Spatial-Temporal Modelling and Analysis of Bacterial Colonies with Phase Variable Genes

OVIDIU PÂRVU and DAVID GILBERT, Brunel University
MONIKA HEINER, Brandenburg University of Technology
FEI LIU, Control and Simulation Center, Harbin Institute of Technology
NIGEL SAUNDERS and SIMON SHAW, Brunel University

This article defines a novel spatial-temporal modelling and analysis methodology applied to a systems biology case study, namely phase variation patterning in bacterial colony growth. We employ coloured stochastic Petri nets to construct the model and run stochastic simulations to record the development of the circular colonies over time and space. The simulation output is visualised in 2D, and sector-like patterns are automatically detected and analysed. Space is modelled using 2.5 dimensions considering both a rectangular and circular geometry, and the effects of imposing different geometries on space are measured. We close by outlining an interpretation of the Petri net model in terms of finite difference approximations of partial differential equations (PDEs). One result is the derivation of the "best" nine-point diffusion model. Our multidimensional modelling and analysis approach is a precursor to potential future work on more complex multiscale modelling.

Categories and Subject Descriptors: I.6.5 [**Simulation and Modeling**]: Model Development; I.6.4 [**Simulation and Modeling**]: Model Validation and Analysis; I.4.6 [**Image Processing and Computer Vision**]: Segmentation

General Terms: Algorithms, Experimentation, Theory

Additional Key Words and Phrases: Coloured stochastic Petri nets, partial differential equations, ordinary differential equations, spatial-temporal modelling, systems biology, pattern analysis, multidimensional, biomodel engineering

ACM Reference Format:
Ovidiu Pârvu, David Gilbert, Monika Heiner, Fei Liu, Nigel Saunders, and Simon Shaw. 2015. Spatial-temporal modelling and analysis of bacterial colonies with phase variable genes. ACM Trans. Model. Comput. Simul. 25, 2, Article 13 (May 2015), 25 pages.
DOI: http://dx.doi.org/10.1145/2742546

This research has been partially funded by British EPSRC Research Grant EP I036168/1 and German BMBF Research Grant 0315449H. F. Liu was supported by the National Natural Science Foundation of China (61273226), and O. Pârvu was supported by a scholarship from Brunel University.
Authors' addresses: O. Pârvu, Room HNZW011B, Heinz Wolff building, Department of Computer Science, Brunel University London, Kingston Lane, Uxbridge, Middlesex, UB8 3PH, London, United Kingdom; email: ovidiu.parvu@brunel.ac.uk; D. Gilbert, Room 301L, Eastern Gateway building, Department of Computer Science, Brunel University London, Kingston Lane, Uxbridge, Middlesex, UB8 3PH, London, United Kingdom; email: david.gilbert@brunel.ac.uk; M. Heiner, Room 2.38, Building VG 1c, Department of Computer Science, Brandenburg University of Technology Cottbus - Senftenberg, Walther-Pauer-Strasse 2, D-03046, Cottbus, Germany; email: monika.heiner@b-tu.de; F. Liu, 92 West Dazhi Street, Control and Simulation Center, Harbin Institute of Technology, Nan Gang District, 150001, Harbin, China; email: liufei@hit.edu.cn; N. Saunders, Room HNZW015, Heinz Wolff building, Department of Biosciences, Brunel University London, Kingston Lane, Uxbridge, Middlesex, UB8 3PH, London, United Kingdom; email: nigel.saunders@brunel.ac.uk; S. Shaw, Room JNCK302, John Crank building, Department of Mathematics, Brunel University London, Kingston Lane, Uxbridge, Middlesex, UB8 3PH, London, United Kingdom; email: simon.shaw@brunel.ac.uk.

ACM 1049-3301/2015/05-ART13 $15.00
DOI: http://dx.doi.org/10.1145/2742546

1. MOTIVATION

Computational models are employed in systems biology [Ideker et al. 2001; Kitano 2002] to explain the mechanisms underlying physiological processes and to predict how the system behaviour changes when the system is perturbed.

The modelling requirements may depend on the considered level of organisation. At intracellular or more fine-grained levels, it is often assumed that species (e.g. proteins/ molecules) are uniformly distributed, and therefore computational models only capture how their average concentration changes over time; explicitly considering individual species and their evolution over both time and space at fine-grained resolutions usually leads to an explosion in the model simulation time. Conversely, at cellular and more coarse-grained levels, it is assumed that the heterogeneity of species (e.g. cells) is important because it can lead to the development of different patterns/structures in space. Therefore, at such higher levels of organisation, computational models record how the number/density of species evolves both over time and space.

To support the development of spatial computational models, corresponding modelling formalisms have been developed; they represent the spatial domain in either a continuous or discrete fashion.

Continuous spatial models are usually encoded as partial differential equations (PDEs) [Schaff et al. 1997] and represent variations of reaction-diffusion [Kondo and Miura 2010] or predator-prey [Arditi et al. 2001] systems, respectively chemotactic movement of cells [Hillen and Painter 2009]. The main reason for modelling processes such as diffusion (reaction-diffusion) or population variation (predator-prey, chemotaxis) using continuous approaches is that only the average density of the species is of interest for each time point and position in space.

Conversely, in case the interactions between individual species are of interest, discrete spatial models could be employed instead. Representative discrete spatial modelling formalisms that employ a lattice-based representation of space and local rules to specify how the system changes from one state to the next are Cellular Automata (CA) [Deutsch and Dormann 2007, Chapters 5–11] and Glazier-Graner-Hogeweg (GGH) [Balter et al. 2007; Graner and Glazier 1992] models (also known as Cellular Potts). In contrast to standard CA and GGH models, individual-based models (IBMs) [An et al. 2009; Thorne et al. 2007] can employ either an on-lattice or off-lattice spatial representation, and their evolution over time is determined by rules specific to individuals (or agents) instead of lattice positions. Modelling formalisms that are not inherently spatial but have been extended with spatial attributes recording the species' position in space (e.g. coordinates in Euclidean space) include process algebras [Feng and Hillston 2014; John et al. 2010], rule-based modelling languages [Blinov et al. 2004; Danos et al. 2007; John et al. 2011; Maus et al. 2011; Nikolić et al. 2012] and P (or membrane) systems [Barbuti et al. 2011; Besozzi et al. 2008].

Similarly, Heiner and Gilbert [2011] and Gilbert et al. [2013] describe how hierarchical and coloured extensions of Petri nets can be employed to construct discrete spatial computational models. One of the main advantages of coloured Petri nets (CPNs) as a modelling formalism, and the reason for employing them here, is that they enable constructing complex stochastic and/or deterministic models using the same framework without the need to integrate multiple approaches in a hybrid model. Moreover, CPNs are well-balanced in their modelling and analysis power; CPNs (with finite discrete colour sets) can be automatically unfolded, and thus all analysis/simulation techniques available for the uncoloured Petri nets can be reused for the analysis/simulation of CPNs.

An initial example of a spatial computational model constructed using hierarchical CPNs is presented in Gao et al. [2013], where the spatial domain is represented at the top level as a hexagonal lattice and a clustering based approach is employed to detect discretised spatial compartments with similar biologically relevant features.

Similarly, it was shown in Liu et al. [2014] that CPNs could be employed for modelling 2D reaction-diffusion systems considering a rectangular lattice.

The preliminary paper [Pârvu et al. 2013] illustrates how the same framework can be employed to construct 2.5D spatial computational models considering both rectangular and circular geometries, and introduces an algorithm for automatically computing the geometric properties (e.g. area) of sector-like patterns in the final state of model simulations.

The main contributions of our article with respect to Pârvu et al. [2013] are as follows:

—A detailed explanation of the motivation for modelling phase variation in bacterial colony growth and its potential practical applications
—An implementation of a spatio-temporal analysis algorithm that enables analysing how the geometric properties of spatial patterns change over time
—An initial attempt to generalise the CPN spatial modelling approach by establishing a connection between CPNs and discretised approximations of PDEs.

We illustrate our approach based on a systems biology case study, namely phase variation patterning in bacterial colony growth. The spatial-temporal modelling and analysis methodology described in this article is a precursor to the development of more complex multiscale models.

The article is organised as follows. The computational motivation, biological background, and basic model are described in Section 2. The representation of space considering circular and rectangular geometries using Cartesian and polar coordinates is described in Section 3, and the analysis is presented in Section 4. The connection between Petri net models and (systems of) PDEs is given in Section 5. We conclude with a summary in Section 6.

Additional information is provided in the Online Appendix, in which models and analysis tools are made available.

2. PHASE VARIATION IN BACTERIAL COLONY GROWTH

2.1. Background

A large proportion of bacterial species use a stochastic gene switching process called *phase variation*, controlled by reversible genetic mutations, inversions, or epigenetic modification [Salaün et al. 2003]. The key features of phase variation are a reversible expression of a gene that occurs without the loss of coded information so that the *on* state is fully and identically recoverable, in a bistable process—distinguishing it from antigenic variation through recombination, and the predictability of the subset of phase varied genes—due to the presence of specific features that confer switchability on the phase variable gene set. When in the *on* configuration, the expression of the gene remains integrated and under the control of traditional transcriptional regulatory processes, but when in the *off* configuration, it cannot be expressed regardless of other influences on expression.

Understanding of the adaptive and evolutionary role of phase variation has traditionally been within the context of “contingency gene theory” [Moxon et al. 1994]. Under this conceptual model, the genes that have evolved to be phase variable mediate adaptations to historically repeated, and therefore future predictable, environmental transitions. This can include functions for adhesion to environmental surfaces versus release into the environment for dissemination, the ability to utilize different sources essential factors for growth such as iron, or supposedly parasitic genes such as restriction enzymes or transposon expression. However, the primary focus of this work to date has been with regard to genes that adapt pathogenic organisms to different hosts, host niches, and the evasion of host defences and immune responses, the latter of which

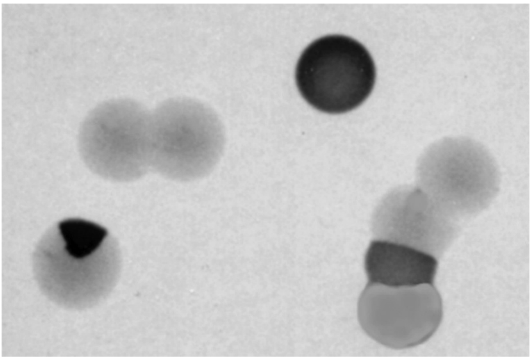


Fig. 1. An illustrative bacterial colony with phase variable genes is depicted in the lower left part of the image. Sector-like patterns (highlighted in black) represent high proportions of "mutant" cells; respectively, areas of the colony highlighted in gray represent high proportions of wild-type cells. (Reproduced with permission from de Vries et al. [2002]. Copyright © 2002, American Society for Microbiology.)

represents a predictable and dynamic environmentally selective pressure within each infected host [Salaün et al. 2005].

In this context, one of the commonly used switching determinants (simple sequence repeats that are subject to changes in length during chromosomal replication) have been used to identify genes likely to be important in host interactions and immune evasion in several species, including *Haemophilus influenzae* [Hood et al. 1996], *Helicobacter pylori* [Salaün et al. 2004; Saunders et al. 1998], and *Neisseria* [Saunders et al. 2000; Snyder et al. 2001]. These and other studies have shown that some species have several tens of such genes, which can control huge numbers of potential gene expression combinations because the genes are switched independently. In the context of this evolutionary and group selection conceptual framework models were developed for the roles of switching rate and selective pressure/relative fitness on population composition [Saunders et al. 2003].

However, there are other potential roles for phase variation. It has been shown that global cellular behaviours are influenced by which iron source is used by *Neisseria meningitidis* [Jordan and Saunders 2009], which depends upon which phase variable receptors are expressed, and that other behaviours depend upon the phase varied expression of restriction enzyme methylases [Srikhanta et al. 2005]. In addition, given the complexity of bacterial structures in vivo, such as biofilms, and the theoretical advantages of cooperativity within bacterial populations from specialization of metabolic functions, we hypothesize an additional role for phase variation: specifically, population functional differentiation and structuring. In this context, the existing models are inadequate for understanding the role of phase variation and the presence of diverse switched phenotypes within populations because they do not include spatial components, which are essential for future considerations of interaction and interdependencies. Progress in this area requires the design of new models, moving from existing models of population proportions in freely competing populations to ones that include and address spatial and structural composition and interfaces. This is the translational focus of the work presented here.

The most readily observable compositional effect of phase variation in cultures grown in vitro is colonial sectoring; Figure 1 provides an example. The relation between the geometric properties (e.g. angle, area, shape) of the sector-like patterns (macroscale) and the mutation and/or fitness rates of the bacteria (microscale) is unknown and could be investigated using a computational model.

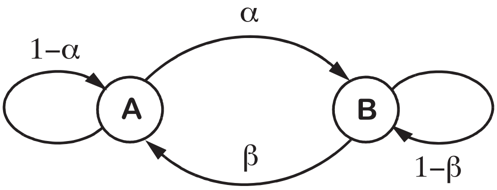


Fig. 2. Phase variation, basic scheme. α/β – forward/backward mutation rates. During cell division, the rate at which type A offspring cells mutate to type B equals α; respectively, the rate at which they preserve their parent's type equals $1 - \alpha$. Similarly, the rate at which type B offspring cells mutate to type A equals β; respectively, the rate at which they preserve their parent's type equals $1 - \beta$.

The problem of modelling the development of bacterial colonies in space has been addressed previously in the literature using two different types of approaches, continuous and discrete, and one or multiple levels of organisation (e.g. intracellular and cellular [Kugler et al. 2010]).

In the continuous case, the spatial domain is usually constrained by a set of boundary conditions and is represented as a system of PDEs (e.g. reaction-diffusion system [Mimura et al. 2000]), which describes how the average number of bacteria changes over time at each position in space.

Conversely, in the discrete case, the continuous spatial domain is approximated by a set of discretised spatial positions. Modelling formalisms employing a discrete representation of space that have been used to model bacterial colony growth include (lattice-gas–based) CA [Ermentrout and Edelstein-Keshet 1993], GGH models [Alber et al. 2003; Poplawski et al. 2008], and IBMs [Ferrer et al. 2008]. Both GGH and CA models represent space as a regular lattice and define local rules that specify how the state (e.g. presence/absence of bacteria) of each lattice position changes from one time point to the other. From a spatial point of view, the main difference between GGH and CA is that in CA models, bacteria are assumed to be of the same shape and size and occupy only one lattice position, whereas in GGH models, bacteria can be of different shape and size and can occupy multiple lattice positions. In contrast to GGH and CA models, in IBMs bacteria can move considering either an on-lattice (e.g. regular rectangular grid) or an off-lattice (e.g. finite subdomain of the Euclidean space) spatial representation [Macal and North 2010].

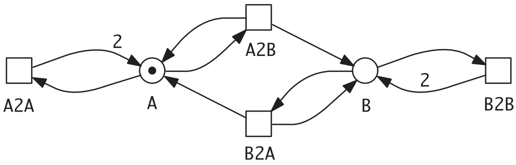
In this article, we introduce a new discrete lattice-based spatial modelling approach for bacterial colony growth based on CPNs. We describe preliminary stochastic models that address colonial patterning, including bidirectional reversible switching between two phenotypes, biologically relevant rates, and differences in the fitness of the two alternate phenotypes. We consider a colony of bacteria with two phenotypes A (*on* state) and B (*off* state), which develop over time by cell division. Cell division may involve cell mutation, and back mutation alternates phenotypes (Figure 2).

During replication, the offspring of type A cells mutate to type B with probability α and preserve their type with probability $1 - \alpha$. Conversely, the offspring of type B cells mutate to type A with probability β and preserve their type with probability $1 - \beta$. Therefore, the rate at which type A/B offspring cells are produced from parent cells of the same type equals the overall birth rate (1) minus the mutation rate (α/β) of type A/B cells.

We are interested in the proportion of phenotypes in the cell generations and how their spatial distribution evolves over time.

2.2. Basic Model of Phase Variation

We start with the equations taken from the previous *deterministic* model of phase variation [Saunders et al. 2003], which describe *synchronous* growth in cell colonies



$$v(A2B) = d_A \alpha A,\ v(A2A) = d_A(1-\alpha)A$$
$$v(B2A) = d_B \beta B,\ v(B2B) = d_B(1-\beta)B$$

Fig. 3. $\mathcal{SPN}$ corresponding to Figure 2. Petri net places A and B record the number of type A, respectively type B, cells. Transitions A2A and A2B represent the division of a type A cell such that an offspring of type A, respectively type B, is produced. Similarly transitions B2A and B2B represent the division of a type B cell such that an offspring of type B, respectively type A, is produced. Arcs without explicit weight values have the default weight value 1 associated. Expressions of the form $v(x)$ (e.g. v(A2A)) encode the marking-dependent stochastic rates of the transition x (e.g. A2A).

with two phenotypes A and B, but no spatial aspects. These equations include the assumption that "if phase variation occurs, the progeny consists of one A and one B." Previously [Saunders et al. 2003], behaviour was explored by iterating the equations on a spreadsheet. We develop a stochastic Petri net ($\mathcal{SPN}$) that is directly executable by playing the token game that facilitates its comprehension and permits the exploration of the behaviour by standard analysis and simulation techniques. Our initial $\mathcal{SPN}$ model (Figure 3) adopts an *asynchronous* modelling approach so that cells divide individually. The model parameters were taken from Saunders et al. [2003]; α and β represent the forward and backward mutation rates, and d_A, d_B represent the fitness of phenotype A and B (i.e. the proportions that survive division).

2.3. Derived Measures of Interest

The n-th generation in a synchronous model yields 2^n bacteria. Vice versa, if we know the total number, *total*, of bacteria generated by asynchronous cell division, then we can obtain the corresponding synchronous generation counter n by

$$n = \log_2 total. \tag{1}$$

For example, 26 synchronous generations (which may develop in about 24 hours) end up with a total population size of approximately $67 \cdot 10^6$. We obtain the proportion of phenotypes A and B modelled by the variables A and B by

$$propA = \frac{A}{A+B};\quad propB = \frac{B}{A+B}. \tag{2}$$

Simulating the stochastic model allows us to observe asynchronous population growth such that cells divide individually. Each event (firing of a transition) corresponds to the division of one cell. Consequently, the size of the population will grow in steps by 1, in contrast to the previous synchronous model.

3. ADDING SPACE

The main limitation of the model depicted in Figure 3 is that it does not represent the evolution of the colony in space. To address this issue, we extend the nonspatial model such that the colony is represented in 2.5 dimensions by an explicit discretised 2D grid with an implicit constant maximal height over all grid positions.

3.1. Alternative Geometries

Previous attempts to model bacterial colony growth have represented space employing a rectangular geometry with Cartesian coordinates. However, starting from a small ini-

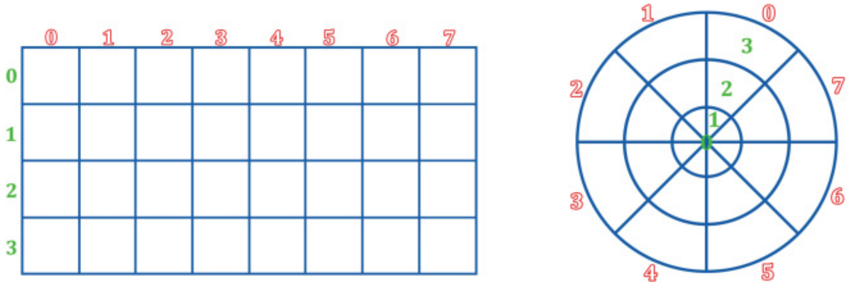


Fig. 4. Discretising space considering Cartesian (left) and polar (right) coordinates. Each annulus in the polar case can be mapped to a row in the grid and each sector to a column such that a position in the grid (left) has one and only one corresponding annular sector (right) and vice versa.

tial population, the colony spreads out as the number of bacteria increases, maintaining a circular shape throughout its development. Therefore, representing space considering a circular geometry with polar coordinates seems to be more appropriate for this particular modelling task.

In addition, the area and shape of the positions in the discretised space varies between geometries as shown in Figure 4. Therefore, depending on the employed spatial representation, the sector-like patterns and/or their geometric properties could potentially differ.

Independently of the chosen spatial representation, each compartment of the discretised space is referenced by a unique tuple (x, y), corresponding to a colour tuple in the model, where x and y are the indices corresponding to the first, respectively second, spatial dimension. Differences between modelling in these two coordinate systems will be highlighted next.

3.1.1. Cartesian Coordinate System. In the Cartesian coordinate system [Weisstein 2013] approach, the 2D space is discretised by splitting it into equally sized rows and columns obtaining a 2D grid as shown in Figure 4 (left). Each grid position is uniquely identified by its corresponding row x and column y, where the rows/columns indices are numbered in increasing top-down/left-right order starting from 0. The area of all positions in the grid is equal. The volume of all grid positions is also equal because their maximal height is the same.

When division occurs, the parent remains in situ and the offspring can either stay with the parent or be displaced to a neighbouring position. The neighbourhood relation between different positions of the grid is represented as a function in the model. The maximum number of neighbours for each position is eight, depending on whether the considered position is in the interior (8) of the grid, at the edge (5), or in the corner (3).

3.1.2. Polar Coordinate System. On the other hand, when considering a polar coordinate system [Weisstein 2013], the 2D space is discretised in a different manner. First of all, the spatial domain is divided into evenly spaced concentric circles. Each one of the concentric circles and its immediate enclosing circle will form an annulus [Weisstein 2013]. All annuli are then split into sectors, obtaining annular sectors as shown in Figure 4 (right).

Each annular sector is uniquely identified by its corresponding annulus x and sector y, where annuli and sectors indices are numbered starting from 0 in increasing inner-outer, respectively anticlockwise, order. The origin of the discretised space is treated as a special annular sector identified by the (x, y) tuple $(0, 0)$. The neighbours of the origin are all of the immediate surrounding annular sectors identified by the set of (x, y) tuples $(1, y)$, $y = \overline{0, \text{number of sectors} - 1}$.

Similar to the neighbourhood relation in a Cartesian coordinate system, all annular sectors except the origin have a maximum of eight neighbours, depending if their position is next to the origin (6), in the interior (8), or at the edge (5).

3.1.3. Comparing the Geometries. One of the differences between the two geometries is that when using the rectangular geometry, the area and volume of all positions in the grid are constant, whereas in the circular geometry, the area and volume are variable. In case of the circular geometry, the variability of the volume of each position in the grid has an effect on the rate functions of the transitions. Conversely, in case of the rectangular geometry, the transition rate function is not influenced by the volume of the positions, as it is constant.

Another important aspect that sets the two geometries apart is the shape of the compartments due to the discretisation process. Let us compare one row from the grid obtained by discretising the space considering a Cartesian coordinate system and the sector obtained similarly by considering a polar coordinate system. The angle described by a row in the grid equals 0 degrees. Conversely, the sides of the sector determine a sharp angle greater than 0 degrees (except when the number of sectors $\rightarrow \infty$).

For this particular case study, we are interested in the angle formed by the patterns of high intensity in the colony. Any sector in the circular geometry will automatically have a nonzero degree angle associated. However, in the rectangular geometry, a nonzero degree angle is formed only if the colony spreads out on multiple rows and columns. To obtain comparable results, we have removed the diagonal movement in the polar coordinates model such that the horizontal spreading of the colony is reduced.

3.1.4. Representing the Geometries Using Colour Sets. The spatial bacterial colony growth models considering both circular and rectangular geometries are represented as CPNs. Despite the multiple differences between the considered geometries, the definition of the colour set for representing space is the same. We denote this colour set as *Grid* and define it as the Cartesian product of the *Grid2D* and *Phenotype* colour sets, where *Grid2D* represents the 2D discretised space and *Phenotype* the type of the bacteria, which in our case is either *A* or *B*; see Online Appendix A for more details.

After unfolding the coloured version of the model, each obtained Petri net place represents a position in the discretised space. The maximum number of bacteria in each place is inversely proportional to the resolution of the grid. Increasing the resolution reduces the maximum capacity of the place, whereas decreasing it makes room for more bacteria.

Cellular division/mutation and displacement to neighbouring positions of the discretised space are encoded in the model by the *division* transition (see Figures 13 and 14, Online Appendix D). The neighbourhood relation between two positions of the discretised space, represented as different Petri net places, depends on the employed geometry. This characteristic is captured by the neighbourhood functions *neighbourhood2D_rectangular* (rectangular geometry) and *neighbourhood2D_circular* (circular geometry) described in Online Appendix B. They define all possible token movements in the net. The neighbourhood function for polar coordinates may appear to be more complicated. However, its length is due to the need for separately considering the neighbours of the origin and not because of an increased complexity.

The rate function for cellular division/mutation depends on the number of parent cells occupying a position of the discretised space. Therefore, the probability of an offspring to be produced is directly proportional to the number of available parent cells. Moreover, in this case study, we are concerned with mutation rates and their influence on the system behaviour. Thus, their total values for each position have to be kept constant irrespective of the number of neighbours. Introducing space means technically multiplying the number of transitions (one for each direction). To counterbalance this

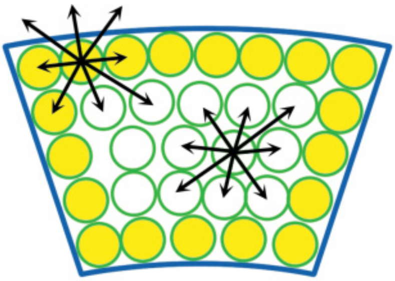


Fig. 5. Interior-edge model used for the circular geometry to represent the probability of a bacterium being displaced to a neighbouring position. Bacteria lying on the edge are highlighted in yellow, bacteria lying in the interior in white, and the annular sector boundary in blue.

effect, we scale the transition rates by dividing them by N, where N is the number of neighbours.

3.2. Controlling the Spatial Dynamic Development of the Clony

3.2.1. Controlling Colony Spreading. The probability of staying with the parent or being displaced to a neighbouring position is modelled differently depending on the spatial representation.

In the circular case, the probability of a bacterium being displaced to a neighbouring position has to take into account the size of the current position, because the area of the annular sectors is variable. We employ the interior-edge model described in Figure 5 to capture this aspect. Considering a particular annular sector, the only bacteria that are able to be displaced from this sector to a neighbouring sector are the ones lying on the edge. Assuming that each bacteria can be displaced in eight directions (N, NW, W, SW, S, SE, E, NE) or remain in situ, only three out of the nine movements of the bacteria on the edge will be to a neighbouring position. The bacteria that lie in the corner are not treated separately in our approach. Thus, the probability of being displaced to a neighbouring position is

$$P = \frac{3}{9} * \frac{\text{Area}_{\text{edge}}}{\text{Area}_{\text{grid position}}} \tag{3}$$

and the probability of staying with the parent is $1-P$. $\text{Area}_{\text{edge}}$ is given by the maximum area that can be occupied by bacteria of size 1×1 μm located around the edge. The difference between the edges and interior of an annular sector is depicted in Figure 5. $\text{Area}_{\text{grid position}}$ is computed as the total area of the annular sector. Both areas depend on the index i of the annulus to which the sector belongs. The value of i is set to 1 for the origin and is incremented with each enclosing annulus. Thus, the values of the areas are

$$\text{Area}_{\text{edge}_i} = \frac{2rN + 2\pi r(2i+1)}{MN}, \ \text{Area}_{\text{grid position}_i} = \frac{\pi r^2(2i+1)}{M^2 N}. \tag{4}$$

where M is the total number of annuli and N is the total number of sectors. A step-by-step description of how the values of $\text{Area}_{\text{edge}}$ and $\text{Area}_{\text{grid position}}$ are computed is given in Online Appendix C.

As the area of annular sectors increases, the ratio between the area on the edge and the total area becomes smaller, which means that the probability of bacteria to be displaced to a neighbouring position decreases.

The reason for considering the area of the overall annular sector edge instead of the (partial) edges shared with individual neighbours is twofold. First of all, explicitly considering edges shared between neighbours requires extending the structure of the model, which would lead to an increased complexity and unfolding/simulation time. Secondly, if the employed spatial resolution is sufficiently high (in our case 101×101) the differences between the areas of the (partial) edges to individual neighbours are small and could be potentially ignored.

In the rectangular case, the area of the grid positions is constant, which means that the model from the circular case would impose a constant probability for all positions in the grid.

To add more flexibility to the model, the probability of staying with the parent or being displaced to one specific neighbouring position is modelled using two preference factors, γ and ω, without changing the total transition rate.

Increasing γ increases the preference to stay with the parent, whereas decreasing γ increases the preference to be displaced. Conversely, increasing ω increases the preference to be displaced to one of the neighbouring positions, whereas decreasing ω increases the preference to stay with the parent. Since the number of neighbours *#neighbours* varies depending on the considered position of the discretised space (e.g. corner), the total preference for a cell to be displaced equals *#neighbours* $\cdot\, \omega$. In the rectangular case, *#neighbours* is equal to 3 if the *grid position* is in the corner, 5 on the edge, and 8 in the interior. Conversely, in the circular case, *#neighbours* is equal to 5 on the edge, 8 in the interior, 6 in the annulus immediately enclosing the origin, and the total number of sectors N for the origin.

The probabilities of staying with the parent or being displaced to a neighbouring position are computed by normalising the corresponding preference factors values as follows:

$$\begin{aligned} P_{\text{stay with parent}} &= \frac{\gamma}{\gamma + (\#neighbours * \omega)} \\ P_{\text{displace to neighbour}} &= \frac{\#neighbours * \omega}{\gamma + (\#neighbours * \omega)}. \end{aligned} \tag{5}$$

All probabilities are encoded in the rate function of the transition *division*, irrespective of the employed geometry.

3.2.2. Controlling Colony Thickness. The bacteria generated by cell division can pile up on top of each other and thus increase the colony thickness at that grid position. This thickness is limited because of the cells' requirements for access to oxygen and nutrients. To control the thickness, we introduce a constant ρ, denoted as $POOLSIZE$ in the $\mathcal{SPN}^{\mathcal{C}}$ model, which limits the maximum number of cells at a certain grid position. This parameter acts as a Boolean condition that enables/prevents cells from replicating without being explicitly considered in rate functions.

For a fixed resolution of the discretised space, the value of ρ is set such that the bacterial colony can develop for 26 generations (i.e. cell count reaches $67 \cdot 10^6 \approxeq 2^{26}$) without reaching the edge of the discretised space. The bacterial colony growth next to the discretised space edge(s) could potentially impact the sector-like patterns development and implicitly alter their geometric properties (e.g. area); this could lead to invalid correlations between the geometric properties and the bacteria' mutation/fitness rates.

The entire set of colour-related definitions common to both circular and rectangular spatial representations and the final version of the models are given in Online Appendices A and D. Rate functions are not described here due to space limitations, but they are defined in the computational models made available as supplementary materials.

The only structural difference between the models is that polar coordinates require additionally one Petri net place and two transitions, which are highlighted in a green colour in the model (see Online Appendix D, Figure 14). The pretransition of the

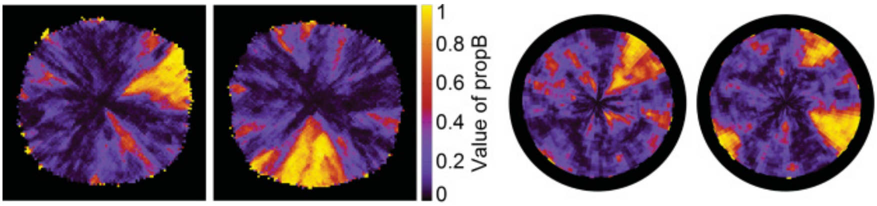


Fig. 6. 2D representation of the final state of four stochastic simulations, two for rectangular (left) and two for circular (right), illustrating the development of sector-like patterns. Due to the stochastic nature of the simulations, the output is different in every run. The value of propB (see Equation (2)) is encoded by colour. Yellow indicates patches with high density of phenotype B, dark purple patches of high density of phenotype A, and red patches of approximately equal proportions. The black background shows the grid area not covered by phenotype B.

place *pool* accounts for the variable pool size (volume) depending on the annulus to which each sector belongs. The extra place *src_index* and its pretransition record to which annulus a given sector belongs—information that is used to adapt the rate of the transition *division*. A future version of our modelling tool will allow specifying a variable initial marking for a coloured place and accessing the index of a position in the grid without the need for additional places and transitions. Henceforth, this overhead should not be taken into consideration when comparing the spatial representations.

4. ANALYSING PHASE VARIATION

4.1. Computational Experiments

The Petri nets were constructed using Snoopy [Rohr et al. 2010], which recently has been extended to support CPNs [Heiner et al. 2012]. Simulations were run with Snoopy's built-in stochastic simulator and MARCIE [Heiner et al. 2013]. Simulation traces have been further processed by customized C++ programs and finally visualised as images or mp4 movies.

All computational experiments were performed on automatically unfolded Petri nets. Unfolding the CPN for a 101 × 101 grid using a rectangular geometry yields an uncoloured Petri net with 30,605 places and 362,405 transitions with an unfolding time of 780 seconds on a regular desktop computer (Intel® Core™ i5-2500 CPU with a 3.30 GHz processor and 2 GB DDR3 RAM). Similarly, unfolding a CPN of the same dimensions using a circular geometry yields an uncoloured Petri net with 40,406 places and 382,191 transitions with an unfolding time of 2,000 seconds. The number of places and transitions is higher in the circular case due to the overhead required by the current Snoopy version for recording to which annulus each sector belongs.

The unfolded Petri net is simulated using the Gillespie algorithm [Gillespie 1977]. The output of the simulation comprises two traces for each grid position, corresponding to the two phenotypes A and B.

The analysis follows the development over time of the proportion of the given phenotype in the total population and the formation of the associated patterns. This requires converting the traces from the stochastic simulations into 2D representations (Figure 6) and analysing the development of the 2D sector-like patterns over time.

We expect that the model will finally allow the prediction of mutation and fitness rates by counting and extracting information from the pattern segments, which in the future could give new insights into the population dynamics of mutation. The rates could be predicted by fitting the model parameters such that the in silico sector-like patterns match the ones observed in vitro/vivo. Similar sector-like patterns supposedly develop due to similar mutation/fitness rates. Therefore, after successfully completing the fitting procedure, the mutation/fitness rates in the model should be representative

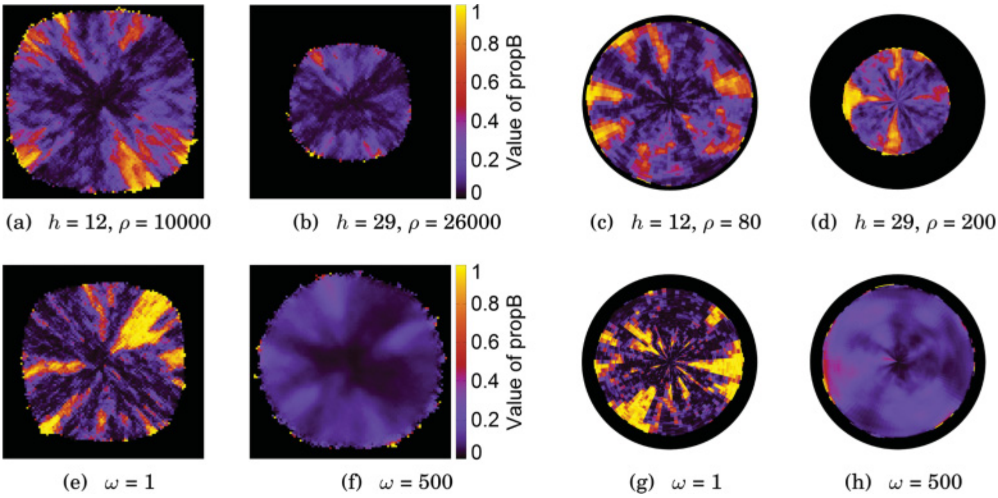


Fig. 7. Different values of the parameter ρ, and implicitly maximum height (h), for the Cartesian coordinate system (a, b) and the polar coordinate system (c, d). Different values of the parameter ω for the Cartesian coordinate system (e, f) and the polar coordinate system (g, h).

for the in vitro/vivo colonies. Currently, the model predicts behaviour that has not been measured so far in the wet lab in the sense that it generates a time series description of the *evolution* of bacterial colony patterns, whereas wet lab data just provide snapshots of final states; it is expected that recording the entire bacterial colony growth in vivo would increase the costs of the experiments without providing additional biologically significant information.

4.2. Parameter Scanning

When the mutation rates are fixed, different combinations of values for parameters ρ and ω will lead to different simulation outcomes. One batch of simulations was run for each parameter ρ and ω by choosing values from the parameter space in order to observe how the behaviour is affected.

4.2.1. Changing ρ. In the first batch of simulations, all parameters were kept constant, except ρ, which had a different value for each run. The values for ρ were selected using a binary search technique considering a geometry-dependent lower bound L and upper bound U. Assuming that avg represents the average value of L and U, if $\rho = avg$ was not providing sufficient room for the development of the colony, then U was set equal to avg and the process was repeated. Conversely, if $\rho = avg$ was providing too much space (i.e. too many cells could pile up on top of each other), then L was set equal to avg and the process was repeated. Otherwise, ρ was set equal to avg and the parameter scanning procedure stopped.

In the rectangular case, the volume or capacity is constant throughout the grid, whereas in the circular case it is not. Therefore, ρ has a different interpretation depending on the chosen spatial representation. For comparison purposes, it is better to consider the *maximum height* of the colony, which is constant throughout the entire grid for both geometries. Experiments with the same heights and corresponding ρ's were carried out for both geometries, and two characteristic results for each one of the geometries are depicted in Figure 7(a) through (d).

Increasing the value of the parameter ρ increases the maximum height of each grid position, which implies that more bacteria can pile up onto each other. Since the number of generations is fixed and the maximum height limit of the colony was increased, it is to be expected that the final width of the colony is reduced; this can be observed in Figure 7(a) through (d). The value of ρ was chosen for both geometries in such a way that the most outward bacteria with respect to the centre do not reach the edge of the grid. The reason for this is that we expected some back propagation of bacteria from the edge of the grid to affect the final outcome of the simulation.

4.2.2. Changing ω. The second batch of simulations changed only the value of ω for each run. In contrast to the selection of values for ρ, the values for ω have been chosen according to the function $f(x) = x^2$, where the initial value of x was 1 ($\omega = 1^2$) and then it was gradually increased, $x_{new} = x_{old} + 1$ ($\omega_{new} = (\sqrt{\omega_{old}} + 1)^2$), until no sectors emerged in the bacterial colony.

Images representing the final states of two simulations for each geometry are given in Figure 7(e) through (h).

The probability of the offspring to stay with the parent or be displaced to a neighbouring position depends on the dimensions of the grid position. All grid positions are equally sized in the model using Cartesian coordinates, which means that the probabilities of staying/being displaced are constant.

However, the area of the grid positions in the model using polar coordinates is different, which means that the probabilities are different as well. The value of ω specified as caption for the polar coordinates model in Figure 7(e) through (h) corresponds to the most outward annular sectors (i.e. annular sectors with the biggest area).

Considering that the value of γ is fixed, the preference of the offspring to be displaced to a neighbouring position is directly proportional to the value of ω. Increasing ω increases the chance of the offspring being displaced, which means that the clear cut between high- and low-density areas in the images fades away. Thus, in Figure 7(e) through 7(h), the images corresponding to a higher value of ω have a more uniform distribution of concentrations than the ones in which ω was smaller.

4.3. Sector Analysis

In the beginning, the analysis of the sectors was done manually by comparing the images of the colony at different time points and deciding if the sector-like patterns are similar to the ones observed in the wet lab.

For the purpose of improving the assessment of results, there was a need to formalise the analysis of sectors. The following set of measures was defined to describe the sector-like patterns: *area*, *angle* described by the sides, *distance* from the centre of the grid, and the *total number* of sectors.

Using specific image processing techniques from the open source computer vision library OpenCV [Bradski and Kaehler 2008], a sector detection and analysis module was implemented. This module takes an image of the colony as input, detects the sector-like patterns in it, and computes the values of the measures of interest. The main steps of the algorithm are given in Algorithm 1. A step-by-step example illustration of the operations performed on each input image is given in Figure 8.

The sector detection and analysis module was implemented in C++. Most image processing functions described in Algorithm 1 are parameterized and can easily be adapted to detect other patterns as well. To avoid recompilation, the parameters' values are loaded at runtime from an XML configuration file. The contents of this XML file can either be changed by hand or via the graphical user interface of the module, which displays in real time how the detected regions change when altering the values of the parameters. The OpenCV function names, descriptions, and default values of the input parameters used for detecting and analysing sectors are given in Online Appendix E.1; see Itseez [2013] for a detailed description of the functions.

ALGORITHM 1: Algorithm for sector analysis

```
Require: $imageFile$ is valid
Ensure: Results written to $outputFile$

Read $imageFile$ and create an instance of type Mat from it called $image$

Convert $image$ to grayscale
Change the contrast and brightness of $image$ for highlighting regions of interest
Apply morphological close operation on $image$ to connect close regions that
  form a sector
Remove noise from $image$ using blur filters
Threshold $image$ and store the result in $binaryImage$
Detect the contours of sectors in $binaryImage$, approximate their polygons and
  $convexHulls$

for all $hull \in convexHulls$ do
    Compute $distance$ from the centre, $area$ and $angle$ using $hull$
    if $area < THRESHOLD$ then
        Consider the sector identified by $hull$ as noise
    end if
end for

Output the measures computed for each sector to $outputFile$
```

A Linux version of the module (x64) and simple usage examples are made available both as supplementary materials and via the Brunel Web site (http://www.brunel.ac.uk/research/centres/cssb/software-systems-and-databases/multiscale). For the non-Linux users, a virtual machine image containing the preconfigured module was created. It can be opened using the freely available, cross-platform virtualization software VirtualBox [Watson 2008].

The advantage of the algorithm working directly with images and not with the raw output of the simulation is that the images can originate from either a dry or wet lab. Thus, our analysis approach is generic. Since wet lab experimental data was not available, the image processing procedure was validated only on in silico–generated images, but our expectation is that the approach should work similarly well on images from the wet lab.

The module can be employed for a single image or repeatedly for a collection of images. Assuming the latter case, if each image corresponds to a different simulation, and within each simulation to the same time point, the extracted measures can be used to compare and contrast two or more simulations. Conversely, in case the collection of images corresponds to the same simulation, it can be used to analyse how the measures change over time. In brief, the potential basic applications of this module are as follows:

—Analyse measures at a particular time point across multiple simulations
—Analyse how measures change over time within one simulation.

By combining these two basic applications, it is possible to analyse how the measures change over time across multiple simulations.

4.4. Results

For both the rectangular and circular model, 1,000 stochastic simulations were run with an average simulation time of 50 minutes.

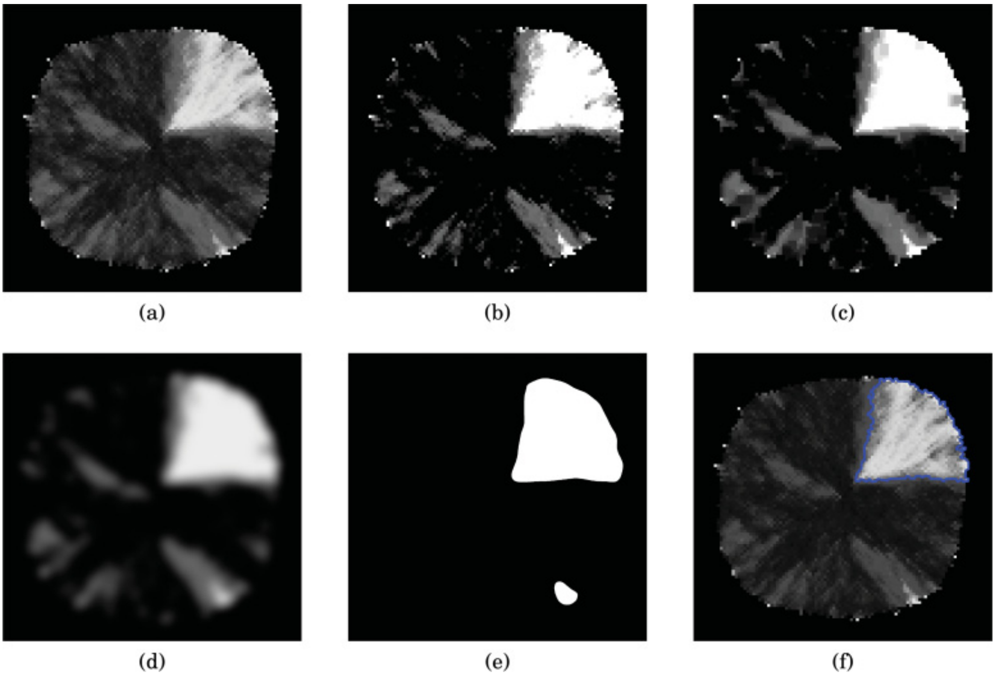


Fig. 8. Main steps for detecting sectors in input images. Each subfigure (SF) corresponds to a specific line (L) from Algorithm 1. Firstly, the input image is converted to grayscale (SF a, L 3). The brightness and contrast of the image are then changed such that the regions of interest are emphasized (SF b, L 4). Next, morphological close operations (SF c, L 5) and Gaussian blur (SF d, L 6) are used to remove the noise. Finally, the image is converted to binary (SF e, L 7) and only the sectors with an area greater than *THRESHOLD* (see Algorithm 1) are considered (SF f, L 12).

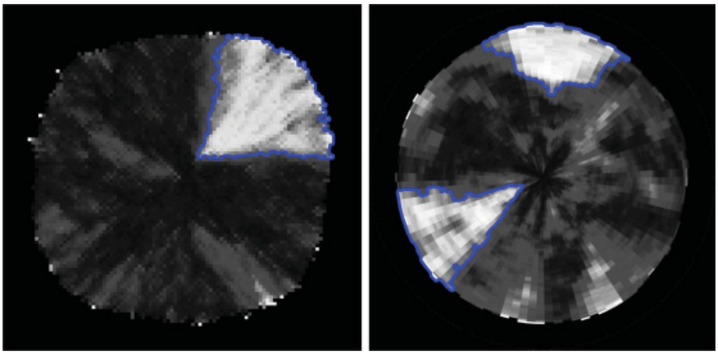


Fig. 9. Images illustrating the detection of sector-like patterns in final state images using rectangular (left) and circular (right) geometries. The border of the detected sectors is highlighted in blue. The area of the sector has to be larger than a threshold value for the sector to be considered.

4.4.1. Final State Analysis. Images were generated from the final states of the simulations, which were then provided as input to the sector analysis module. An example of the result of the sector detection procedure for each geometry is depicted in Figure 9.

The output of the analysis procedure are CSV files containing information about the area, angle, distance from the centre, and number of detected sectors. The averaged results from all simulations for both the rectangular and circular case are described in Table I. We employed a two-sample statistical test for comparing the results. The

Table I. Rectangular (□) and Circular (○) Sector Analysis with μ – Mean, σ – Standard Deviation, c_v – Coefficient of Variation

	Area		Distance		Angle		Sectors	
Measures	□	○	□	○	□	○	□	○
μ	3%	5%	41%	39%	56°	78°	1.47	1.78
σ	2%	2%	17%	16%	18°	25°	1.14	1.03
c_v	0.93	0.62	0.40	0.41	0.32	0.32	0.77	0.58

Note: Area and distance (from the centre) are given with respect to total grid area and maximum distance from the centre.

Table II. The *p*-Value Obtained from the Statistical Tests in Rectangular (□) and Circular (○) Case for All Measures (Area, Distance from the Centre, Angle, Number of Sectors)

	Shapiro-Wilk		
Measures	□	○	Mann-Whitney
Area	<2.2e-16	<2.2e-16	<2.2e-16
Distance	1.334e-07	5.398e-12	5.92e-05
Angle	1.272e-14	<2.2e-16	<2.2e-16
Sectors	<2.2e-16	<2.2e-16	6.68e-13

Note: The Shapiro-Wilk test was used to check the hypothesis that the sample data was drawn from a normal distribution. The Mann-Whitney test was used to check the hypothesis that the sample data for both geometries was drawn from the same distribution.

data corresponding to all measures and both geometries was tested for normality using the Shapiro-Wilk [Shapiro and Wilk 1965] and the Q-Q plot [Wilk and Gnanadesikan 1968] methods. In all cases, the null hypothesis (i.e. that the sample data is drawn from a normal distribution) was rejected. Thus, we tested if the sample data for both geometries is drawn from the same distribution using the Mann-Whitney [Mann and Whitney 1947; Wilcoxon 1945] nonparametric test. Similarly, the null hypothesis (i.e. that the sample data are drawn from the same distribution) was rejected. The *p*-values obtained for all tests are given in Table II.

Both area and angle have higher values in the circular case than in the rectangular case, which is to be expected due to the different 2D space discretisation. Sectors in the circular geometry inherently have a nonzero degree angle associated, whereas rows in a rectangular geometry do not. Moreover, the area of the annular sectors is increasing as they are farther away from the centre of the grid. Conversely, the area of all positions in the rectangular geometry is constant. The number of sectors is slightly bigger in the circular case because the bacteria from the starting position can be displaced in maximum "number of sectors" directions, whereas in the rectangular case only in maximum eight.

Finally, the distance of the sectors from the grid centre is approximately equal for both geometries. Thus, according to these results, the distance from the centre is the only reliable measure that has similar values for both geometries. Running batches of more simulations will increase the accuracy of the results, and more fine-grained conclusions can be drawn. Histograms and corresponding normal distribution curves for all measures have been plotted and added to Online Appendix E.2 to complement the analytical comparison of the results described earlier.

4.4.2. Time Series Analysis. The number of considered time points per model simulation *#timepoints* was chosen such that

(1) no significant changes occur between any two consecutive time points (relative to the geometric properties of the sector-like patterns) and
(2) the value of *#timepoints* is minimal.

After running multiple model simulations, the value of *#timepoints* was set to 101. For comparability purposes, the same *#timepoints* value was employed across all forthcoming simulations. If either the start, end, or number of considered time points would differ, the simulation output would no longer be directly comparable since each time point could potentially correspond to a different stage of the bacterial colony development.

Out of the 1,000 simulations for each geometry, only 500 had the same simulation start time, end time, and number of time points, and were considered for the time series analysis. The remaining simulations were comparable with respect to final state images only.

The configuration for each considered simulation was as follows:

—Simulation start time: 0
—Simulation end time: 100
—Number of considered simulation time points (*#timepoints*): 101.

Each simulation is described by 101 time points, and each time point contains zero or more sectors. A time point is described by the number of sectors and the measures specific to these sectors (angle, area, distance from the centre). Since the number of sectors might differ at the same time point from one simulation to another, the output is not directly comparable. Therefore, instead of considering the measures of all sectors independently, only the average measures' values are taken into account. In this new setting, each time point is characterized by the number of sectors and the average values of the area, angle, and distance from the centre.

A set of 500 time series was analysed for both the rectangular and circular geometry with an average analysis time of 5 seconds/time series. The variability of the measures between different simulations over time is depicted in Figure 10.

The first conclusion drawn from Figure 10 is that the values of neither measure changes after time point 50. To test this conclusion, the standard deviation of the measures was computed for each simulation. The first time point after which the standard deviation of a measure is equal to zero for both rectangular and circular geometries is given in Table III. The standard deviation of all measures irrespective of the employed geometry is equal to zero starting from time point 47 onwards. Therefore, the simulation end time could be reduced to more than half in the future from 100 to 47. Moreover, a steady state termination criteria could be designed for stopping model simulations early whenever the sectors' geometric properties remain constant for a sufficiently long period of time. This could be achieved by integrating the sector detection and analysis module with existing model simulation workflows.

Secondly, all measures increase exponentially on average over time. The area is expected to increase since sectors continue developing outward after they are initially formed. Both the angle and distance from the centre maintain approximately constant values for each sector after the sector is formed. However, they increase on average since the number of sectors increases over time, which means that new sectors are formed. This also means that sectors are formed at varying distances from the centre of the grid. Unfortunately, for our case, this implies that simulations have to be run until the simulation end time is reached to be certain that all possible sectors were captured.

Conversely, if all sectors would emerge early and none of them would disappear, the simulation time could be reduced even further. The analysis could be executed on the fly and the simulation could be stopped after all sectors have started forming. In this scenario, the number of sectors and the values of the angle and distance from the centre would be constant throughout the entire simulation. The only potential issue is the area of the sectors. To compute it, the angle of the sectors and the width of the colony should be known. Assuming that there is prior knowledge regarding the final width of the colony, the computation of the area is trivial. Alternatively, the width of

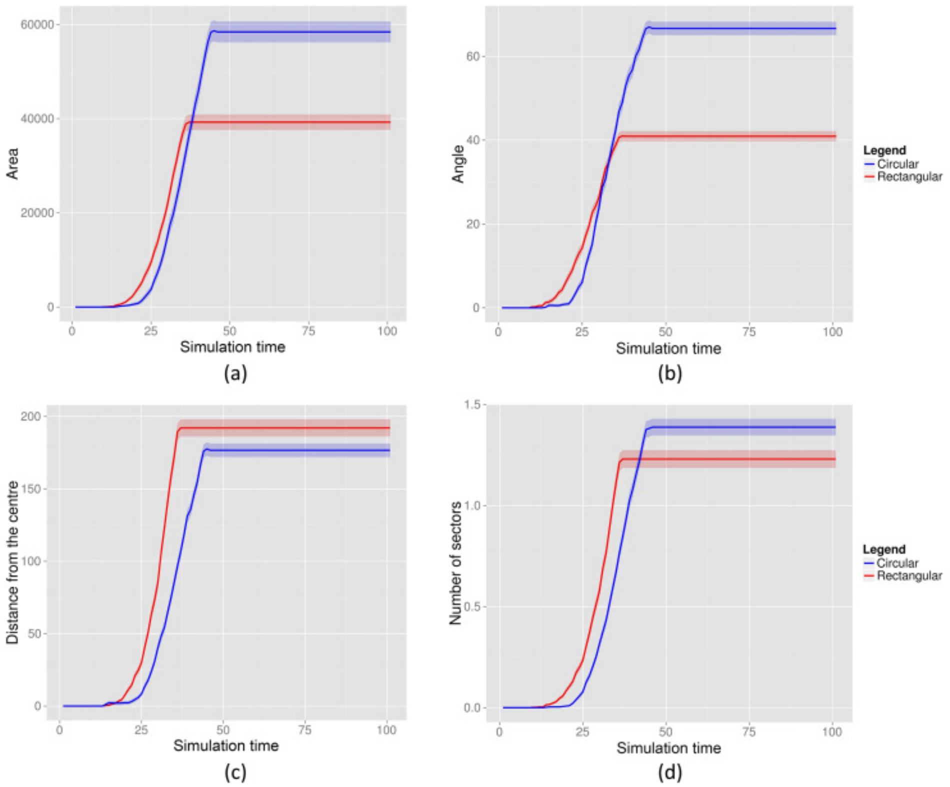


Fig. 10. Variability of measures over simulation time considering all simulations. The solid line represents the mean of the measure. The shaded area is bounded below by (mean – standard error) and above by (mean + standard error). Measures values were recorded for each simulation time point t, $t = \overline{0, 100}$.

Table III. The First Time Point After Which the Standard Deviation of the Measures Is Equal to Zero for Both Rectangular (□) and Circular (○) Geometries

	Area	Angle	Distance from Centre	Number of Sectors
○	47	47	46	46
□	39	39	38	37

the colony can be calculated from a training set of simulations and used afterwards for the simulations that follow. Although it is not applicable to the phase variation case study, this property might be useful when simulating and/or analysing other models.

5. ON THE CONNECTION TO PARTIAL DIFFERENTIAL EQUATIONS

The CPN spatial modelling methodology, as it was presented here, is currently limited to specific spatial representations (2.5D Cartesian and polar) considering flat surfaces.

In contrast, the PDE literature contains more generic techniques and algorithms that could be potentially employed to represent space considering different numbers of dimensions, systems of coordinates, and (non)flat surfaces. To benefit from such approaches in the CPN spatial modelling methodology we will begin a study that has a long-term aim to establish a connection between Petri nets and discrete approximations to PDEs. The aim for the moment is deliberately modest in that we will only consider how a simplified phase variation model can be translated between its continuous (PDE)

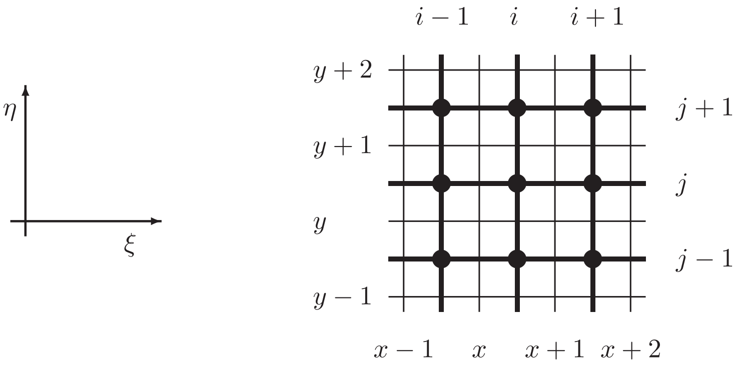


Fig. 11. The underlying (ξ, η) Cartesian coordinate system (left) and the (i, j) finite difference grid offset from the (x, y) "place" grid (right).

and discrete representation. This is done to not to cloud the argument at this stage with cumbersome and unnecessary mathematical notation and detail.

5.1. Notations

Our starting point is the left schematic in Figure 4, where the species is constrained to be displaced in 2D space only according to a predefined indexed rectangular geometry. However, to match with the PDE literature, it is convenient to index grid lines rather than the places between the lines. Thus, for each place defined by the diagonal joining (x, y) to $(x+1, y+1)$, we will identify the midpoint of that place as the point (i, j). This defines an offset grid according to Figure 11 (right) and permits us to identify places with grid intersections. Further, since (x, y) has already been used to denote discrete positions, we use (ξ, η) to refer to the underlying continuous Cartesian coordinate system of 2D space.

5.2. Modelling Lateral Displacement

Now consider the very simplest case where a species, originally at a position (i, j) in this regular Cartesian lattice, can at the next timestep diffuse one spatial step either left, right, up, or down. The concentration of the species at each grid location is denoted by $c_{i,j}$, and in the limit of continuous time, with the assumption that the diffusion rates follow mass action kinetics with the common rate parameter k, we get the following system of ordinary differential equations (ODEs):

$$\frac{1}{k}\frac{dc_{i,j}}{dt} = c_{i+1,j} + c_{i-1,j} + c_{i,j+1} + c_{i,j-1} - 4c_{i,j}. \tag{6}$$

Denoting the grid spacing as $h = (x+1) - x$ (constant, and the same in each direction) and equipped with a standard reference to finite difference approximations of PDEs such as Smith [1965, Chapter 5], we can then recognise this system of ODEs as a *method of lines* algorithm. Indeed, identifying $k = \sigma/h^2$, where σ is a physical rate parameter independent of the grid, these equations are a *second-order finite difference approximation* to the canonical parabolic PDE problem:

$$\frac{\partial c}{\partial t} = \sigma \nabla^2 c, \tag{7}$$

with the Laplacian defined in the standard Cartesian way as $\nabla^2 c = \partial^2 c/\partial\xi^2 + \partial^2 c/\partial\eta^2$. This PDE is to be solved for times $t > 0$, within the domain covered by the grid, subject to known boundary conditions on the grid boundary, and a specified *initial* distribution of c at $t = 0$.

To verify this connection between (6) and (7), let us adopt the notation $c_{ij} = c(\xi_i, \eta_j)$ and then use a Taylor series expansion in the coordinate directions about (ξ_i, η_j) to get

$$c_{i\pm1,j} = \sum_{n=0}^{3} \frac{(\pm h)^n}{n!} \frac{\partial^n c_{ij}}{\partial \xi^n} + O(h^4) \text{ and } c_{i,j\pm1} = \sum_{n=0}^{3} \frac{(\pm h)^n}{n!} \frac{\partial^n c_{ij}}{\partial \eta^n} + O(h^4),$$

where, as usual for the "big Oh" notation, $O(A) \le CA$ for some constant C. Adding these four equations together and noting the cancellations results in

$$\nabla^2 c = \frac{c_{i+1,j} + c_{i-1,j} + c_{i,j+1} + c_{i,j-1} - 4c_{i,j}}{h^2} + O(h^2),$$

and dropping the $O(h^2)$ error terms then reveals (6) to be a second-order approximation of (7). The "second order" terminology arises because the approximation has a leading order error term of size $O(h^2)$ (i.e. the "second" power).

5.3. Modelling Diagonal Displacement

Such a model is restrictive in that it allows species migration only along the coordinate directions. We can easily imagine a model that allows for only diagonal migration, and then we arrive at

$$\frac{1}{k}\frac{dc_{i,j}}{dt} = c_{i+1,j-1} + c_{i-1,j-1} + c_{i+1,j+1} + c_{i-1,j+1} - 4c_{i,j}. \tag{8}$$

Both (6) and (8) can be conveniently represented using the inverted square brackets of *finite difference stencil notation*. In this shorthand notation, they become

$$\frac{1}{k}\frac{dc_{i,j}}{dt} = \left]\begin{array}{ccc} & 1 & \\ 1 & -4 & 1 \\ & 1 & \end{array}\right[c_{i,j} \qquad \text{and} \qquad \frac{1}{k}\frac{dc_{i,j}}{dt} = \left]\begin{array}{ccc} 1 & & 1 \\ & -4 & \\ 1 & & 1 \end{array}\right[c_{i,j}. \tag{9}$$

The stencil notation is designed to be intuitive. For example, the one on the right of (9) tells us that we are to operate on c, centered at (i, j) using diagonally adjacent neighbours and the coefficients placed in the stencil at those grid points. The aim, of course, is to recover (8). In the same way, the left of (9) corresponds to a compact way of writing (6).

Now, a Taylor series expansion along the diagonals that cross at (ξ_i, η_j) will reveal that (8) is also a second-order approximation. In other words, we need to neglect an error term of magnitude $O(h^2)$ to arrive at (8) from (7).

5.4. Modelling Diagonal and Lateral Displacement

The model can be further improved by allowing diffusion in all eight directions. In fact, although there is not space here to give the details, it is a routine calculation to show that by adding four of (6) to one of (8), the $O(h^2)$ error terms present in each *cancel out*, leaving an error term of order $O(h^4)$. This combination results in the nine-point stencil

$$\frac{5}{k}\frac{dc_{i,j}}{dt} = \left]\begin{array}{rrr} 1 & 4 & 1 \\ 4 & -20 & 4 \\ 1 & 4 & 1 \end{array}\right[c_{i,j}, \tag{10}$$

and, for dense grids where $h \ll 1$, it is clear that $h^4 \ll h^2$, so this nine-point stencil can be considered a much more accurate approximation of the diffusion process.

The details of this argument can be found in the classic text by Smith [1965], and from there, for example, we can see that the generalisation to three spatial dimensions, as well as polar and spherical polar coordinate systems, is straightforward. Furthermore, analogous formulations on nonflat surfaces may also be possible, but as that is an evolving area in PDE research, we mention it here only as an interesting possibility for future work.

In contrast to the deterministic PDE model described here, the CPN phase variation model is stochastic. To add stochasticity to a deterministic PDE model, established methods (e.g. Gunzburger et al. [2014]) for dealing with stochastic partial differential equations (SPDEs) can be employed where the stochasticity is introduced through uncertainty in the coefficients, or through the addition of a random driving noise.

5.4. Modelling Cell Division and Mutation

Finally, we speculate on how we might incorporate spatial variation into the phase switching and cell division ideas described earlier around Figure 3. For this, we consider the coupled PDE system,

$$\frac{\partial A}{\partial t} = (1-\alpha)d_A\nabla^2 A + d_B\beta\nabla^2 B + f_A(A, B), \tag{11}$$

$$\frac{\partial B}{\partial t} = (1-\beta)d_B\nabla^2 B + d_A\alpha\nabla^2 A + f_B(A, B), \tag{12}$$

where f_A and f_B are (possibly nonlinear) source terms that can be designed, through parameter identification and calibration for example, to allow for production via cell division. The approximation of the Laplacian operator ∇^2 can now be carried out in exactly the same way as described earlier. For example, if we choose to use the four-point stencil in (6), we will obtain

$$\frac{dA_{i,j}}{dt} = \left] \begin{matrix} & 1 & \\ 1 & -4 & 1 \\ & 1 & \end{matrix} \right[\left(\frac{(1-\alpha)d_A}{h^2} A_{i,j} + \frac{d_B\beta}{h^2} B_{i,j} \right) + f_A(A_{i,j}, B_{i,j}), \tag{13}$$

$$\frac{dB_{i,j}}{dt} = \left] \begin{matrix} & 1 & \\ 1 & -4 & 1 \\ & 1 & \end{matrix} \right[\left(\frac{d_A\alpha}{h^2} A_{i,j} + \frac{(1-\beta)d_B}{h^2} B_{i,j} \right) + f_B(A_{i,j}, B_{i,j}). \tag{14}$$

We could also use the four-point stencil in (8) in the obvious way and even the nine-point stencil as earlier. Indeed, because the mutation and fitness rates (α_A, α_B and d_A, d_B) are constants, the error terms are unaffected and the nine-point stencil remains a higher-order approximation.

Although these observations have been kept at the level of overview, they suggest a potentially fruitful line of research that can connect discretised approximations of PDEs to CPNs and systems biology. The plethora of computational algorithms from PDE research may then be profitably applied to the study of, for example, bacterial colony growth.

6. SUMMARY

In this article, we have described a methodology of modelling bacterial colonies that evolve in time and space using rectangular and circular geometries, as well as a procedure for sector-like patterns detection and analysis.

Currently, it is not possible to state which geometry is more appropriate for the phase variation case study, because there are not sufficient images from the wet lab against which to validate our results. The emphasis of this work is on the generic methodologies that we developed and that can be employed for different case studies modelled using CPNs. When images of actual bacterial colonies are available, they could be used as targets for model fitting to generate more accurate computational models for describing bacterial colony growth under different conditions.

In the future, we plan to extend our spatial modelling framework from 2.5D (i.e. 2D and implicitly modelling height) to full 3D representation, which would allow the simulation and observation of more detailed aspects of bacterial colonies. Moreover, we would like to further investigate the connection between Petri net models and PDEs to develop a unifying framework for efficient simulation of multidimensional (potentially multiscale) models. To improve the scalability of the approach, CPN models will be simulated directly at the coloured level using improved parallelisation support. We also want to extend our sector detection and analysis procedure from working with 2D sector-like patterns to linear and nonlinear 3D surfaces.

All supplementary materials are made available online and via the Brunel Web site (http://www.brunel.ac.uk/research/centres/cssb/software-systems-and-databases/multiscale).

APPENDIX

A. CONTRIBUTION OF AUTHORS

The first author, Ovidiu Pârvu, is the lead author of this submission. He is currently registered as a full-time PhD student at Brunel University, College of Engineering, Design and Physical Sciences, Department of Computer Science.

The contributions of the individual authors stated both in words and percentages of effort can be split as follows:

—Ovidiu Pârvu (60% of effort) developed the methodology for modelling using $\mathcal{SPN}^{\mathcal{C}}$ and polar coordinates, implemented the visualisation and the sector detection and analysis modules, ran the parameter scan procedures and created most of the figures, executed the batches of 1,000 simulations on the cluster, analysed the simulation output, and performed the statistical analysis of the results. Moreover, he wrote Sections 3 and 4 of the article.

—David Gilbert (15% of effort) suggested the development of the grid with polar coordinates and supervised the implementation of the analysis techniques. He also wrote core parts of Sections 1 and 6.

—Monika Heiner (10% of effort) built the original $\mathcal{SPN}^{\mathcal{C}}$ model using Cartesian coordinates and wrote Section 2, excluding the Background subsection.

—Fei Liu (5% of effort) developed the modelling language and supported the model development.

—Nigel Saunders (5% of effort) provided the biological case study and is one of the authors of the publication [Saunders et al. 2003] on which our article builds. He wrote the Background subsection in Section 2.

—Simon Shaw (5% of effort) provided the mathematical formalisation of the connection between Petri nets and PDEs. He wrote Section 5 of the article.

All authors read and approved the entire article.

ELECTRONICS APPENDIX

The electronic appendix for this article can be accessed in the ACM Digital Library.

ACKNOWLEDGMENTS

The authors gratefully acknowledge the reviewers' insightful comments, which helped improve the manuscript.

REFERENCES

Mark S. Alber, Maria A. Kiskowski, James A. Glazier, and Yi Jiang. 2003. On cellular automaton approaches to modeling biological cells. In *Mathematical Systems Theory in Biology, Communications, Computation, and Finance.* IMA Volumes in Mathematics and Its Applications, No. 134. Springer, 1–39. http://link.springer.com/chapter/10.1007/978-0-387-21696-6_1.

Gary An, Qi Mi, Joyeeta Dutta-Moscato, and Yoram Vodovotz. 2009. Agent-based models in translational systems biology. *Wiley Interdisciplinary Reviews: Systems Biology and Medicine* 1, 2, 159–171. DOI:http://dx.doi.org/10.1002/wsbm.45

Roger Arditi, Yuri Tyutyunov, Andrey Morgulis, Vasiliy Govorukhin, and Inna Senina. 2001. Directed movement of predators and the emergence of density-dependence in predator–prey models. *Theoretical Population Biology* 59, 3, 207–221. DOI:http://dx.doi.org/10.1006/tpbi.2001.1513

Ariel Balter, Roeland M. H. Merks, Nikodem J. Popławski, Maciej Swat, and James A. Glazier. 2007. The Glazier-Graner-Hogeweg model: Extensions, future directions, and opportunities for further study. In *Single-Cell-Based Models in Biology and Medicine*, A. R. A. Anderson, M. A. J. Chaplain, and K. A. Rejniak (Eds.). Birkhäuser Basel, 151–167. http://link.springer.com/chapter/10.1007/978-3-7643-8123-3_7.

Roberto Barbuti, Andrea Maggiolo-Schettini, Paolo Milazzo, Giovanni Pardini, and Luca Tesei. 2011. Spatial P systems. *Natural Computing* 10, 1, 3–16. DOI:http://dx.doi.org/10.1007/s11047-010-9187-z

Daniela Besozzi, Paolo Cazzaniga, Dario Pescini, and Giancarlo Mauri. 2008. Modelling metapopulations with stochastic membrane systems. *Biosystems* 91, 3, 499–514. DOI:http://dx.doi.org/10.1016/j.biosystems.2006.12.011

Michael L. Blinov, James R. Faeder, Byron Goldstein, and William S. Hlavacek. 2004. BioNetGen: Software for rule-based modeling of signal transduction based on the interactions of molecular domains. *Bioinformatics* 20, 17, 3289–3291. DOI:http://dx.doi.org/10.1093/bioinformatics/bth378

Gary Bradski and Adrian Kaehler. 2008. *Learning OpenCV: Computer Vision with the OpenCV Library*. O'Reilly, Cambridge, MA.

Vincent Danos, Jérôme Feret, Walter Fontana, Russell Harmer, and Jean Krivine. 2007. Rule-based modelling of cellular signalling. In *CONCUR 2007—Concurrency Theory*. Lecture Notes in Computer Science, Vol. 4703. Springer, 17–41. http://link.springer.com/chapter/10.1007/978-3-540-74407-8_3.

Nicolette de Vries, Dirk Duinsbergen, Ernst J. Kuipers, Raymond G. J. Pot, Patricia Wiesenekker, Charles W. Penn, Arnoud H. M. van Vliet, Christina M. J. E. Vandenbroucke-Grauls, and Johannes G. Kusters. 2002. Transcriptional phase variation of a type III restriction-modification system in Helicobacter pylori. *Journal of Bacteriology* 184, 23, 6615–6623.

Andreas Deutsch and Sabine Dormann. 2007. *Cellular Automaton Modeling of Biological Pattern Formation: Characterization, Applications, and Analysis*. Springer Science and Business Media.

G. Bard Ermentrout and Leah Edelstein-Keshet. 1993. Cellular automata approaches to biological modeling. *Journal of Theoretical Biology* 160, 1, 97–133. DOI:http://dx.doi.org/10.1006/jtbi.1993.1007

Cheng Feng and Jane Hillston. 2014. PALOMA: A process algebra for located Markovian agents. In *Quantitative Evaluation of Systems*, G. Norman and W. Sanders (Eds.). Springer International Publishing, 265–280. http://link.springer.com/chapter/10.1007/978-3-319-10696-0_22.

Jordi Ferrer, Clara Prats, and Daniel López. 2008. Individual-based modelling: An essential tool for microbiology. *Journal of Biological Physics* 34, 1–2, 19–37. DOI:http://dx.doi.org/10.1007/s10867-008-9082-3

Qian Gao, David Gilbert, Monika Heiner, Fei Liu, Daniele Maccagnola, and David Tree. 2013. Multiscale modeling and analysis of planar cell polarity in the Drosophila wing. *IEEE/ACM Transactions on Computational Biology and Bioinformatics* 10, 2, 337–351. DOI:http://dx.doi.org/10.1109/TCBB.2012.101

David Gilbert, Monika Heiner, Fei Liu, and Nigel Saunders. 2013. Colouring space—a coloured framework for spatial modelling in systems biology. In *Application and Theory of Petri Nets and Concurrency.* Lecture Notes in Computer Science, Vol. 7927. Springer, 230–249. http://link.springer.com/chapter/10.1007/978-3-642-38697-8_13.

Daniel T. Gillespie. 1977. Exact stochastic simulation of coupled chemical reactions. *Journal of Physical Chemistry* 81, 25, 2340–2361.

Franois Graner and James A. Glazier. 1992. Simulation of biological cell sorting using a two-dimensional extended Potts model. *Physical Review Letters* 69, 13, 2013–2016. DOI:http://dx.doi.org/10.1103/PhysRevLett.69.2013

Max D. Gunzburger, Clayton G. Webster, and Guannan Zhang. 2014. Stochastic finite element methods for partial differential equations with random input data. *Acta Numerica* 23, 521–650. DOI:http://dx.doi.org/10.1017/S0962492914000075

Monika Heiner and David Gilbert. 2011. How might Petri nets enhance your systems biology toolkit. In *Applications and Theory of Petri Nets*. Lecture Notes in Computer Science, Vol. 6709, Springer, 17–37.

Monika Heiner, Mostafa Herajy, Fei Liu, Christian Rohr, and Martin Schwarick. 2012. Snoopy—a unifying Petri net tool. In *Application and Theory of Petri Nets.* Lecture Notes in Computer Science, Vol. 7347. Springer, 398–407.

Monika Heiner, Christian Rohr, and Martin Schwarick. 2013. MARCIE-model checking and analysis done efficiently. In *Application and Theory of Petri Nets and Concurrency*. Lecture Notes in Computer Science, Vol. 7927. Springer, 389–399. DOI:http://dx.doi.org/10.1007/978-3-642-38697-8_21

Thomas Hillen and Kevin J. Painter. 2009. A user's guide to PDE models for chemotaxis. *Journal of Mathematical Biology* 58, 1–2, 183–217. DOI:http://dx.doi.org/10.1007/s00285-008-0201-3

Derek W. Hood, Mary E. Deadman, Michael P. Jennings, Marina Bisercic, Robert D. Fleischmann, J. Craig Venter, and E. Richard Moxon. 1996. DNA repeats identify novel virulence genes in Haemophilus influenzae. *Proceedings of the National Academy of Sciences of the United States of America* 93, 20, 11121–11125.

Trey Ideker, Timothy Galitski, and Leroy Hood. 2001. A new approach to decoding life: systems biology. *Annual Review of Genomics and Human Genetics* 2, 1, 343–372. DOI:http://dx.doi.org/10.1146/annurev.genom.2.1.343

Itseez. 2013. OpenCV documentation. Retrieved March 31, 2015, from http://docs.opencv.org.

Mathias John, Cédric Lhoussaine, Joachim Niehren, and Adelinde M. Uhrmacher. 2010. The attributed Pi-calculus with priorities. In *Transactions on Computational Systems Biology XII*, C. Priami, R. Breitling, D. Gilbert, M. Heiner, and A. M. Uhrmacher (Eds.). Springer, 13–76. http://link.springer.com/chapter/10.1007/978-3-642-11712-1_2.

Mathias John, Cédric Lhoussaine, Joachim Niehren, and Cristian Versari. 2011. Biochemical reaction rules with constraints. In *Programming Languages and Systems*. Lecture Notes in Computer Science, Vol. 6602. Springer, 338–357. http://link.springer.com/chapter/10.1007/978-3-642-19718-5_18.

Philip W. Jordan and Nigel J. Saunders. 2009. Host iron binding proteins acting as niche indicators for Neisseria meningitidis. *PloS One* 4, 4, e5198. DOI:http://dx.doi.org/10.1371/journal.pone.0005198

Hiroaki Kitano. 2002. Computational systems biology. *Nature* 420, 6912, 206–210. DOI:http://dx.doi.org/10.1038/nature01254

Shigeru Kondo and Takashi Miura. 2010. Reaction-diffusion model as a framework for understanding biological pattern formation. *Science* 329, 5999, 1616–1620. DOI:http://dx.doi.org/10.1126/science.1179047

Hillel Kugler, Antti Larjo, and David Harel. 2010. Biocharts: A visual formalism for complex biological systems. *Journal of the Royal Society Interface* 7, 48, 1015–1024. DOI:http://dx.doi.org/10.1098/rsif.2009.0457

Fei Liu, Mary Ann Blätke, Monika Heiner, and Ming Yang. 2014. Modelling and simulating reaction–diffusion systems using coloured Petri nets. *Computers in Biology and Medicine* 53, 297–308. DOI:http://dx.doi.org/10.1016/j.compbiomed.2014.07.004

Charles M. Macal and Michael J. North. 2010. Tutorial on agent-based modelling and simulation. *Journal of Simulation* 4, 3, 151–162. DOI:http://dx.doi.org/10.1057/jos.2010.3

Henry B. Mann and Donald R. Whitney. 1947. On a test of whether one of two random variables is stochastically larger than the other. *Annals of Mathematical Statistics* 18, 1, 50–60. DOI:http://dx.doi.org/10.2307/2236101

Carsten Maus, Stefan Rybacki, and Adelinde M. Uhrmacher. 2011. Rule-based multi-level modeling of cell biological systems. *BMC Systems Biology* 5, 1, 166. DOI:http://dx.doi.org/10.1186/1752-0509-5-166

Masayasu Mimura, Hideo Sakaguchi, and Mitsugu Matsushita. 2000. Reaction–diffusion modelling of bacterial colony patterns. *Physica A: Statistical Mechanics and Its Applications* 282, 1–2, 283–303. DOI:http://dx.doi.org/10.1016/S0378-4371(00)00085-6

E. Richard Moxon, Paul B. Rainey, Martin A. Nowak, and Richard E. Lenski. 1994. Adaptive evolution of highly mutable loci in pathogenic bacteria. *Current Biology: CB* 4, 1, 24–33.

Durica Nikolić, Corrado Priami, and Roberto Zunino. 2012. A rule-based and imperative language for biochemical modeling and simulation. In *Software Engineering and Formal Methods.* Lecture Notes in Computer Science, Vol. 7504. Springer, 16–32. http://link.springer.com/chapter/10.1007/978-3-642-33826-7_2.

Ovidiu Pârvu, David Gilbert, Monika Heiner, Fei Liu, and Nigel Saunders. 2013. Modelling and analysis of phase variation in bacterial colony growth. In *Computational Methods in Systems Biology.*

Lecture Notes in Computer Science, Vol. 8130. Springer, 78–91. http://link.springer.com/chapter/10.1007/978-3-642-40708-6_7.

Nikodem J. Poplawski, Abbas Shirinifard, Maciej Swat, and James A. Glazier. 2008. Simulation of single-species bacterial-biofilm growth using the Glazier-Graner-Hogeweg model and the CompuCell3D modeling environment. *Mathematical Biosciences and Engineering: MBE* 5, 2, 355–388. http://www.ncbi.nlm.nih.gov/pmc/articles/PMC2547990/.

Christian Rohr, Wolfgang Marwan, and Monika Heiner. 2010. Snoopy—a unifying Petri net framework to investigate biomolecular networks. *Bioinformatics* 26, 7, 974–975.

Laurence Salaün, Sarah Ayraud, and Nigel J. Saunders. 2005. Phase variation mediated niche adaptation during prolonged experimental murine infection with Helicobacter pylori. *Microbiology (Reading, England)* 151, 3, 917–923. DOI:http://dx.doi.org/10.1099/mic.0.27379-0

Laurence Salaün, Bodo Linz, Sebastian Suerbaum, and Nigel J. Saunders. 2004. The diversity within an expanded and redefined repertoire of phase-variable genes in Helicobacter pylori. *Microbiology (Reading, England)* 150, 4, 817–830.

Laurence Salaün, Lori A. Snyder, and Nigel J. Saunders. 2003. Adaptation by phase variation in pathogenic bacteria. *Advances in Applied Microbiology* 52, 263–301.

Nigel J. Saunders, E. Richard Moxon, and Mike B. Gravenor. 2003. Mutation rates: Estimating phase variation rates when fitness differences are present and their impact on population structure. *Microbiology* 149, 2, 485–495.

Nigel J. Saunders, Alex C. Jeffries, John F. Peden, Derek W. Hood, Herve Tettelin, Rino Rappuoli, and E. Richard Moxon. 2000. Repeat-associated phase variable genes in the complete genome sequence of Neisseria meningitidis strain MC58. *Molecular Microbiology* 37, 1, 207–215.

Nigel J. Saunders, John F. Peden, Derek W. Hood, and E. Richard Moxon. 1998. Simple sequence repeats in the Helicobacter pylori genome. *Molecular Microbiology* 27, 6, 1091–1098.

Jim Schaff, Charles C. Fink, Boris Slepchenko, John H. Carson, and Leslie M. Loew. 1997. A general computational framework for modeling cellular structure and function. *Biophysical Journal* 73, 3, 1135–1146. DOI:http://dx.doi.org/10.1016/S0006-3495(97)78146-3

Samuel S. Shapiro and Martin B. Wilk. 1965. An analysis of variance test for normality (complete samples). *Biometrika* 3, 52.

Gordon D. Smith. 1965. *Numerical Solution of Partial Differential Equations*. Oxford University Press, London, England.

Lori A. Snyder, Sarah A. Butcher, and Nigel J. Saunders. 2001. Comparative whole-genome analyses reveal over 100 putative phase-variable genes in the pathogenic Neisseria spp. *Microbiology (Reading, England)* 147, 8, 2321–2332.

Yogitha N. Srikhanta, Tina L. Maguire, Katryn J. Stacey, Sean M. Grimmond, and Michael P. Jennings. 2005. The phasevarion: A genetic system controlling coordinated, random switching of expression of multiple genes. *Proceedings of the National Academy of Sciences of the United States of America* 102, 15, 5547–5551. DOI:http://dx.doi.org/10.1073/pnas.0501169102

Bryan C. Thorne, Alexander M. Bailey, and Shayn M. Peirce. 2007. Combining experiments with multi-cell agent-based modeling to study biological tissue patterning. *Briefings in Bioinformatics* 8, 4, 245–257. DOI:http://dx.doi.org/10.1093/bib/bbm024

Jon Watson. 2008. VirtualBox: Bits and bytes masquerading as machines. *Linux Journal* 2008, 166, Article No. 1. http://dl.acm.org/citation.cfm?id=1344209.1344210.

Eric W. Weisstein. 2013. Wolfram MathWorld Home Page. Retrieved March 31, 2015, from http://mathworld.wolfram.com.

Frank Wilcoxon. 1945. Individual comparisons by ranking methods. *Biometrics Bulletin* 1, 6, 80–83. DOI:http://dx.doi.org/10.2307/3001968

Martin B. Wilk and Ramanathan Gnanadesikan. 1968. Probability plotting methods for the analysis of data. *Biometrika* 55, 1, 1–17.

Received January 2014; revised November 2014; accepted February 2015