string of vertices already belonging to the first daughter is assigned to either daughter with equal probabilities. Red (dark gray) and blue (light gray) circles indicate vertices assigned to the first and the second daughters whereas arrows show the direction of the neighbor assignment in each step; dashed lines show the mitotic plane. This procedure is carried out until all vertices are divided between the two daughters, and the final partitions in this scheme are highlighted by bold contours.

By construction, this algorithm respects the topological necessity that all the vertices belonging to either daughter are contiguous and form a string. Once all possible outcomes of the algorithm are identified, we need to count the number of instances when the result of the division of an i-sided mother is a j-sided daughter, and then we calculate the respective frequencies which represent the entries P_{ij} of the P matrix. The P matrix of the R_3 scheme is spelled out in Appendix A together with the matrices of all other schemes for the case where the smallest polygon class are triangles. (We briefly refer to the lower-end cutoff of the polygon classes allowed, which is indicated in the subscript in the abbreviated name of a scheme, as the smallest polygon class.)

The version of this scheme presented in Fig. 1 and denoted by R_3 produces daughters with three or more vertices, but it can be generalized so as to exclude triangular cells by

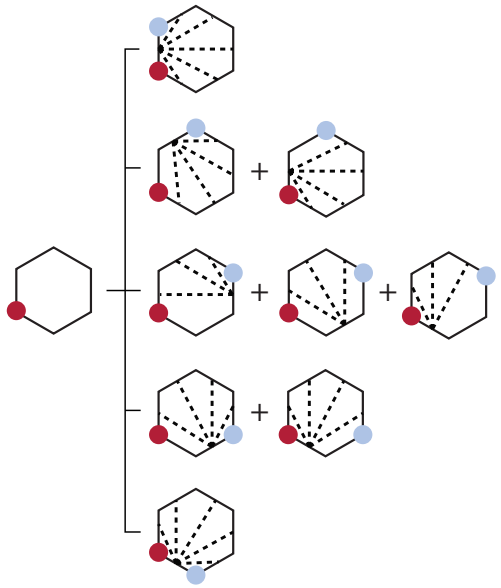

FIG. 2. Partly random division scheme, here illustrated by the patterns resulting from the PR_3 division of a hexagonal mother, begins with a vertex assigned to the first daughter [red (dark gray) circle] and a vertex randomly selected from the remaining vertices [blue (light gray) circle]. Based on these two vertices, we construct all topologically allowed division patterns (dashed lines).

assigning two rather than just one vertex to the first daughter and by terminating the algorithm when two vertices are left for the second daughter; this may be thought to represent the detachment or death of three-sided cells. This variant, which allows polygons with four or more sides, is referred to as R_4. In an analogous fashion we can also adapt the algorithm such that the smallest number of sides is 5 (R_5).

B. Partly random division (PR_m)

In the second scheme considered here, an arbitrary vertex of an i-sided mother cell is assigned to the first daughter. Then, one of the remaining $i-1$ vertices is selected at random, i.e., with a probability of $1/(i-1)$, and assigned to the second daughter. In the next step, the remaining $i-2$ vertices are distributed among the daughters, allowing all divisions that result in two contiguous strings of vertices (Fig. 2). All of the thus obtained patterns are assumed to be equiprobable. Like the random scheme, this algorithm exists in three variants depending on the smallest polygon class (3, 4, or 5; referred to as PR_3, PR_4, and PR_5, respectively); the PR_4 and PR_5 schemes start with 2 and 3 contiguous vertices for each daughter, respectively, rather than with 1 like PR_3.

C. Symmetric division (S_m)

Experiments show that often each of the two most distant vertices of the mother belongs to a different daughter and that the more symmetrical divisions are more common than the less symmetric ones [4,18]. To account for these facts, we construct a scheme where each daughter is assigned 1 of the 2 vertices that are topologically farthest from each other, and the dividing line is determined by the most symmetric division. Thus, a mother cell with an even number of sides i divides into daughters with $i/2+2$ sides, whereas those with an odd number of sides produce daughters with $i/2+3/2$ and $i/2+5/2$ sides (Fig. 3).

FIG. 3. Symmetric division scheme of cells with three, four, five, and six sides.

D. Near-symmetric division (S^2_m and S^3_m) and near-symmetric topological division (ST^2_m and ST^3_m)

The symmetric-division scheme can be extended so as to also include the second and the third most symmetric patterns; these two generalizations are briefly referred to as S^2_m and S^3_m, respectively. If we allow the first and the second most symmetric divisions, then a mother cell with an even number of sides i can produce daughters with $i/2+1$, $i/2+2$, or $i/2+3$ sides (except if the mother is a triangle, in which case the daughters can be only triangular and quadrilateral) and that with an odd i can produce daughters with $i/2+1/2$, $i/2+3/2$, $i/2+5/2$, or $i/2+7/2$ sides. Figure 4(a) shows the two possible patterns in a hexagonal mother cell, which can divide into two pentagonal cells or into a hexagonal and a quadrilateral cell. These modes are then subject to further restrictions imposed by the smallest polygon class permitted (triangles, quadrilaterals, or pentagons) and are assumed to occur with equal probabilities. In a similar fashion, the third most symmetric division can be included too. In the near-symmetric topological division scheme (ST^2_m and ST^3_m), the probability for a given division pattern is proportional to the number of times the pattern appears [Fig. 4(b)].

E. Topological division (T_m)

In this scheme, two topologically most distant vertices are assigned to the daughters and all possible divisions are allowed, the frequency of a given pattern being proportional to the number of its occurrences. As a result, the more symmetric divisions are more frequent than the less symmetric ones (Fig. 5). This approach is the same as that in Ref. [12], except that it respects the topological limitations; as all of the other schemes, it appears in three variants depending on the smallest polygon class.

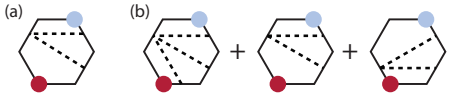

FIG. 4. Division of a hexagonal mother cell in the near-symmetric division scheme allowing the first and the second most symmetric patterns (a) which are assumed to occur with equal probabilities. In the near-symmetric topological scheme ST^2_m, the probabilities of these modes are not the same but proportional to the number of times each pattern appears. A hexagonal mother produces a quadrilateral, a pentagonal, and a hexagonal daughter with a probability of $\frac{2}{7}$, $\frac{3}{7}$, and $\frac{2}{7}$, respectively (b).

FIG. 5. Division patterns of a hexagonal mother cell allowed by the topological scheme T_3. Two most distant vertices are assigned to the daughters, and each possible division indicated by the position of the mitotic plane (dashed line) occurs with a frequency proportional to the number of instances where it appears.

F. Gaussian scheme

The P matrices of the above schemes share several features best appreciated when looking at them; those corresponding to the case where the smallest polygon class are triangles are listed in Appendix A. The number of nonzero entries in a row depends on the row in question (i.e., on the number of sides of the mother cell) and on the smallest polygon class. If the smallest polygon class are triangles, the polygons arising from the division of an i-sided cell range from triangles to $(i+1)$-sided polygons. If the smallest polygon class are quadrilaterals, four- to i-sided cells can result from division, and if this class are pentagons daughter cells can have between five and $i-1$ sides. By comparing the matrix in Eq. (A7) to that spelled out in the non-normalized form in Box 1 in Ref. [12], we also observe that in case where the smallest polygon class are quadrilaterals, the P matrix is lower triangular whereas if the smallest class are triangles it contains an additional upper bidiagonal. The rows of the P matrix are symmetric in that the first nonzero value in a row is the same as the last one, etc.; naturally, each row must be normalized (see the P matrices in Appendix A).

In the phenomenological Gaussian scheme, we dispose of the rules behind the five schemes and we propose P matrices with entries in each row distributed according to the Gaussian function. The number of nonzero entries in each row is limited depending on the row number and on the smallest polygon class. The entries of the Gaussian P matrix are given by

$$P_{ij} = \frac{\exp\left(-x_{ij}^2/2\sigma^2\right)}{\sum_j \exp\left(-x_{ij}^2/2\sigma^2\right)}, \tag{5}$$

where

$$x_{ij} = -i/2 + j - 2 \tag{6}$$

and $2 < j < i+2$ if the smallest polygon class considered are triangles; if it is quadrilaterals and pentagons, then $3 < j < i+1$ and $4 < j < i$, respectively. The Gaussian function respects the symmetry of entries in the rows of the P matrix, and the position of the maximum of the Gaussian function in Eq. (6) is chosen so as to correspond to all P matrices constructed based on the division rules. The denominator in Eq. (5) ensures that the rows are normalized. The Gaussian scheme is controlled solely by σ, which is taken to be the same for all rows in the P matrix. If σ is small, then the distribution of entries in a row is narrow, whereas if σ is large it is broad.

G. Stationary distributions of polygon classes

The different cell division schemes proposed above produce different stationary distributions of polygon classes (Fig. 6), which can be obtained either iteratively or by finding the dominant eigenvector. Interestingly, one can quickly see that the distributions can be divided into three groups color coded in the figure using shades of gray, with some variations within each group. Those in the first group have the lowest frequency of hexagons (less than about 30%) and are broad, containing polygon classes ranging from quadrilaterals to nonagons or even decagons. These distributions are rather skewed. The second group consists of distributions that are somewhat narrower and have between 35% and 50% of hexagons. Here, skewness is present too but in the two distributions from this group shown in the left column it is rather small. The third group represents completely ordered tissues consisting exclusively of hexagons. Here, the hexagonal cells still divide so as to produce pentagons, e.g., in the symmetric scheme [Eq. (A3)], but the stationary distribution is also determined by the S matrix. The S matrix promotes all pentagons produced by the division of the hexagons to hexagons so that the tissue consisting solely of hexagonal cells is indeed the fixed point of the algorithm.

The left, middle, and right columns in Fig. 6 correspond to division schemes where the smallest polygon class is triangles, quadrilaterals, and pentagons, respectively. It is evident that the three groups are largely although not exclusively determined by the choice of the smallest polygon class: all members of the first group are in the left column and all distributions in the right column belong to the third group. We also observe that the stationary distributions are correlated with the distribution of coefficients in the row of the P matrix representing the division of heptagons, which are shown in the inset of each panel in Fig. 6. In schemes with only a few nonzero coefficients in this row then the stationary distribution is narrow. The most evident examples are the cases in the right column, where the smallest polygon class are pentagons, and the symmetric division schemes S_m, where heptagonal cells divide into pentagonal and hexagonal daughters with equal probabilities; in both cases the stationary state consists solely of hexagons. On the other hand, if there are many nonzero coefficients in the row encoding the division of hexagons (i.e., if there exist many division topologies), the distribution of polygon classes in the tissue will be broad and the frequency of hexagons will be typically considerably smaller than 100%.

In the Gaussian scheme, the form of matrix P is controlled by a continuous parameter σ, and hence there exists a continuous set of stationary distributions of polygon classes which depend both on σ and on the choice of smallest polygon class. Representative examples are shown in Fig. 7. Like in the rule-based schemes, the distribution of polygon classes is broad if σ is large enough and if the smallest polygon class are triangles. On the other hand, sufficiently small σ, say 0.1, result in a highly ordered, hexagons-only tissue as shown in the bottom row of Fig. 7. The Gaussian scheme generally reproduces all types of distributions obtained using rule-based schemes, and it also gives a narrow symmetric distribution containing pentagons, hexagons, and heptagons with $p_5 = p_7$ seen, e.g., at $\sigma = 0.5$ if the smallest polygon class are either triangles or quadrilaterals (Fig. 7). Such distributions do not result from any set of rules described in Sec. II. All Gaussian-scheme distributions satisfy Eq. (1) with an accuracy of at least 10^{-7}.

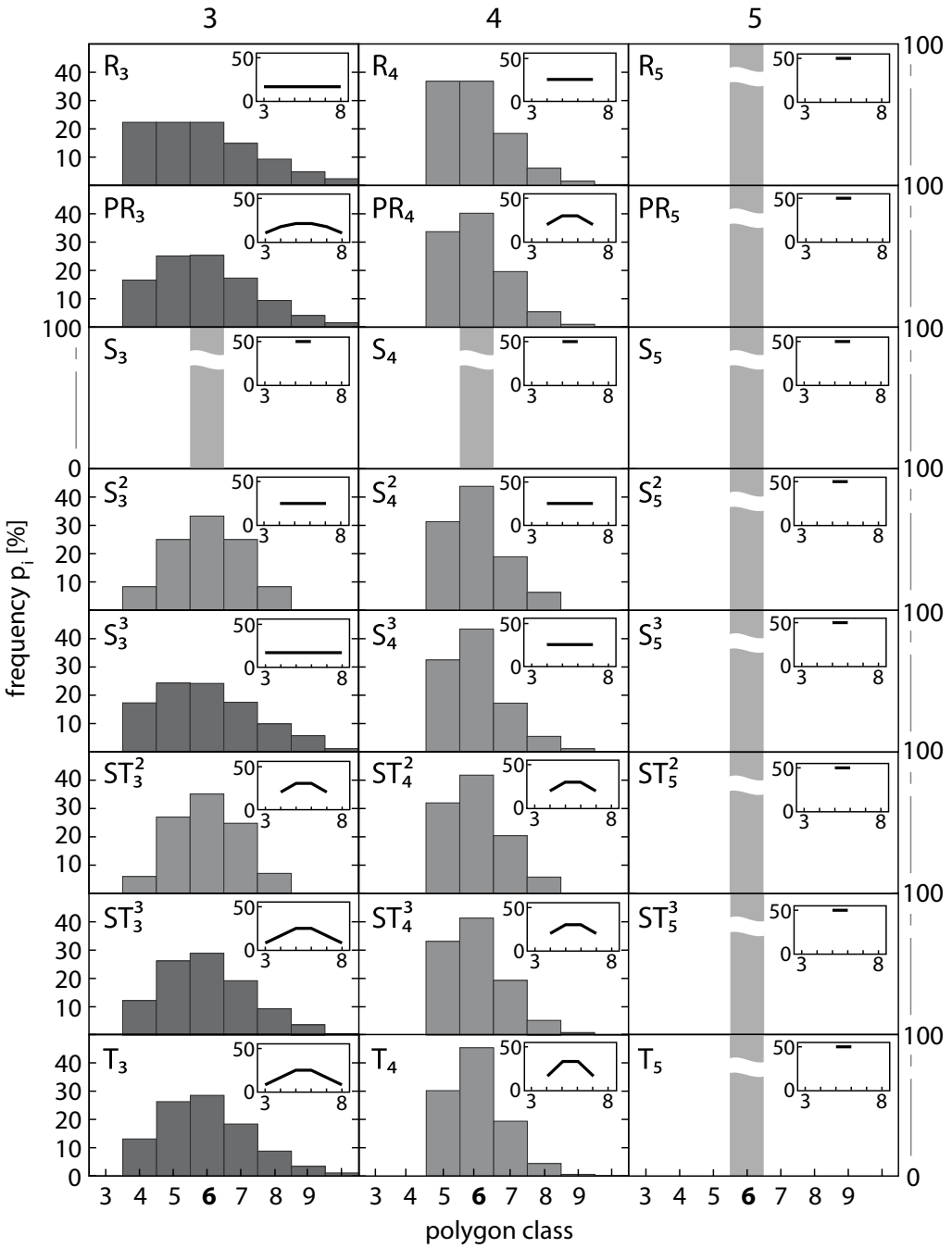

FIG. 6. Stationary distributions of polygon classes obtained using rule-based cell division schemes from Sec. II. In the left, middle, and right columns, the smallest polygon class are triangles, quadrilaterals, and pentagons, respectively. The insets show the coefficients in the row of the P matrix representing the probabilities for the division of heptagons, with lines connecting points for better readability. The three groups of distributions are color coded: in the dark-gray ones the frequency of hexagons is below 30%, in the medium-gray ones it is between 30% and 100%, and in the light-gray ones it is 100%.

Some of our schemes can be compared directly to those obtained in simulations within the topological model of Ref. [18]. In this model, the tissue is represented by a finite tiling but there is no mechanism controlling the geometrical properties of the tiles so that the tiling is really a topological object. Like in Ref. [12] all cells divide in each generation, and the location of the mitotic plane is determined sequentially by various rules. This leads to many modes of cell division, two of which are nominally equivalent to our R_3 and S_3 schemes; a third mode is partly related to the T_3 scheme. With the stationary set of frequencies $\mathbf{p} = [p_3, p_4 \ldots p_{10}] = [22.3\%, 22.3\%, 22.3\%, 14.9\%, 9.3\%, 4.8\%, 2.4\%]$ (here we do not include $p_{i>10}$), the R_3 scheme compares rather well to the "random and unequal split" mode where $\mathbf{p} \approx [26.9\%, 27.2\%, 18.4\%, 10.4\%, 6.3\%, 4.1\%, 2.8\%]$. The S_3 scheme, which is characterized by a symmetric division of polygons and predicts that the only nonzero frequency is $p_6 = 100\%$, corresponds to the "random and equal split" mode

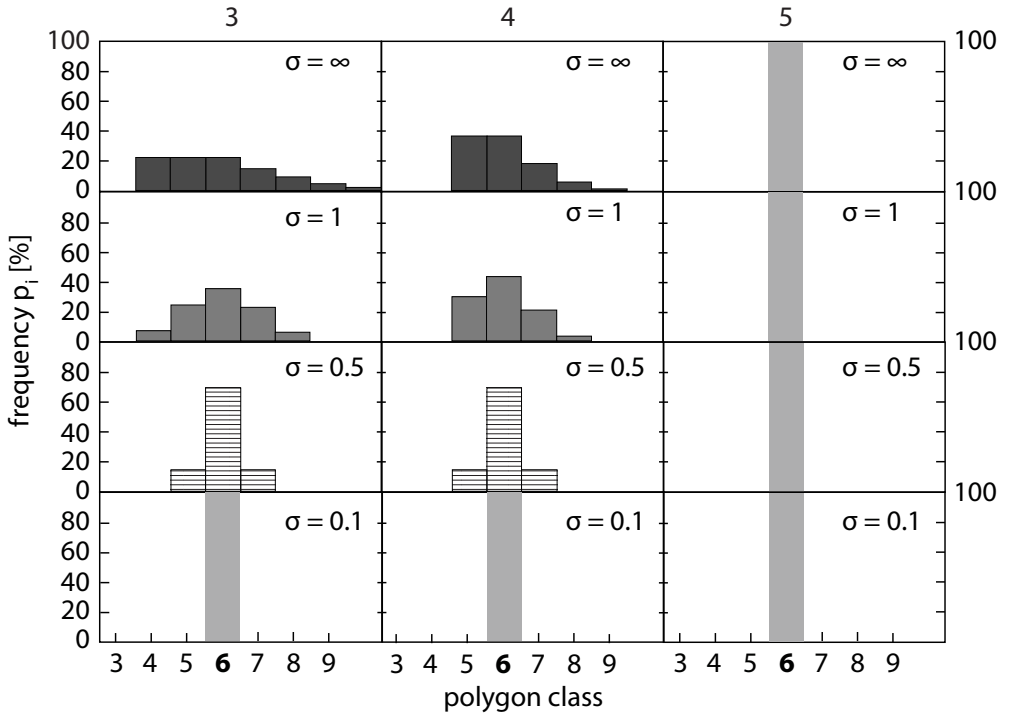

FIG. 7. Representative examples of stationary distributions of polygon classes within the Gaussian scheme for $\sigma = \infty$, 1, 0.5, and 0.1 (top to bottom) and for three choices of the smallest polygon class (triangles, quadrilaterals, and pentagons in left, middle, and right columns, respectively). The color code is the same as in Fig. 6 except in two of the $\sigma = 0.5$ distributions containing only pentagons, hexagons, and heptagons with identical frequencies of pentagons and heptagons (hatched).

with $\mathbf{p} \approx [7.9\%, 29.4\%, 34.2\%, 19.0\%, 6.8\%, 2.1\%, 0.3\%]$, and here the stationary distributions are completely different. On the other hand, our T_3 scheme is not exactly analogous to the "random and binomial split" mode but their distributions of polygon classes, $\mathbf{p} = [13.1\%, 26.3\%, 28.5\%, 18.4\%, 8.8\%, 3.4\%, 1.1\%]$ and $\mathbf{p} \approx [18.7\%, 27.5\%, 22.6\%, 15.2\%, 8.2\%, 4.4\%, 1.7\%]$, respectively, are reasonably close to each other. This comparison is encouraging but not good enough so as to validate the schemes from Sec. II. At the same time, there is no guarantee that the topological representation of Ref. [18] can really be used as a reference. With many spiky cells, its physical form clearly departs from the tilings seen in real tissues, and any shape relaxation would surely affect the topology.

III. SIZE-DEPENDENT ASYNCHRONOUS DIVISION

By accounting for the two-dimensional topology and by extending the cell division rules so as to include certain empirical observations as well as some elements of mechanics and energetics of division, the above modifications bring the model from Ref. [12] closer to real tissues as evidenced by the diverse range of stationary distributions of polygon classes obtained. However, the approach still suffers from inherent drawbacks. First, the evolution of the tissue is divided into time steps corresponding to generations so that in each step, all cells in the tissue undergo exactly one round of division irrespective of any factors that control the process. The most important physical feature that affects the probability for division is cell size, which is known to increase prior to division [4] so that the probability should be larger in the larger cells. Cell size does not directly appear in the model but it does correlate with the number of sides: The Lewis law states that the average projected (i.e., apical) area of cells in an epithelium is a linear function of the number of sides i [4].

Second, the version of the shift matrix S proposed in Ref. [12] is obtained by averaging over the whole epithelium, and all cells in a generation gain an additional vertex due to division of their neighbors. However, the number of sides of neighbors of a given cell is anticorrelated with the number of sides of the cell itself: According to Aboav-Weaire law cells with few sides are, on average, surrounded by cells with many sides and vice versa [6,7]. As a result, a mitotic event is more likely to take place in the neighbors of a cell with a small number of sides i than in the neighbors of a cell with a large i, and thus the reference cell is more likely to gain a side due to the division of its neighbors if it has only a few sides. The S matrix should thus depend on i.

A. Formulation of mean-field asynchronous-division model

The two deficiencies can be remedied within a refined mean-field framework. Instead of considering cell divisions at the level of a single generation where all of the cells divide, we consider the evolution of the tissue in a time interval Δt which is fixed by the fastest dividing cells, i.e., those with the largest number of sides. For this we need to define the largest polygon class considered and thus restrict the dimension of the vector of frequencies of the polygon classes $\mathbf{p}$. Here, we take that i should be no larger than $i_{\max} = 10$ because we know

of no experimental report of a tissue where the frequency of cells with more than 10 sides would be sizable [19]; the smallest polygon class allowed are triangles. The duration of the Markov step Δt is thus given by the average time required for the division of a decagonal cell. During this period, fewer than 100% of nonagonal, octagonal, ..., triangular cells divide so that the probability for division in a time step of Δt denoted by q_i depend on the number of sides i, with $q_{i<i_{\max}} < 1$ and $q_{i_{\max}} = 1$. This is where the Lewis law implicitly resides: cells with many sides are on average larger than those with few sides, so that in them the probability for division within the period Δt should be larger than in the smaller cells.

A convenient ansatz for the dependence of q_i on i is

$$q_i = \frac{1 + \tanh[\alpha(i - \beta)]}{2\sigma_{i_{\max}}}, \tag{7}$$

where $\sigma_{i_{\max}} = \{1 + \tanh[\alpha(i_{\max} - \beta)]\}/2$ ensures that $q_{i_{\max}} = 1$. This ansatz has two parameters. Roughly speaking, α controls how strongly q_i depends on i and for $\alpha = 0$, $q_i = 1$ irrespective of i so that the model, save for the upper cutoff of the basis at $i_{\max}$ polygons, reduces to its synchronous variant. On the other hand, β represents a threshold for i such that $q_{i<\beta} < 0.5$ and $q_{i<\beta} > 0.5$; in the limit of $\alpha \gg 1$, $q_{i<\beta} \to 0$ whereas $q_{i>\beta} \to 1$. Experiments show that the average number of sides of a dividing cell is about seven and that the frequency of eight- and nine-sided cells among the dividing population is much larger than among the resting cells [4,20], and thus choosing β around six seems reasonable. As per α, we are not aware of any experimental facts that can be used to estimate it; evidently, it can be considered as a fitting parameter. Our ansatz qualitatively agrees with the *Drosophila* sensory organ precursor cells data from Ref. [21] analyzed in Ref. [17]; depending on parameters, the agreement can even be quantitative.

Now we can spell out the asynchronous-division matrix $\widetilde{\mathsf{P}}$ corresponding to one Markov step:

$$\widetilde{P}_{ij} = (1 - q_i)\delta_{ij} + q_i P_{ij}; \tag{8}$$

here δ_{ij} is the Kronecker symbol and P_{ij} is a suitable rule-based division matrix from Sec. II. This expression states that within one Markov step, an i-sided cell either does not divide with a probability of $1 - q_i$ or it divides with a probability of q_i and produces j-sided daughters as specified by the P matrix.

The asynchronous-division shift matrix $\widetilde{\mathsf{S}}$, which represents the redistribution of frequencies of polygon classes due to division of neighbors of the reference cell, is based on the empirical Aboav-Weaire law mentioned above. This law states that the average number of sides of neighbors of an i-sided cell is given by

$$m(i) \approx 6 - a + \frac{6a + \mu_2}{i}, \tag{9}$$

where $a \approx 1$ is a constant and μ_2 is the second moment of polygon frequencies p_i which is about 1 in the disordered tissues, smaller than 1 in the more ordered ones, and 0 in tissues consisting solely of hexagonal cells. We note that $m(i)$ given by Eq. (9) is restricted to a rather narrow interval, for $a = 1$ and $\mu_2 = 1$, $m(3) \approx 7.3$ and $m(i \gg 1) \approx 5$, so that using it within too broad a range of i would be questionable but within our scheme where i is between 3 and 10 it makes sense. In a mean-field spirit, we assume that the probability that a neighbor of a cell with i sides divides in one Markov step is given by $q_{m(i)}$. The shift matrix is now given by

$$\widetilde{S}_{ij} = [1 - q_{m(i)}]\delta_{ij} + q_{m(i)}\delta_{i+1,j}. \tag{10}$$

Here, the first term represents the decrease of the frequency of i-sided cells due to the division of their i neighbors, which convert i-sided cells into those with more than i sides, and the second term is the increase of the frequency of $(i + 1)$-sided cells due to the promotion of i-sided cells because one of their $m(i)$ neighbors has divided and thus the i-sided cells gained 1 side. The shift matrix $\widetilde{\mathsf{S}}$ conserves the probability when applied to a vector of frequencies of polygon classes unrestricted from above; a more detailed discussion of its form is presented in Appendix B.

The size-dependent asynchronous-division model based on $\widetilde{\mathsf{P}}$ and $\widetilde{\mathsf{S}}$ defined by Eqs. (8) and (10), respectively, reduces to the variant proposed in Ref. [12] if $\alpha = 0$ and we solve it in the same manner except that now in each time step we need to find the value of μ_2 self-consistently as μ_2 is both a parameter of the shift matrix $\widetilde{\mathsf{S}}$ through $m(i)$ [Eq. (9)] and the second moment of the ensuing distribution of frequencies $\mathbf{p}$. This is done iteratively, with three to four iterations needed to converge the procedure. On the technical side, we note that as the basis is cut at $i_{\max}$, the shift matrix $\widetilde{\mathsf{S}}$ in general results in a leakage from the $\mathbf{p}$ vector at its $i_{\max}$th (i.e., 10th) component. This is a small effect because p_{10} is generally rather small. The probability leakage can be repaired either by the renormalization of the last column of the $\widetilde{\mathsf{S}}$ matrix or by the renormalization of the $\mathbf{p}$ vector after each iteration. In any case, the exact nature of dealing with this problem is unimportant for the salient features of the stationary distribution of polygon classes.

Figure 8(b) shows the stationary distributions of the asynchronous-division model based on the near-symmetric topological rule ST_3^2 (Sec. II D) for several values of α at $\beta = 5.5$; Fig. 8(a) contains the corresponding division probabilities q_i. The stationary distributions demonstrate that the effect of the variation of q_i across the polygon classes can be rather strong depending on α, and that the ensuing distributions are generally broader and shifted to smaller polygon classes compared to that obtained using the synchronous-division model. This is best appreciated by comparing the $\alpha = 0.25$ case [fourth panel from left in Fig. 8(b)] to the $\alpha = 0$ case, which reduces to the synchronous-division model [22] [first panel from left in Fig. 8(b)]. In the former, the frequency of pentagons, quadrilaterals, and triangles are considerably larger than in the latter whereas the frequencies of hexagons, heptagons, and octagons are much smaller; in fact, pentagons are the most abundant polygon class. The shift is clearly visible in Fig. 8(c) where we plot the mean of the stationary distributions μ_1, which drops to a value as low as 5.4 at $\alpha \approx 0.25$. In turn, for $\alpha \gg 1$ where the difference in division probabilities of the smaller and the larger polygon classes is essentially 100%, the mean again approaches that at $\alpha = 0$, that is $\mu_1 = 6$. Figure 8(c) also shows the mean for other,

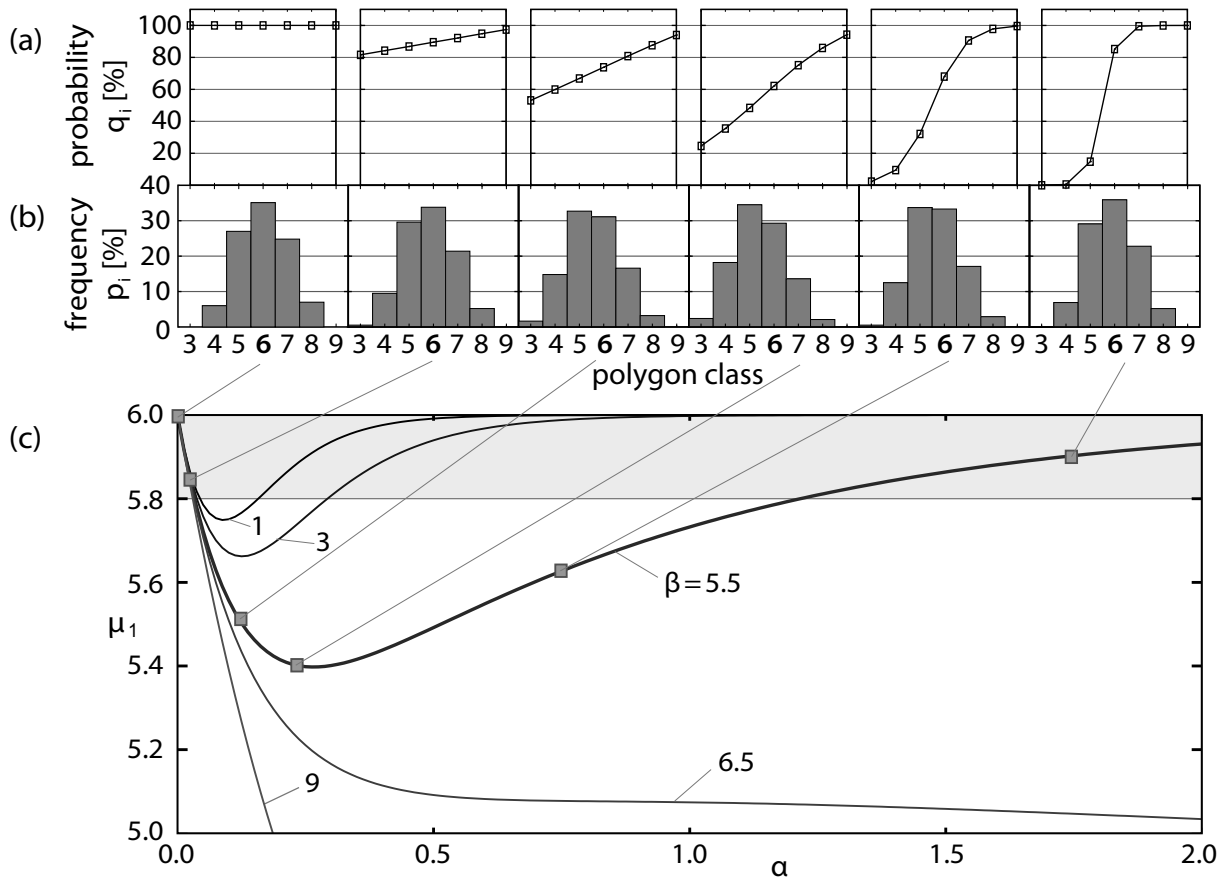

FIG. 8. Stationary distributions of frequencies of polygon classes within the size-dependent asynchronous-division model for $\beta = 5.5$ and $\alpha = 0, 0.03, 0.1, 0.25, 0.75$, and 1.75 (b) together with the division probabilities for the different classes (a). (c) Shows the mean of the distributions vs α for $\beta = 1, 3, 5.5, 6.5$, and 9; squares indicate the data points corresponding to items in (b). The gray band represents the range of μ_1 seen in experiments [1].

less realistic values of β; those with large β (e.g., $\beta = 9$) are definitely biased by the cutoff value $i_{\max} = 10$.

B. Violation of Euler sum formula and four-way vertices

Any nonzero value of α produces stationary distributions with a mean smaller than 6, which can be interpreted as follows. For any $\alpha > 0$, the division probability q_i increases with i and so polygons with a small number of sides remain in the tissue for a longer time before they divide than those with a large number of sides. In addition, the dividing smaller polygon classes do not directly decay into the larger ones (e.g., triangles necessarily produce only triangular and quadrilateral daughters in all schemes as shown by the matrices in Appendix A) but they do result from the division of the larger polygon classes as implied by the form of the P matrices which is generally lower triangular with an additional first diagonal above the main one (Appendix A). At finite but small α, this results in a shift of the distribution. On the other hand, if α is much larger than unity, then the smaller polygon classes barely divide but they do gain new sides because of the division of their neighbors, which divide with a very high probability; in turn, the larger classes do not gain any new sides from the division of their neighbors, which have fewer sides. In this regime, the $\widetilde{\mathsf{S}}$ matrix depopulates the smaller polygon classes but does not affect the larger ones.

The above effects all stem from the well-established fact that the probability of division is not the same in all polygon classes, which is included in our asynchronous-division model but not in the synchronous-division model of Ref. [12] and evidently leads to the shift of the stationary distributions of the asynchronous-division model. Moderate shifts where the mean of the distribution is between 5.8 and 6 are consistent with experimental observations [1] whereas large ones with a mean below 5.8 are not. In this respect, the asynchronous-division model is generally less realistic than the more coarse synchronous-division model which produces distributions with a mean of 6, thereby satisfying the sum rule [Eq. (1)]. To better understand this point, we examine the structure of the latter model and we find that stationary distributions with a mean of 6 appear because of the symmetry of the elements of each row of the P matrix and because of the specific index-raising form of the S matrix (Appendix C). The synchronous-division model meets these two requirements, and so does our Gaussian model which is completely phenomenological and does not rely on any explicit cell division topology.

The symmetry of the rows of the P matrix is quite natural as shown by the various division schemes proposed in Sec. II, but the form of the S matrix, which promotes every polygon class to that just above it in each generation, relies on patently

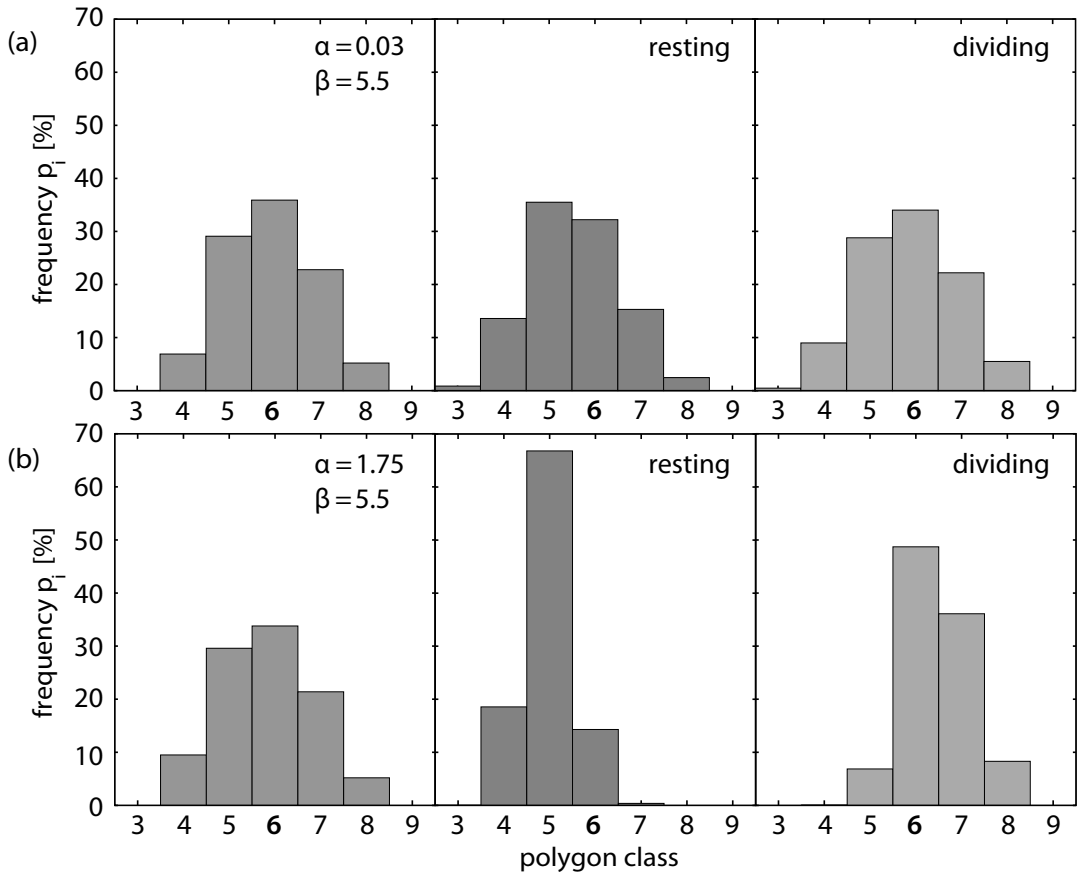

FIG. 9. Distributions of frequencies of polygon classes of the resting (middle) and dividing cell populations (right) within the size-dependent asynchronous-division model for $\beta = 5.5$ and $\alpha = 0.03$ and 1.75 [(a) and (b), respectively]. Also shown are the distributions of the combined cell population (left).

mean-field arguments, and fortuitously yields reasonable results. The shift matrix of the asynchronous-division model [Eq. (10)] goes a step beyond these arguments, building on the Aboav-Weaire law, but leads to stationary distributions with a mean smaller than 6. This may suggest that there exist additional mechanisms that control cell division, addressing, e.g., the orientation of the mitotic plane; this could affect the $\tilde{\mathsf{S}}$ matrix as well as the $\tilde{\mathsf{P}}$ matrix. We relegate the discussion of such an extension to a future study.

As the theoretical framework explored in Ref. [12] and here is based on an abstract rather than a concrete representation of epithelia, we can interpret the obtained stationary distributions solely by noting that those with a mean of 6 are consistent with tilings with three-way vertices whereas those with a mean smaller than 6 correspond to tilings which also contain a fraction of four-way vertices. The latter are seen in some tissues although typically in small numbers, and in the scope of the present framework they may appear within a single generation if the mitotic plane of the reference cell agrees with that of a dividing neighbor. (This, of course, can happen both within the original model of Ref. [12] and its generalizations presented in Sec. II and within the asynchronous-division model.) If the fraction of four-way vertices denoted by f_4 is small, then [1]

$$\mu_1 \approx 6 - 4f_4. \tag{11}$$

This formula can be used to estimate the fraction of four-way vertices corresponding to the distributions in Fig. 8. For example, a 5% fraction of four-way vertices ($f_4 = 0.05$) yields $\mu_1 = 5.8$. This value of f_4 is not unreasonable either because of unresolved pairs of nearby three-way vertices or because of true four-way vertices, and in real tissues μ_1 is indeed smaller than 6 although not smaller than 5.8 [1]. This value may be considered as the lower bound defining the range of α and β where the asynchronous-division model produces results that are consistent with experimental observations [shaded band in Fig. 8(c)].

C. Resting and dividing cell populations in asynchronous-division model

The asynchronous-division model also allows us to distinguish between the resting and the dividing cell populations, which are computed by multiplying the stationary distribution with the respective probabilities $1 - q_i$ and q_i and then renormalizing the two partial distributions for easier comparison with experiments. In Fig. 9, we show the two populations for $\alpha = 0.03$ and 1.75 and $\beta = 5.5$; the model parameters were chosen such that the combined distributions are very similar. As anticipated, for $\alpha = 0.03$ the distributions of polygon classes in the resting and the dividing population are quite close to each other [Fig. 9(a)], with means at 5.544 and 5.847, respectively. But at the larger $\alpha = 1.75$ where the cell division

probability varies quite dramatically with i [Fig. 8(a)], the distributions are markedly different from each other and both narrower than the combined distribution. The resting population is dominated by pentagonal cells with $p_5 \approx 69\%$ and the mean is 4.968 whereas in the dividing population the mean is 6.458. The fact that the shapes of the two distributions are different disagrees with, e.g., Lewis' experimental studies of cucumber epidermis [4] and with the *Drosophila* wing disk epithelium [20], but we expect that by varying α and β and by using division rules other than ST_3^2 (Sec. II D) a range of outcomes may be obtained.

The results shown in Fig. 9 demonstrate that although the stationary distributions produced by the different models can be quite similar, this does not necessarily apply to the partial distributions of resting and dividing cells.

IV. DISCUSSION

To assess the relevance of the models from Secs. II and III, one should compare them to experiments much like in Ref. [12]. To this end, we collected data from various dividing tissues, using the frequency of hexagons as the order parameter that characterizes a given distribution; the tissues analyzed are summarized in Table I. The data can be divided into two categories. The first one contains tissues where growth is characterized by an unbiased mitosis with no preferred orientation of the mitotic plane [4,9,11,12,18,23,24], which appears to be the primary process at the stage where the epithelium is first formed. In these tissues, the frequency of hexagons is typically between 38% and 48% (Fig. 10 and Table I). This category of distributions can be interpreted using rule-based schemes which include certain topological limitations, are restricted to a few near-symmetric division patterns, or account for the combinatorics of division which increases the probability of the near-symmetric patterns. Such schemes produce theoretical frequencies of hexagons typically a little larger than 40%.

The second category consists of tissues whose structure is not controlled only by unbiased mitosis but also by other ordering processes, which lead to tissue specialization in later developmental stages. These processes include oriented mitosis which takes place preferentially in a specific direction [24,26] and tissue rearrangements due to cell disappearance or neighbor exchange [8,25]. Most of them make the tissue more ordered [9], but the processes that lead to cell maturation and differentiation may result in decreased order signaled by a small frequency of hexagons [26].

TABLE I. Epithelial tissues represented in Fig. 10.

Tissue	Process involved	Mitosis only	Ref.
Xenopus laevis (tadpole tail epidermis)	Rapid proliferation, few cell rearrangements	•	[12]
Hydra vulgaris (outer epidermis)	Rapid proliferation, few cell rearrangements	•	[12]
Drosophila melanogaster (larval imaginal wing disk columnar epithelium and peripodial epithelium)	Rapid proliferation, few cell rearrangements	•	[12]
Drosophila melanogaster (third instar larval wing disk)	Rapid proliferation, cell division with few T2 transitions	•	[11]
Drosophila melanogaster (mid-third instar larva and early prepupa)	Proliferation	•	[23]
Drosophila melanogaster (developing wing epithelium: imaginal disk and early prepupa)	Proliferation	•	[9]
Drosophila melanogaster (developing wing epithelium: pupal development before hair formation)	Uniform orientation by planar cell polarity pathway involving neighbor exchanges		[9]
Drosophila melanogaster (embryonic epithelium at stages 6, 7, 8, and late stage 8)	Neighbor exchange during germband extension		[8]
Anagalis arvensis (meristem)	Rapid proliferation, few cell rearrangements proliferation	•	[18]
Cucumis (epidermis)		•	[4]
Allium cepa (abaxial epidermis of scales)	Uniform tissue proliferation	•	[24]
Euonymus fortunei var. *Vegetus* (adaxial epidermis of leaf)	Uniform tissue proliferation	•	[24]
Dryopteris filix-mas (gametophytic cells)	Uniform tissue proliferation	•	[24]
Anarchis densa (surface cells of bud)	Uniform tissue proliferation	•	[24]
Anarchis densa (abaxial and adaxial surface of leaf)	Proliferation in preferred direction		[24]
Volvox aureus (colony)	Proliferation, alternating mitotic planes		[24]
Homo sapiens (epidermis)	Division and cell detachment		[25]
Gallus gallus (spinal cord neuroepithelium)	Proliferation with wide range of mitotic plane orientations (perpendicular and parallel to apical surface) and cell detachment		[26]

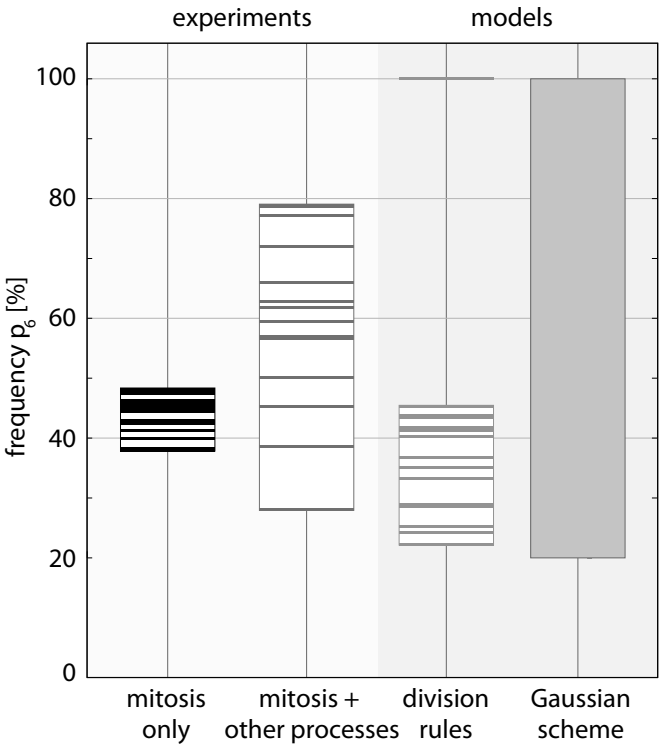

FIG. 10. Experimental frequencies of hexagons p_6 in tissues characterized by (i) mitosis with unbiased orientation of the mitotic plane and by (ii) mitosis complemented by other processes that contribute to cell rearrangement compared to the results of (iii) the rule-based models and (iv) the phenomenological Gaussian scheme. Each bar represents one set of experimental data from tissues listed in Table I or one theoretical stationary distribution as appropriate.

As a result, the spread of p_6 in tissues from the second category is much bigger than in tissues from the first category (Fig. 10), the former including tissues where the frequency of hexagons is between 28% and 78%. Those second-category cases where p_6 exceeds 45% cannot be accounted for by any of our rule-based division schemes, in which p_6 ranges between 22% and 45% or equals 100% depending on the scheme. This should not be too surprising as these schemes do not include any information on the tissue aside from the topological and statistical description of mitosis and the implementation of any geometrical feature of the whole tissue such as a preferred orientation of the mitotic plane is simply beyond their scope.

As far as reproducing the more ordered, large-p_6 distributions from the second category is concerned, the size-dependent asynchronous-division version of the model fares no better than the original scheme. We implemented it using the rule-based schemes that give hexagons-only stationary distributions (e.g., the S_3 symmetric division scheme), expecting that the asynchronous-division protocol will broaden it to some extent. This indeed happens at a finite α but the stationary distributions obtained (not shown) are all rather anomalous, containing quadrilaterals, pentagons, and hexagons but no heptagons and octagons. While a mean $\mu_1 < 6$ is seen in the more ordered epithelia with a large frequency of hexagons, a complete absence of heptagons is not consistent with experiments [8,9]. We conclude that the asynchronous-division model cannot provide an interpretation of the structure of these more ordered types of epithelia. At the same time, the phenomenological Gaussian scheme can reproduce stationary distributions with an arbitrary p_6 solely by varying σ (Fig. 10), but does not offer much new insight.

The above outcome is still a lesson learned, namely, that additional processes must be included in the model or that cell division must be biased, say by planar polarity of the tissue. Some of these features can conceivably be incorporated in the topological model of Ref. [18].

V. CONCLUSIONS

The schemes described here extend and complement the original model of Ref. [12], showing how the different cell division rules affect the distribution of polygon classes in an epithelium. Our rule-based schemes lead to a group of distributions qualitatively similar to that of Ref. [12] but also to many broader ones, which represent less ordered tissues, and to the hexagon-only distribution corresponding to a topologically perfectly ordered tissue. These schemes offer a more comprehensive interpretation of the structure of real proliferating tissues unaffected by any process other than mitosis with an unbiased orientation of the mitotic plane. They also cover some but not all tissues where cell division is biased or accompanied by one or more modes of cell rearrangement.

Representing the tissue by the frequencies of polygon classes rather than by a physical structure is an extreme case of coarse graining, and the fact that a theory revolving solely around the frequencies works is rather surprising at first sight. At the same time, any physical tiling representing a tissue is subject to constraints such the no-gap and no-overlap rule applying to all edge-to-edge tilings and the sum rule [Eq. (1)], which impose a strong restriction on the distribution of the frequencies. As these constraints can be included in a model based on the frequencies-only representation of the tissue, such as model can produce meaningful results together with a more transparent insight into the actual processes that determine the tissue structure. Our asynchronous-division model shows how the original idea from Ref. [12] can be refined so as to include well-known empirical facts pertaining to most random tilings, which then provides an *a posteriori* validation of the original model.

Neglecting the cell-level mechanics (forces and physical constraints) may indeed seem a gross simplification. On the other hand, there exists evidence that several features characteristic of epithelial tissues do not depend on the underlying energy functional; for example, the height of the energy barrier and the energy gain after a T1 neighbor-exchange transition in an epithelium as a function of projected edge length [27,28] are essentially universal. Together with the fact that cell division virtually completely overrides the minimal-energy structure of the tissue [11], it suggests that models such as that from Ref. [12], its generalizations proposed here, and other approaches stripped of cell geometry such as the topological model of Ref. [18] are well worth exploring further. One of the next steps in this direction may be the continuous-time variant of the asynchronous-division model.

ACKNOWLEDGMENTS

This project has received funding from the European Union's Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie Grant Agreement No. 642774. The authors acknowledge the financial support from the Slovenian Research Agency (research core funding Grants No. P1-0055 and No. P1-0208).

APPENDIX A: P MATRICES FOR DIVISION SCHEMES R_3, PR_3, S_3, S_3^2, AND T_3

Here, we spell out the P matrices for the division schemes from Secs. II A–II E for the case where the smallest polygon class is triangles. All matrices begin with the P_{33} entry in the top-left corner, which gives the probability that a three-sided mother cell produces a three-sided daughter, and the last row shown corresponds to the probabilities of division of a seven-sided mother into daughters with three, four, . . . , eight sides.

In the random division scheme R_3,

$$\mathsf{P} = \begin{bmatrix} 1/2 & 1/2 & & & & \\ 1/3 & 1/3 & 1/3 & & & \\ 1/4 & 1/4 & 1/4 & 1/4 & & \\ 1/5 & 1/5 & 1/5 & 1/5 & 1/5 & \\ 1/6 & 1/6 & 1/6 & 1/6 & 1/6 & 1/6 \\ & & \cdots & & & \end{bmatrix}. \tag{A1}$$

In the partly random division scheme PR_3,

$$\mathsf{P} = \begin{bmatrix} 1/2 & 1/2 & & & & \\ 3/10 & 4/10 & 3/10 & & & \\ 2/10 & 3/10 & 3/10 & 2/10 & & \\ 15/105 & 24/105 & 27/105 & 24/105 & 15/105 & \\ 3/28 & 5/28 & 6/28 & 6/28 & 5/28 & 3/28 \\ & & \cdots & & & \end{bmatrix}. \tag{A2}$$

In the symmetric division scheme S_3,

$$\mathsf{P} = \begin{bmatrix} 1/2 & 1/2 & & & \\ & 1 & & & \\ & 1/2 & 1/2 & & \\ & & 1 & & \\ & & 1/2 & 1/2 & \\ & & \cdots & & \end{bmatrix}. \tag{A3}$$

In the near-symmetric division scheme allowing the first and the second most symmetric patterns with equal probabilities S_3^2,

$$\mathsf{P} = \begin{bmatrix} 1/2 & 1/2 & & & \\ 1/3 & 1/3 & 1/3 & & \\ 1/4 & 1/4 & 1/4 & 1/4 & \\ & 1/3 & 1/3 & 1/3 & \\ & 1/4 & 1/4 & 1/4 & 1/4 \\ & & \cdots & & \end{bmatrix}. \tag{A4}$$

In the topological division scheme T_3,

$$\mathsf{P} = \begin{bmatrix} 1/2 & 1/2 & & & & \\ 1/4 & 2/4 & 1/4 & & & \\ 1/6 & 2/6 & 2/6 & 1/6 & & \\ 1/9 & 2/9 & 3/9 & 2/9 & 1/9 & \\ 1/12 & 2/12 & 3/12 & 3/12 & 2/12 & 1/12 \\ & & \cdots & & & \end{bmatrix}; \tag{A5}$$

one can immediately see that the fourth row corresponding to the division of hexagons is consistent with Fig. 5. For comparison, the entries in the P matrix based on Eq. (4) adapted so as to include triangles are given by

$$P_{ij} = \frac{C_{\mathrm{comb}}(i-2, j-3)}{2^{i-2}} \tag{A6}$$

because each daughter cell must inherit 1 vertex from the mother so that the total number of vertices of the i-sided mother that can be divided randomly is $i-2$ rather than $i-4$ like in Eq. (4). This gives

$$\mathsf{P} = \begin{bmatrix} 1/2 & 1/2 & & & & \\ 1/4 & 2/4 & 1/4 & & & \\ 1/8 & 3/8 & 3/8 & 1/8 & & \\ 1/16 & 4/16 & 6/16 & 4/16 & 1/16 & \\ 1/32 & 5/32 & 10/32 & 10/32 & 5/32 & 1/32 \\ & & \cdots & & & \end{bmatrix}. \tag{A7}$$

APPENDIX B: FORM OF ASYNCHRONOUS-DIVISION SHIFT MATRIX $\widetilde{\mathsf{S}}$

The first term in Eq. (10) depends on the probability that the reference i-sided cell does not gain an additional side due to the division of its i neighbors, which have $m(i)$ sides on average, and reads as $1 - q_{m(i)}$. One may expect that this probability should be given by $1 - iq_{m(i)}$ instead as there are i neighbors rather than just 1, but such a conclusion is incorrect because not all divisions of a neighbor increase the number of sides of the reference cell as the mitotic plane may not intersect the side shared by the cell and its dividing neighbor. If the orientation of the mitotic plane is isotropic, then only $1/m(i)$ divisions of each of the i neighbors will increase the number of sides of the reference cell by 1.

More precisely, if the exact numbers of sides of neighbors of the reference cell were known, then the total probability that the reference cell will gain a vertex because its neighbors divide would be

$$\sum_{l=1}^{i} \frac{q_{k_l}}{k_l}, \tag{B1}$$

where $k_1, k_2, \ldots, k_i$ are the numbers of sides of the neighbors and $q_{k_1}, q_{k_2}, \ldots, q_{k_i}$ are their respective division probabilities. Since in the statistical representation of the dividing tissue k_l's are unknown, we could try approximating expression (B1) in a mean-field manner by

$$i\frac{q_{m(i)}}{m(i)}, \tag{B2}$$

where $m(i)$ is the average number of sides of the neighbors of the reference i-sided cell, which is given by the Aboav-Weaire law [Eq. (9)].

In the present version of the asynchronous-division model, this effect cannot be readily included because it leads to negative entries in the $\widetilde{\mathsf{S}}$ matrix, at least for some values of α [Eq. (7)]. For example, at $\alpha = 0$, $q_i = 1$ for all i and thus the magnitude of the first term in Eq. (10) would be $1 - i/m(i)$, which is negative for $i > 6$ where $m(i) < 6$. Naturally, this unphysical situation arises because the time step of our

asynchronous-division algorithm is chosen such that within one step, the largest polygon classes divide with a probability of 100%. In a continuous-time version with an infinitesimally short time step, the division probability within a step would be much smaller than 100% for all polygon classes, and this issue would not arise.

In this paper, we do not explore such a model; instead, we replace the expression (B1) by $q_{m(i)}$ so as to arrive at the form of the $\widetilde{\mathsf{S}}$ matrix in Eq. (10). We note that with this choice our asynchronous-division model reduces to the original model of Ref. [12] for $\alpha = 0$, i.e., when all $q_i = 1$.

APPENDIX C: SUM RULE AND FORM OF P AND S MATRICES

The model of Ref. [12] leads to stationary distributions which satisfy the sum rule [Eq. (1)] automatically due to the specific, index-raising form of the S matrix, the triangular nature of the P matrix, and the symmetry of the nonzero elements in the columns of the P matrix [Eq. (6)], the latter being a consequence of the symmetry of polygon division, i.e., the fact that each daughter with j sides requires a complementary daughter with $i + 4 - j$ sides. To show this, we spell out Eq. (3) for a stationary state where $\mathbf{p}_t = \mathbf{p}_{t-1}$ row by row:

$$\begin{aligned} p_5 &= P_{44}p_4 + P_{54}p_5 + P_{64}p_6 + P_{74}p_7 + \cdots, \\ p_6 &= 0 + P_{55}p_5 + P_{65}p_6 + P_{75}p_7 + \cdots, \\ p_7 &= 0 + 0 + P_{66}p_6 + P_{76}p_7 + \cdots, \\ p_8 &= 0 + 0 + 0 + P_{77}p_7 + \cdots, \\ \ldots &= \cdots + \cdots + \cdots + \cdots + \cdots; \end{aligned} \tag{C1}$$

here the effect of the S matrix is already taken into account and we assume that the smallest polygon class are quadrilaterals.

Now, we sum the columns of these equations. The sum of all left-hand sides is $p_5 + p_6 + p_7 + \cdots = 1$ as the $\mathbf{p}$ vector is normalized; note that in Ref. [12] $p_4 = 0$ by construction because triangular cells are not included in the scheme and all quadrilateral cells are promoted to pentagonal cells upon the action of the S matrix. Because all rows of the P matrix are normalized [12], the sum of the first column on the right-hand side is $P_{44}p_4 = p_4$, the sum of the second column is $P_{54}p_5 + P_{55}p_5 = p_5$, the sum of the third column is $P_{64}p_6 + P_{65}p_6 + P_{66}p_6 = p_6$, etc.

The mean of the distribution $\mu_1 = \sum_i i p_i$ can be written by adding Eqs. (C1) multiplied by an appropriate i (i.e., the first equation by 5, the second by 6, etc.). After collecting the terms with the same p_i we find that

$$\begin{aligned} \mu_1 = {} & 5P_{44}p_4 \\ & + (5P_{54} + 6P_{55})p_5 \\ & + (5P_{64} + 6P_{65} + 7P_{66})p_6 \\ & + (5P_{74} + 6P_{75} + 7P_{76} + 8P_{77})p_7 \\ & + \cdots. \end{aligned} \tag{C2}$$

Due to (i) the fact that the entries in each row of the P matrix in the triangle below the diagonal are symmetric ($P_{44} = 1$, $P_{54} = P_{55}$, $P_{64} = P_{66} < P_{65}$, etc.) and to (ii) the normalization of each row of this matrix we can sum the terms in each bracket to find that

$$\begin{aligned} \mu_1 &= 5p_4 + 5.5p_5 + 6p_6 + 6.5p_7 + 7p_8 + 7.5p_9 + \cdots \\ &= \sum_i i p_i + \frac{1}{2}\sum_i (6 - i)p_i \\ &= \mu_1 + \frac{1}{2}\sum_i (6 - i)p_i. \end{aligned} \tag{C3}$$

Evidently, the second term on the right-hand side must vanish, which immediately gives the sum rule [Eq. (1)].

[1] M. P. Miklius and S. Hilgenfeldt, Eur. Phys. J. E **34**, 50 (2011).
[2] J. T. Blankenship, S. T. Backovic, J. S. P. Sanny, O. Weitz, and J. A. Zallen, Dev. Cell **11**, 459 (2006).
[3] F. T. Lewis, Anat. Rec. **33**, 331 (1926).
[4] F. T. Lewis, Anat. Rec. **38**, 341 (1928).
[5] D. Sánchez-Guitiérrez, M. Tozluoglu, J. D. Barry, A. Pascual, Y. Mao, and L. M. Escudero, EMBO J. **35**, 77 (2016).
[6] D. A. Aboav, Metallography **3**, 383 (1970).
[7] D. Weaire, Metallography **7**, 157 (1974).
[8] J. A. Zallen and R. Zallen, J. Phys.: Condens. Matter **16**, S5073 (2004).
[9] A. K. Classen, K. I. Anderson, E. Marois, and S. Eaton, Dev. Cell **9**, 805 (2005).
[10] R. Blumenfeld and S. F. Edwards, Eur. Phys. J. E **19**, 23 (2006).
[11] R. Farhadihar, J. C. Röper, B. Aigouy, S. Eaton, and F. Jülicher, Curr. Biol. **17**, 2095 (2007).
[12] M. C. Gibson, A. B. Patel, R. Nagpal, and N. Perrimon, Nature (London) **442**, 1038 (2006).
[13] S. A. Sandersius, M. Chuai, C. J. Weijer, and T. J. Newman, PLoS One **6**, e18081 (2011).
[14] A. Hočevar and P. Ziherl, Phys. Rev. E **80**, 011904 (2009).
[15] M. L. Manning, R. A. Foty, M. S. Steinberg, and E.-M. Schoetz, Proc. Natl. Acad. Sci. USA **107**, 12517 (2010).
[16] P. Gómez-Gálvez, P. Vicente-Munuera, A. Tagua, C. Forja, A. M. Castro, M. Letrán, A. Valencia-Expósito, C. Grima, M. Bermúdez-Gallardo, Ó. Serrano-Pérez-Higueras, F. Cavodeassi, S. Sotillos, M. D. Martín-Bermudo, A. Márquez, J. Buceta, and L. M. Escudero, Nat. Commun. **9**, 2960 (2018).
[17] A. Abdullah, D. Avraam, O. Chepizhko, T. Vaccari, S. Zapperi, C. A. M. La Porta, and B. Vasiev, arXiv:1710.08527.
[18] A. B. Patel, W. T. Gibson, M. C. Gibson, and R. Nagpal, PLoS Comput. Biol. **5**, e1000412 (2009).
[19] In our experience, the numerical results do not significantly depend on $i_{\max}$ if $\alpha \ll 1$ or $\alpha \gg 1$, i.e., in the two realistic regimes of α.
[20] T. Aegerter–Wilmsen, A. C. Smith, A. J. Christen, C. M. Aegerter, E. Hafen, and K. Basler, Development **137**, 499 (2010).

[21] C. Besson, F. Bernard, F. Corson, H. Rouault, E. Reynaud, A. Keder, K. Mazouni, and F. Schweisguth, Curr. Biol. **25**, 1104 (2015).

[22] This is not necessarily the case for every division rule because of the cutoff at $i_{\max}$ used in the asynchronous-division model; e.g., for the T_3 scheme we observe small deviations manifested in the first moment of the distribution, the mean, which is 5.953 at $\alpha = 0$ and not exactly 6 like in the synchronous-division model.

[23] D. Sánchez-Guitiérrez, A. Sáez, A. Pascual, and L. M. Escudero, PLoS One **8**, e79227 (2013).

[24] R. W. Korn and R. M. Spalding, New Phytol. **72**, 1357 (1973).

[25] B. Dubertret and N. Rivier, Biophys. J. **73**, 38 (1997).

[26] A. C. Wilcock, J. R. Swedlow, and K. G. Storey, Development **134**, 1943 (2007).

[27] S. Kim, Y. Wang, and S. Hilgenfeldt, Phys. Rev. Lett. **120**, 248001 (2018).

[28] M. Krajnc, S. Dasgupta, P. Ziherl, and J. Prost, Phys. Rev. E **98**, 022409 (2018).