Modelling and Analysis of Phase Variation in Bacterial Colony Growth

Ovidiu Pârvu¹, David Gilbert¹, Monika Heiner², Fei Liu³, and Nigel Saunders¹

¹ School of Information Systems, Computing and Mathematics {ovidiu.parvu, david.gilbert, nigel.saunders}@brunel.ac.uk
² Computer Science Institute, Brandenburg University of Technology monika.heiner@informatik.tu-cottbus.de
³ Harbin Institute of Technology liufei@hit.edu.cn

Abstract. We describe an investigation into spatial modelling by means of an ongoing case study, namely phase variation patterning in bacterial colony growth, forming circular colonies on a flat medium. We explore the application of two different geometries, rectangular and circular, for modelling and analysing the colony growth in 2.5 dimensions. Our modelling paradigm is that of coloured stochastic Petri nets and we employ stochastic simulation in order to generate output which is then analysed for sector patterning. The analysis results are used to compare the two geometries, and our multidimensional approach is a precursor to more work on detailed multiscale modelling.

Keywords: Coloured stochastic Petri nets; spatial modelling; Systems Biology; pattern analysis; multidimensional; BioModel Engineering.

1 Motivation

This paper builds on [5], where we have introduced our methodology for the use of a structured family of Petri net classes which enables the investigation of biological systems using complementary modelling abstractions comprising the qualitative and quantitative, i.e., stochastic, continuous, and hybrid paradigms.

We extend our spatial modelling approach introduced in [2,3] where we discretise space within a geometrical framework exploiting finite discrete colour sets embedded in coloured Petri nets. We motivate our work by describing an investigation into spatial modelling by means of an ongoing case study, namely phase variation patterning in bacterial colony growth, forming circular colonies on a flat medium. In order to illustrate the power and flexibility of our approach we explore the application of two different geometries, rectangular and circular, for modelling and analysing the colony growth in 2.5 dimensions: the 2 dimensions of the surface of the colony are modelled explicitly while the height is modelled implicitly. In order to capture the stochastic properties of the case study, we have chosen coloured stochastic Petri nets (SPN^{C}) as our modelling paradigm and employ stochastic simulation in order to generate output which

is then analysed for sector patterning. The analysis results are used to compare the two geometries, and our multidimensional approach is a precursor to more work on detailed multiscale modelling. The main contributions of our paper are

- a detailed model of phase variation in bacterial colony growth, in two geometries (rectangular and circular),
- the development and application of techniques to analyse the properties of the patterns generated by phase variation,
- a comparison of the application of the two geometries.

This paper is organised as follows. The biological background and the basic model are described in Section 2. In Section 3 we explore the application of the two alternative geometries using Cartesian and polar coordinates, and the analysis is presented in Section 4. We conclude our paper with a brief summary in Section 5. Some additional data are given in the Appendix.

2 Phase Variation in Bacterial Colony Growth

Background. Microbial populations commonly use a stochastic gene switching process called phase variation, controlled by reversible genetic mutations, inversions, or epigenetic modification [14]. Understanding of its adaptive role has traditionally been within the context of "contingency gene theory" [10] in which populations will predictably include variants adapted to "foreseeable" frequently encountered environmental or selective conditions [12]. The mechanistically most common switches are mediated by random mutations in simple sequence repeats, as exemplified by H. influenzae [8], H. pylori [13, 15] and Neisseria [18]. Recent reconsideration suggests a different and additional role for phase variation in the generation of predictable functional diversity within multicellular microbial populations, providing differentiated sub-specializations within structured and predictable communities. 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.

The most readily observable compositional effect of phase variation in cultures grown in vitro is colonial sectoring. In this paper we present preliminary stochastic models that address colonial patterning including bi-directional 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 and B, which develop over time by cell division. Cell division may involve cell mutation, and back-mutation alternates phenotypes; see Fig. 1. We are interested in the proportion of phenotypes in the cell generations, and how their spatial distribution evolves over time.

Basic model of phase variation. We start with the equations taken from the previous *deterministic* model of phase variation [16], which describe *synchronous* growth in cell colonies with two phenotypes A and B, but no spatial



Fig. 1: Phase variation, basic scheme. α / β – forward/backward mutation rate.

aspects. These equations include the assumption that "if phase variation occurs, the progeny consists of one A and one B." Previously [16], behaviour was explored by iterating the equations on a spreadsheet. We develop a stochastic Petri net (SPN) that is directly executable by playing the token game which facilitates its comprehension, and permits the exploration of the behaviour by standard analysis and simulation techniques. Our initial SPN model, see Fig. 2, adopts an *asynchronous* modelling approach so that cells divide individually. The model parameters were taken from [16]; α and β represent the forward and backward mutation rates, and d_A , d_B the fitness of phenotype A and B, i.e. the proportions that survive to division.



Fig. 2: SPN corresponding to Fig. 1; v – marking-dependent stochastic rates.

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

$$propA = \frac{A}{A+B}; \ propB = \frac{B}{A+B}$$
 (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 with the previous synchronous model.

Folding. To prepare for spatial modelling of cell colonies we fold our first (uncoloured) Petri net. We define two colour sets, *Phenotype* and *DivisionType*, see Appendix A, to fold the two places A and B into one coloured place *cell* with the colour set *Phenotype*, and to fold the four transitions into the coloured transition *division*. We obtain the model in Fig. 3. The derivation of our final model requires three further steps: adding space, controlling colony spreading, and controlling thickness, which we discuss in the next section.



Fig. 3: SPN^{C} as SPN short-hand notation; unfolding this SPN^{C} generates the SPN in Fig. 2.

3 Adding space

The colony is represented in 2.5 dimensions by an explicit 2D grid with an implicit constant maximal height over all grid positions.

3.1 Alternative geometries

Starting from a small initial population the colony spreads out as the number of bacteria increases maintaining a circular shape throughout its development. Thus, a circular geometry with polar coordinates for representing space seems to be most appropriate for this particular modelling task. However, previous attempts to model bacteria colony growth have represented space employing a rectangular geometry with Cartesian coordinates. Independently of the chosen spatial representation, the 2D space is discretised in compartments which are then mapped to a grid. Each position of the grid is referenced by a unique tuple (x, y), corresponding to a colour tuple in the model, where x is the index of the row and y of the column in the grid, respectively. Differences between modelling in these two coordinate systems will be highlighted next.

Cartesian coordinate system. In the Cartesian coordinate system [19] approach, the 2D space is discretised by splitting it into equally sized rows and columns obtaining a 2D grid similar to a matrix as shown in Fig. 4. The mapping

between this matrix and the compartments of the 2D grid is direct, because each position in the matrix corresponds to a compartment in the grid. The area of all the 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 neighbour. The neighbourhood relation between different positions of the grid is represented as a function. The maximum number of neighbours for each position is up to 8 depending on whether the considered position is in the interior of the grid, at the edge or in the corner.

Polar coordinate system. On the other hand, when considering a polar coordinate system [19], the 2D space is discretised in a different manner. First of all, the space is divided into evenly spaced concentric circles. Each one of the concentric circles and its immediate enclosing circle will form an annulus [19]. All annuli are then split into sectors obtaining annular sectors like the ones presented in Fig. 4.



Fig. 4: Discretising space considering Cartesian (left) and polar (right) coordinates. Each annulus in the polar case is 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.

When running a simulation from the centre of the discretised space, it is important that the offsprings are able to be displaced with equal chance in either of the directions identified by the sectors. For this purpose the origin of the space is considered as a position in the grid which has as neighbours all the immediate surrounding annular sectors. Therefore, the first row of the 2D grid will contain only one entry, the origin.

The number of neighbours for the origin is equal to the number of sectors. Similar to the neighbourhood relation in a Cartesian coordinate system, all other annular sectors have maximum 8 neighbours, depending if their position is next to the origin, in the interior or at the edge.

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 while 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 transition rate function. Conversely, in case of the rectangular geometry, the transition rate function is not influenced by the volume of the positions since it is constant.

Another important aspect which 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 form 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 non-zero degrees angle associated. However, in the rectangular geometry a non-zero degrees angle is formed only if the colony spreads out on multiple rows and columns. In order 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.

Representing the geometries using colour sets. In spite of the multiple differences between the rectangular and circular geometries, the definition of the colour sets used for each Petri net is the same. The *Grid* colour set is equal to the Cartesian product of the *Grid2D* and *Phenotype* colour sets where *Grid2D* represents the two-dimensional grid and *Phenotype* the type of the bacteria; in our case either A or B.

Each Petri net place represents a subset of 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, while decreasing it makes room for more bacteria.

One crucial difference between the geometries consists of the neighbourhood relation between two positions. This characteristic is captured by the neighbourhood functions described in Appendix B, *neighbourhood2D_rectangular* and *neighbourhood2D_circular*. They define all possible movements in the net. The neighbourhood function for polar coordinates may appear to be more complicated. However, its length is due to the need of separately considering the neighbours of the origin and not because of an increased complexity.

In this case study we are concerned with mutation rates and their influence on the system behaviour. Therefore, 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 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 colony

Controlling colony spreading. The probability of staying with the parent or being displaced to a neighbouring position is modelled differently depending on the representation of space.

In the circular case the probability of a bacteria to be 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 Fig. 5 to capture this aspect. Considering a particular annular sector, the only bacteria which are able to be displaced from this sector to a neighbouring sector are the ones lying on the edge. Assuming that each bacteria can move in 8 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 which 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. Area_{edge} is given by the maximum area which can be occupied by bacteria of size 1 by 1 lying on the edge. The difference between the edges and interior of an annular sector is depicted in Fig. 5. Area_{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:

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

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



Fig. 5: Interior-edge model used for the circular geometry in order to represent the probability of a bacterium to be 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.

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 a

bacterium to be displaced to a neighbouring position decreases. On the other hand, 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 a 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, while decreasing γ increases the preference to be displaced. Conversely, increasing ω increases the preference to be displaced, while decreasing ω increases the preference to stay with the parent. In the general case, the probabilities of staying with the parent or being displaced to a neighbouring position are:

$$P_{stay \ with \ parent} = \frac{\gamma}{\gamma + (\#neighbours * \omega)}$$

$$P_{displace \ to \ neighbour} = \frac{\omega}{\gamma + (\#neighbours * \omega)}$$
(5)

In the rectangular case #neighbours is equal to 3 if the position on the grid 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 number of sectors" for the origin.

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

Controlling 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. In order to control the thickness we introduce a constant ρ , denoted as *POOLSIZE* in the SPN^{C} model, which limits the maximum number of cells at a certain grid position. The constant ρ is set to give room for 26 generations. 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 Appendix A and D, respectively. *Due to space limitations rate functions are not described here, 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 green colour in the model (see Appendix D, Fig. 10). The pre-transition of the 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 pre-transition record to which annulus a given sector belongs, information which 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 of additional places and transitions. Henceforth, this overhead should not be taken into consideration when comparing the spatial representations.

9

4 Analysing Phase Variation

4.1 Computational experiments

The Petri nets were constructed *using* Snoopy [11], recently extended to support coloured Petri nets [6]. Simulations were run with Snoopy's built-in stochastic simulator and Marcie [7]. 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 coloured Petri net 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(R) Core(TM) i5-2500 CPU @ 3.30 GHz processor, 2 GB DDR3 RAM). Similarly, unfolding a coloured Petri net 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 2000 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 [4]. 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, see Fig. 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 rates and fitness by counting and extracting information from the pattern segments, which in the future could give new insights into the population dynamics of mutation. Currently, the model predicts behaviour which has not been measured so far in the wet lab in the sense that it generates a time series description of the evolution of the patterns in the bacteria colony, while wet lab data just provide snapshots of final states.



Fig. 6: 2D representation of the final state of 4 stochastic simulations, 2 for rectangular (left) and 2 for circular (right), illustrating the development of sectorlike 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, red patches of approximately equal proportions. The black background shows the grid area not covered by phenotype B.

4.2 Parameter scanning

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

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 by starting with an initial value and linearly increasing it after each run.

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 Fig. 7a-7d.

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 width of the colony is reduced; this can be observed in Fig. 7a-7d. The value of ρ was chosen for both geometries in such a way that the most outwards bacteria with respect to the centre do not reach the edge of the grid. The reason for this is that we expected some backpropagation of bacteria from the edge of the grid to affect the final outcome of the simulation.

Changing ω . The second batch of simulations changed only the value of ω for each run. Similar to the selection of values for ρ , the values for ω have been randomly selected from the entire search space. Images representing the final states of two simulations for each geometry are given in Fig. 7e-7h. 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 Fig. 7e-7h 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 to be displaced which means that the clear cut between high and low density areas in the images fades away. Thus, in Fig. 7e-7h the images corresponding to a higher value of ω have a more uniform distribution of concentrations than the ones in which ω was smaller.



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).

4.3 Sector Analysis

In the beginning, the analysis of the sectors was done by looking at the images of the colony at different time points and deciding if the sector-like patterns are similar to the ones in the wet lab. Unfortunately, only few images from the wet lab are available. New wet lab experiments are ongoing, but images of the colonies can not be provided yet.

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 patterns from the final state of the simulation: *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 [1], a sector detection module was implemented which takes images as input. The main steps of the algorithm are given in Appendix E.1.

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 dry or wet lab. Thus, our analysis approach is generic. Since the experiments in the wet lab are still ongoing, the image processing procedure was validated only on in silico generated images, but our expectation is that the approach should similarly work well on images from the wet lab.

Results. One thousand stochastic simulations were run for both the rectangular and circular model with an average simulation time of 50 minutes. 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 Appendix E.2.

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 1. We employed a two-sample statistical test for comparing the results. The data corresponding to all measures and both geometries was tested for normality using the Shapiro-Wilk [17] and the Q-Q plot [21] 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 [9,20] non-parametric 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 2, Appendix E.3.

Both area and angle have higher values in the circular 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 non-zero degree angle associated, while 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, while in the rectangular case only in maximum 8. 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 which 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 Appendix E.3 in order to complement the analytical comparison of the results described above.

Measures	Area		Distance		Angle		Sectors	
		\bigcirc		\bigcirc		\bigcirc		\bigcirc
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

Table 1: Rectangular (\Box) and circular (\bigcirc) sector analysis with μ – mean, σ – standard deviation, c_v – coefficient of variation. Area and distance (from the centre) are given wrt total grid area and maximum distance from the centre.

5 Summary

In this paper we have described a methodology of modelling bacterial colonies which evolve in time and space using rectangular and circular geometries, and 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 paper is on the generic methodologies which we developed and which can be employed for different case studies modelled using coloured Petri nets. Work is ongoing in the wet lab to generate images of actual bacterial colonies which will then be used as targets for model fitting in order 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. We also want to extend our sector detection and analysis procedure from working with 2D sector-like patterns to linear and non-linear 3D surfaces.

All supplementary materials are made available at http://people.brunel. ac.uk/~cspgoop/data/cmsb2013 for the interested reader.

Acknowledgments

This research has been partially funded by the British EPSRC Research Grant EP I036168/1 and the German BMBF Research Grant 0315449H. Ovidiu Pârvu is supported by a scholarship from Brunel University. We would like to gratefully acknowledge the comments of the reviewers which have helped us improve the quality of the paper.

References

- Bradski, G., Kaehler, A.: Learning OpenCV: Computer Vision with the OpenCV Library. O'Reilly, Cambridge, MA (2008)
- Gao, Q., Gilbert, D., Heiner, M., Liu, F., Maccagnola, D., Tree, D.: Multiscale Modelling and Analysis of Planar Cell Polarity in the Drosophila Wing. IEEE/ACM Transactions on Computational Biology and Bioinformatics 99(PrePrints), 1–1 (2012)
- Gilbert, D., Heiner, M., Liu, F., Saunders, N.: Colouring Space A Coloured Framework for Spatial Modelling in Systems Biology. In: Proc. PETRI NETS 2013. p. to appear. LNCS, Springer (June 2013)
- 4. Gillespie, D.: Exact stochastic simulation of coupled chemical reactions. The Journal of Physical Chemistry 81(25), 2340–2361 (1977)
- Heiner, M., Gilbert, D.: How Might Petri Nets Enhance Your Systems Biology Toolkit. In: Proc. PETRI NETS 2011. pp. 17–37. LNCS 6709, Springer (2011)
- Heiner, M., Herajy, M., Liu, F., Rohr, C., Schwarick, M.: Snoopy a unifying Petri net tool. In: Proc. PETRI NETS 2012. LNCS, vol. 7347, pp. 398–407. Springer (June 2012)

- Heiner, M., Rohr, C., Schwarick, M.: MARCIE Model checking And Reachability analysis done efficiently. In: Proc. PETRI NETS 2013. p. to appear. LNCS, Springer (June 2013)
- Hood, D.W., Deadman, M.E., Jennings, M.P., Bisercic, M., Fleischmann, R.D., Venter, J.C., Moxon, E.R.: 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 (Oct 1996), PMID: 8855319
- Mann, H.B., Whitney, D.R.: On a Test of Whether one of Two Random Variables is Stochastically Larger than the Other. The Annals of Mathematical Statistics 18(1), 50–60 (1947)
- Moxon, E.R., Rainey, P.B., Nowak, M.A., Lenski, R.E.: Adaptive evolution of highly mutable loci in pathogenic bacteria. Current biology: CB 4(1), 24–33 (Jan 1994), PMID: 7922307
- Rohr, C., Marwan, W., Heiner, M.: Snoopy a unifying Petri net framework to investigate biomolecular networks. Bioinformatics 26(7), 974–975 (2010)
- Salan, L., Ayraud, S., Saunders, N.J.: Phase variation mediated niche adaptation during prolonged experimental murine infection with helicobacter pylori. Microbiology (Reading, England) 151(Pt 3), 917–923 (Mar 2005), PMID: 15758236
- Salan, L., Linz, B., Suerbaum, S., Saunders, N.J.: The diversity within an expanded and redefined repertoire of phase-variable genes in helicobacter pylori. Microbiology (Reading, England) 150(Pt 4), 817–830 (Apr 2004), PMID: 15073292
- Salan, L., Snyder, L.A., Saunders, N.J.: Adaptation by phase variation in pathogenic bacteria. Advances in applied microbiology 52, 263–301 (2003), PMID: 12964248
- Saunders, N.J., Peden, J.F., Hood, D.W., Moxon, E.R.: Simple sequence repeats in the helicobacter pylori genome. Molecular microbiology 27(6), 1091–1098 (Mar 1998), PMID: 9570395
- Saunders, N., Moxon, E., Gravenor, M.: Mutation rates: estimating phase variation rates when fitness differences are present and their impact on population structure. Microbiology 149(2), 485–495 (2003)
- Shapiro, S.S., Wilk, M.B.: An analysis of variance test for normality (complete samples). Biometrika 3(52) (1965)
- Snyder, L.A., Butcher, S.A., Saunders, N.J.: Comparative whole-genome analyses reveal over 100 putative phase-variable genes in the pathogenic neisseria spp. Microbiology (Reading, England) 147(Pt 8), 2321–2332 (Aug 2001), PMID: 11496009
- 19. Weisstein, E.W.: Wolfram MathWorld. "http://mathworld.wolfram.com/", http://mathworld.wolfram.com
- 20. Wilcoxon, F.: Individual Comparisons by Ranking Methods. Biometrics Bulletin 1(6), 80-83 (1945)
- Wilk, M.B., Gnanadesikan, R.: Probability plotting methods for the analysis of data. Biometrika 55(1), 1–17 (Mar 1968)

A Colour-related definitions

The colour-related definitions common to both circular and rectangular geometries:

```
const D1 = int with 101;
const D2 = D1;
colorset Phenotype = enum with A, B;
colorset DivisionType = enum with replicate, mutate;
colorset CD1 = int with 1-D1;
colorset CD2 = int with 1-D2;
colorset Grid2D = product with CD1 x CD2;
colorset Grid = product with Grid2D x Phenotype;
var c : Phenotype;
var div : DivisionType;
var x,xn : CD1;
var y,yn : CD2;
```

B Neighbourhood functions

The neighbourhood function for the Cartesian coordinate system is:

fun neighbour2D_rectangular(CD1 x,CD2 y,CD1 xn,CD2 yn) bool:
// (xn,yn) is one of the up to eight neighbours of (x,y)
// or (x,y) itself
 (xn=x-1 | xn=x | xn=x+1)
 & (yn=y-1 | yn=y | yn=y+1)
 & (1<=xn & xn<=D1) & (1<=yn & yn<=D2);</pre>

Similarly, the neighbourhood function for the polar coordinate system is defined as follows:

```
fun neighbour2D_circular(CD1 x,CD2 y,CD1 xn,CD2 yn) bool:
// (xn,yn) is one of the up to eight neighbours of (x,y)
// or (x,y) itself
(1<=xn & xn<=D1 & 1<=yn & yn<=D2) &
((x=1 & y=1 & xn=1 & yn=1) |
(x=2 & ((
    (y=1 & yn=D2) |
    (y=D2 & yn=1) |
    (yn=y-1 | yn=y+1 | yn=y)
) & (xn=x+1 | xn=x)) | (xn=1 & yn=1)) |
```

```
(y=D2 \& yn=1) |

(yn=y-1 | yn=y+1 | yn=y)

\& (2 < x) \& (xn=x | xn=x-1 | xn=x+1))
```

C Probability of a bacterium to be displaced to a neighbouring position in the circular case

As stated in Section 3, the probability of a bacterium to be displaced to a neighbouring position is:

$$P = \frac{3}{9} * \frac{\text{Area}_{\text{edge}}}{\text{Area}_{\text{grid position}}}$$
(6)

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. Let us denote by M the number of annuli, by N the number of sectors and by r the radius of the actual bacterial colony.

The value of $Area_{edge}$ is given by the maximum area which can be occupied by bacteria of size 1 by 1 lying on the edge. Considering the annular sector AS in Fig. 8:

$$Area_{edge_i} = Area_{left_i} + Area_{top_i} + Area_{right_i} + Area_{bottom_i}$$
(7)



Fig. 8: $Area_{edge}$ is equal to the sum of the four subareas: left, top, right and bottom. The value of the subarea left is equal to the value of subarea right.

The value of $Area_{left_i}$ is equal to $Area_{right_i}$ and they are given by the area of the rectangle of width equal to the width of a bacterium (i.e. 1) and the height equal to the absolute difference of the radii defining the annular sector. Considering that all concentric circles defining the annuli are equidistant the height of the rectangle is equal to r divided by the total number of annuli. Thus:

$$Area_{left_i} = Area_{right_i} = 1 * \frac{r_{colony}}{M} = \frac{r_{colony}}{M}$$
(8)

The value of $Area_{top_i}$ is equal to the area of the annular sector defined by the outer radius r_{outer} of the annular sector AS and the inner radius r_{inner} , such that $r_{outer} - r_{inner}$ is equal to the height of a bacterium (i.e. 1). Thus, we obtain:

$$Area_{top_{i}} = \frac{\pi r^{2}(i+1)^{2}}{M^{2}N} - \frac{\pi}{N} \left(\frac{r(i+1)}{M} - 1\right)^{2}$$
$$= \frac{\pi r^{2}(i+1)^{2}}{M^{2}N} - \frac{\pi}{M^{2}N} \left(r^{2}(i+1)^{2} - 2r(i+1)M + M^{2}\right) \qquad (9)$$
$$= \frac{\pi (2r(i+1) - M)}{MN}$$

The value of $Area_{bottomi}$ is equal to the area of the annular sector defined by the inner radius r_{inner} of the annular sector AS and the outer radius r_{outer} , such that $r_{outer} - r_{inner}$ is equal to the height of a bacterium (i.e. 1). Thus, we obtain:

$$Area_{bottom\,i} = \frac{\pi}{N} \left(\frac{ri}{M} + 1\right)^2 - \frac{\pi r^2 i^2}{M^2 N} = \frac{\pi}{M^2 N} \left(r^2 i^2 + 2riM + M^2\right) - \frac{\pi r^2 i^2}{M^2 N} = \frac{\pi (2ri + M)}{M N}$$
(10)

In conclusion, after summing up all elements:

$$Area_{edge_i} = \frac{2rN + 2\pi r(2i+1)}{MN}$$
(11)

The value of $Area_{grid position}$ is computed as the value of the annulus to which the sector belongs divided by the total number of sectors:

$$\operatorname{Area}_{\operatorname{grid position}_{i}} = \frac{\pi r^{2}(2i+1)}{M^{2}N}$$
(12)

D Coloured Stochastic Petri Net models



Fig. 9: $\mathcal{SPN}^{\mathcal{C}}$ model based on the Cartesian coordinates system.



Fig. 10: SPN^{C} model based on the polar coordinates system. Immediate transitions are highlighted in green.

E Sector detection and analysis

E.1 Sector analysis algorithm

Algorithm 1 Algorithm for sector analysis

Require: *imageFile* is valid

Ensure: Results written to *outputFile*

- 1: Read imageFile and create an instance of type Mat from it called image
- 2: Change the contrast and brightness of image for highlighting regions of interest
- 3: Remove noise from image using blur filters
- 4: Apply morphological close operation on *image* in order to connect close regions which form a sector
- 5: Threshold *image* and store the result in *binaryImage*
- 6: Detect the contours of sectors in binaryImage, approximate their polygons and convexHulls
- 7: for all $hull \in convexHulls$ do
- 8: Compute *distance* from the centre, *area* and *angle* using *hull*
- 9: end for
- 10: Print results into *outputFile*

E.2 Sector detection



Fig. 11: 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 coloured with blue. The size of the sector area has to be higher than a threshold value in order for the sector to be considered.

E.3 Statistics

Table 2: The p-value obtained from the statistical tests in rectangular (\Box) and circular (\bigcirc) case for all measures (area, distance from the centre, angle, number of sectors). 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.

Measures	Shapir	o-Wilk	Mann-Whitney		
		\bigcirc			
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	$<\!\!2.2e\text{-}16$	6.68e-13		

The following figures describe the distribution of the mean of one of the measures (area, distance from the centre, angle, number of sectors) using a circular (left) and rectangular (right) geometry. The frequency represents the number of sectors in each bin of the histogram. Each curve follows a normal distribution having the mean and standard deviation equal to the mean and standard deviation of the sample data used to create the histogram.



Fig. 12: Measure: Area of sectors.



Fig. 13: Measure: Distance from the centre of the grid.



Fig. 14: Measure: Angle of sectors expressed in degrees.



Fig. 15: Measure: Number of sectors.

F Contribution of authors

The first author, Ovidiu Pârvu, is leading author of this submission. He is registered as a full time PhD student at Brunel University, School of Information Systems, Computing and Mathematics.

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 SPN^{C} and polar coordinates, implemented the visualisation and the sector detection and analysis modules, run the parameter scan procedures and created most of the figures, executed the batches of one thousand simulations on the cluster and did the statistical analysis of the results. Moreover he wrote Sections 3 and 4 of the paper.
- David Gilbert (20% 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 5 of the paper.
- Monika Heiner (10% of effort) built the original SPN^{C} model using Cartesian coordinates and wrote the Section 2 of the paper 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 [16] on which our paper builds. He wrote the "Background" subsection in Section 2 of the paper.

All authors read and approved the entire paper.