VRIJE UNIVERSITEIT AMSTERDAM

MASTER THESIS

Investigating travelling waves in a system of communicating cells

Author:
Douwe A.J. Grundel, BSc

Supervisors:
Yiteng Dang, MSc
Dr. Hyun Youk
Prof. Dr. Frank J. Bruggeman

MSc thesis of minor internship (XM_405032, 30 ECTS)

in the

Youk Lab - Physics of Cellular Systems
Department of Bionanoscience, Technische Universiteit Delft

February 1, 2019

Abstract

In multicellular systems cells coordinate their behaviour in both space and time by sensing and secreting signalling molecules, thereby creating dynamic patterns. These patterns are traditionally explained by reaction-diffusion models. However, a disadvantage of these reaction-diffusion models is that they do not take into account cells explicitly. Therefore, a new model, a cellular automaton that takes into account diffusion, was introduced to address this problem. In this thesis I use this model to investigate a specific dynamic pattern: the travelling wave (TW). The goal is to identify properties required for and properties enabling the prediction of TW formation and propagation.

In particular, I investigate the cellular properties TW require. These cellular properties are specified by a genetic circuit that determines how a cell responds to the concentration of signalling molecule that it senses. I show that only specific circuits produce TWs. Furthermore, the effect of initial conditions is investigated. The initial fraction of active genes can largely determine the amount of TW that will form. The initial amount of spatial order may also play a role. Finally, I quantify how much any given pattern resembles a TW. This quantification indicates that TWs form suddenly and that predicting the time of TW formation is difficult.

In conclusion, the cellular properties and initial conditions together can likely be used to estimate the probability that a TW will form if some additional research is done. However, the exact timing of TW formation is more difficult to predict, especially because some evidence points to chaos-like behaviour in TW formation.

This thesis is a sub-part of work that may help increase the understanding of dynamic patterns in multicellular systems.

Acknowledgements

I would like to thank all members of the Youk lab for their helpful comments during the group meeting. I am also grateful to my supervisors: Yiteng Dang, Dr. Hyun Youk, and Prof. Frank Bruggeman. I especially would like to thank my Delft supervisors: Hyun and Yiteng.

Hyun, thanks for your insightful comments and suggestions during our meetings. I also appreciate your kindness in general.

Yiteng, I am especially grateful for your support, suggestions, and ideas. You were always very generous with your time. Even when I had many questions, or in case of long discussions, you remained kind and considerate. Thank you!

Contents

Chapter 1

Introduction

In many biological systems, ranging from the ecological scale to the single cell scale, spatiotemporal processes need to be coordinated if the system is to function correctly. In order to achieve this coordination subparts of the systems send signals to other subparts, thereby creating dynamic patterns.

An example of such a dynamic pattern on the large, ecological scale is a phenomenon called waves of agitation. These waves are observed in populations of swarming animals (across different species: insects, fish, birds) in reaction to an attack by a predator. Several individual animals close to the predator will perform escape manoeuvres. This manoeuvre is a signal for neighbouring individuals, who then copy the behaviour and consequently the behaviour spreads as a wave through the entire swarm (Hemelrijk et al., 2015).

On the other end of the spectrum of scales are intracellular dynamic patterns. One example is the travelling waves observed in proteins of the Min system in the bacterium E. coli. The proteins of the Min system ensure that cell division occurs in the middle of the cell by positioning the Z-ring (a protein complex) correctly. The position of the Z-ring is determined by a concentration minimum of one of the Min proteins. This concentration minimum is a consequence of wave patterns of the Min proteins (Loose et al., 2011).

The system scale of interest for this thesis lies between these two extremes. Here I will focus exclusively on dynamic patterns in multicellular systems. Most relevant examples for this scale come from developmental biology. I will briefly mention three such examples here. First, cell cycle synchronization of cells in the Drosophila embryo is regulated by travelling waves of the Cdk1 protein (Deneke et al., 2016). Second, the branched structure of the lungs in mice can be explained by the pattern that arises from the interaction between sonic hedgehog (SHH) and fibroblast growth factor (FGF) 10 (Menshykau et al., 2012). Third, the rugae of the palate in mice also obtain their striped structure through an interaction of SHH and FGF (Economou et al., 2012). Interestingly, the patterns that formed in these three examples can all be explained by a common mechanism: a Turing type reaction-diffusion mechanism (Deneke et al., 2016; Menshykau et al., 2012; Economou et al., 2012). Moreover, many other (dynamic) patterns in different (developmental) systems on the multicellular scale can be explained by reaction-diffusion mechanisms (Deneke and Di Talia, 2018; Green and Sharpe, 2015; Marcon and Sharpe, 2012).

The Turing mechanism, first described by Alan Turing, explains how (dynamic) patterns can arise as a consequence of diffusion from an initially homogeneous system. When this homogenous initial system is only slightly perturbed at any location, diffusion could cause the system to move away from the homogeneous steady state and form patterns. More specifically, the pattern is formed through an interaction between, at least, two signalling molecules (morphogens). In the simplest case one molecule, the activator, promotes both its own synthesis as well as the synthesis of the second molecule, the inhibitor. The inhibitor molecule inhibits the synthesis of the activator. If the inhibitor diffuses away faster than the activator, patterns with constant wave length emerge (Gierer and Meinhardt, 1972; Turing, 1952).

Patterns on the scale of multicellular (developmental) systems can therefore be explained eloquently in terms of two (or more) signalling molecules by reaction-diffusion models. However, such models do not explicitly take into account how these signalling molecules affect cells. Moreover, the models do not take cells into account explicitly at all. Instead they assume that there are infinitely many cells in any given space (Turing, 1952). Therefore, traditional reaction-diffusion models do not provide insight into the properties individual cells of the system require to generate patterns at the multicellular level. This is unfortunate, because in these multicellular systems the cells are the subparts of the systems that coordinate their behaviour. Learning more about these cell properties is likely valuable for our understanding of multicellular systems.

While cellular responses can be taken into account in reaction-diffusion models (Gomez and Arcak, 2017), such models are still most accurate in the limit of large population sizes. In contrast, developmental systems have small populations of dozens or hundreds of cells in their early stages. At such small population scales, stochastic effects are likely to be important. It is known that cell-to-cell variability and stochasticity play important roles at the single-cell level (Raj and van Oudenaarden, 2008). To take such effects into account requires an individual-based model where cells are modelled explicitly.

In order to get insight in the cellular properties required to sustain patterns in a multicellular system, a new computational model was developed by Maire & Youk and Olimpio, Dang and Youk (Maire and Youk, 2015; Olimpio et al., 2018; Dang et al., prep). This model addresses the problems introduced above. Instead of an infinite number of spatially undefined cells, there is a finite number of cells at a specific location which take up a discrete amount of space. Each cell can both sense and secrete two distinct signalling molecules (fig 1.1). Whether or not a cell secretes a signalling molecule is dependent on the concentration of the signalling molecule the cell senses and the genetic circuit of the cell. The genetic circuit could, for instance, resemble the previously discussed activator-inhibitor system: sensing molecule 1 would then promote the transcription and translation of both gene 1 and gene 2, resulting in secretion of both molecules. Sensing molecule 2 would then inhibit the transcription and translation of molecule 1. When a signalling molecule is secreted it diffuses away and may be sensed by other cells (or the same cell).

It is important to stress that this model does not seek to explain a specific system. Instead, the hope is that general properties on the scale of multicellular systems can be identified by taking into account individual cells in the model. In other words, the goal is to investigate what can happen in such systems. Therefore, no specific signal transduction pathways are assumed. The sensitivity of a cell to signals is tuned with a Hill-type equation (Weiss, 1997).

My role in this project is to use this newly developed model and investigating a particular type of dynamic pattern: the travelling wave. Such waves are clearly distinguishable and often fulfil important signalling functions in biological systems (Deneke and Di Talia, 2018). More specifically, there are four main goals. First, investigate which cellular properties (system parameters) are able to produce travelling waves. Second, investigate to what extent the initial condition - the initial spatial composition of cellular states -affects the formation of travelling waves. Third, investigate how stable travelling waves are once they have formed. Fourth, investigate means of quantifying how much a system resembles a wave, at any given time. Together, these aspects may be used to predict whether a specific situation will result in a travelling wave or not.

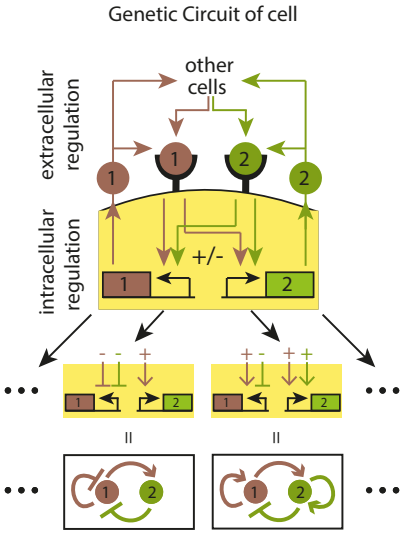


FIGURE 1.1: Figure taken from Dang et al. (prep). It shows that each cell can both sense and secrete two distinct signalling molecules. The genetic circuit of the cell determines whether each signalling molecule acts as activator, inhibitor or does not affect the cell.

Chapter 2

Methods

2.1 Model description

The multicellular system is modelled as a cellular automaton in Matlab. Instead of arbitrarily selecting a number of neighbours that can influence a cell, diffusion and degradation of the signal molecules is explicitly modelled. This determines how far secreted signal molecules travel. It is assumed that the diffusion happens on a faster time scale than cells can measure the concentration of signalling molecules in their environment. In other words, the signalling molecule concentrations are at steady state. Cells update the status of their genes based on the steady state concentration of signal molecule they sense on their surface. The cells are placed on a hexagonal grid and periodic boundary conditions are used.

In the remainder of this subsection the update rule (relation of the state of all cells at time t to the state of all cells at time t + 1) for the cellular automaton will be derived. This is not something I worked on myself. I adapted this subsection from Yiteng Dang's work. More details of the model may be found in other publications (Maire and Youk, 2015; Olimpio et al., 2018; Dang et al., prep).

The state of the system at time t, $X(t)$, is specified by $X(t) = \{X_k(t)\}_{i=1}^{N}$, where N is the number of cells in the system and X_k is the state of cell k, described by an activation value for each of the two genes: $X_k = (X_k^1, X_k^2)$. In the majority of this thesis cells are assumed to respond to signals with a step function (infinite Hill coefficient). Consequently, X_k^i can only be 0 (gene off) or 1 (gene on) in those cases.

Cells secrete signalling molecules at a rate $C_k^{(i)}$, where i indicates the signalling molecule (1 or 2) and k indicates which cell (1 to N). The secretion rate has bounds: $C_{Off}^{(i)} \leq C_k^{(i)} \leq C_{On}^{(i)}$, where $C_{Off}^{(i)}$ is the secretion rate of signal molecule i when gene i is off, and $C_{On}^{(i)}$ is the secretion of signal i when gene i is on. The rates are normalised with respect to $C_{Off}^{(i)}$, so $C_{Off}^{(i)}$ = 1. The relationship between the cell state and secretion rate is given by the following equation:

$$C^{(i)}(X_k^i) \equiv (C_{On}^{(i)} - C_{Off}^{(i)})X_k^i + C_{Off}^{(i)} \tag{2.1}$$

In the common scenario in this thesis, where X_k^i is binary, this means: $C^{(i)}(X_k^{(i)} = 1) = C_{On}^{(i)}$ and $C^{(i)}(X_k^{(i)} = 0) = C_{Off}^{(i)}$. In the other cases, where a finite hill coefficient (no step function) is used, cell states are continues: $X_k^{(i)} \in [0, 1]$.

Using the steady state assumption, the concentration of signal molecule secreted by a cell decays with the distance to that cell according to:

$$c^{(i)}(r) = C_k^{(i)} f^{(i)}(r) \tag{2.2}$$

$$f^{(i)}(r) = \frac{\lambda^{(i)}}{r} e^{\frac{r-R}{\lambda^{(i)}}} sinh(R) \tag{2.3}$$

where r is the distance from cell k, R is the radius of cell k, and $\lambda^{(i)}$ is the signalling length of molecule i. $\lambda^{(1)}$ is set to 1 and $\lambda^{(2)}$ is expressed in relation to $\lambda^{(1)}$.

Now it is possible to calculate the concentration cell k senses (Y) at any given time:

$$Y_k^{(i)}(t) = \sum_{m=1}^{N} f_{km}^{(i)} C_m^{(i)}(t) \tag{2.4}$$

where $f_{km}^{(i)}$ is the distance-dependent interaction strength between cells k and m. Explicitly,

$$f_{km}^{(i)} \equiv \begin{cases} f^{(i)}(r_{km}), & \text{if } k \neq m. \\ 1, & \text{if } k = m. \end{cases} \tag{2.5}$$

with r_{km} as the distance between cells k and m and $f^{(i)}$ as in equation 2.3.

Now the concentration of signal molecule i any given cell k senses has been specified. The next step is considering how this concentration influences the cell.

The effect of signalling molecule j on the secretion of signalling molecule i can be one of three possibilities: 1) stimulating, 2) inhibiting, or 3) non-existent. This is specified in the interaction matrix:

$$M_{int}^{(ij)} \equiv \begin{cases} 1, & \text{j activates i} \\ -1, & \text{j inhibits i} \\ 0, & \text{no interaction between j and i} \end{cases} \tag{2.6}$$

Next, it is assumed that the two signalling molecule species have to simultaneously fulfil certain constrains to turn a gene on or off. For instance, in case of an activator and inhibitor, the gene will only be activated when the activator is present in high concentration and the inhibitor in low concentrations. Therefore, the update rule for cell states is multiplicative:

$$X_k^{(i)}(t+1) = \prod_{j=1}^{2} g_k^{ij}(X(t)) \tag{2.7}$$

where $X_k^{(i)}$ is the state of gene i in cell k, and g_k^{ij} specifies how gradual the cellular response of gene i to signal molecule j is. In most cases cells are considered to have binary genes (either on or off). In that case g_k^{ij} is given by:

$$g_k^{ij} = \begin{cases} \theta(Y_k^{(j)} - K^{(ij)}) & \text{activation, } M_{int}^{(ij)} = 1 \\ \theta(K^{(ij)} - Y_k^{(j)}) & \text{inhibition, } M_{int}^{(ij)} = -1 \\ 1 & \text{no interaction, } M_{int}^{(ij)} = 0 \end{cases} \tag{2.8}$$

where $Y_k^{(j)}$ the concentration of signal molecule j that cell k senses, $K^{(ij)}$ is the threshold concentration of gene i to signal molecule j that needs to be exceeded to trigger a cellular reaction, the threshold is assumed to be the same for all cells in the system. θ ensures a step-function is in place (rounds positive values to 1, and negative values to 0). This ensures that the multiplicative behaviour, as discussed above, is in place. In case of a activator-inhibitor interaction a gene will only be switched on (state = 1) if both the concentration of the activator exceeds its threshold and the concentration of the inhibitor does not exceeds its threshold.

In case of a more gradual response to sensed concentrations, continues cell-states (ranging from 0 to 1) can be used. This is implemented using Hill-equations. In that case g_k^{ij} is given by:

$$g_k^{ij} = \begin{cases} \frac{(Y^{(j)})^h}{(K^{(ij)})^h + (Y_k^{(j)})^h} & \text{activation, } M_{int}^{(ij)} = 1 \\ \frac{(K^{(ij)})^h}{(K^{(ij)})^h + (Y_k^{(j)})^h} & \text{inhibition, } M_{int}^{(ij)} = -1 \\ 1 & \text{no interaction, } M_{int}^{(ij)} = 0 \end{cases} \tag{2.9}$$

where h is the Hill-coefficient, taken to be the same for both signalling molecules.

2.2 Defining travelling waves

Travelling waves (TWs) are signal pulses that move across the entire cellular system. Based on observations it is possible to say that the vast majority of TWs consists of bands of three different cell

states, the remaining state is the background through which the wave moves over time. Only in very rare cases waves of two states in a single background are observed. Therefore, TWs are defined as consisting of bands of three different cell states.

TWs can be spatially oriented in three different ways, for each of these orientations two possible travelling directions are possible: horizontal (up or down), vertical (left or right), or diagonal (left,up or right,down). TWs can also be bended (not in straight line). Examples are shown in figure 2.1 and movies attached to this report (folder: Examples_TW_orientation).

TWs can be defined as waves with a period equal to $\sqrt{N}$, where N is the number of cells. The amplitude can be defined in proportion to the difference of the cellular states in the wave to the cellular state of the background.

2.3 Algorithms for automatic wave detection

Travelling waves (TWs) are detected automatically. The algorithm first selects patterns of a period that could be a TW. Then it is checked whether the spatial order and number of ON-genes remains consistent. If so, a TW is found and further details of the TW can be established. More details in appendix A.

2.4 Spatial Symmetry Score: quantifying wave resemblance using Fourier Transform

In order to be able to quantify how much the system resembles a travelling wave at any given time, a reliable quantification of a travelling wave is required. An important property of travelling waves is that they move over time. At each time step within the period of a wave, it occupies a different location in space. Yet, the quantification for this wave must be the same, regardless of its position in space. This can be achieved using a discrete spatial Fourier transform. The input for this transform is the 2D cellular system at time t, and the output is the so-called Power Spectrum (PS) of time t. Each unique travelling wave will have an unique PS, regardless of its position in the original space. More specifically, each original grid of cells either yields two distinct PS, one for each gene, or the original grid yields four distinct PS, one for each cell state ((ON,ON), (ON,OFF), (OFF,ON), (OFF,OFF)). Both approaches are used.

In order to obtain a PS from the original 2D cell grid, two vectors (x and y) that describe this original 2D space are defined. Consequently, each cell can be described by a unique coordinate (x,y) (visualised in fig B.1). Next, all cells states in the system are normalised (mean-centered and scaled):

$$X(t)^{(i)}_{norm} = \frac{X(t)^{(i)} - < X(t)^{(i)} >}{\sigma(X(t)^{(i)})} \tag{2.10}$$

where $X(t)^{(i)}_{norm}$ is the vector with normalised values of gene/state [1] i, $X(t)^{(i)}$ is the vector containing not normalises values of gene/state i,<> indicates the mean, N the number of cells in the system, and σ the standard deviation.

This system of normalised cells is transformed to Fourier space using the Fast Fourier Transform (as implemented in Matlab):

$$Y^{(i)}_{p+1,q+1}(t) = \sum_{j=0}^{l_x-1} \sum_{k=0}^{l_y-1} \omega_x^{jp} \omega_y^{kq} X^{(i)}_{j+1,k+1}(t) \tag{2.11}$$

with:

$$\begin{cases} \omega_x = e^{\frac{-2\pi i}{x}} \\ \omega_y = e^{\frac{-2\pi i}{y}} \end{cases} \tag{2.12}$$

, where i is the imaginary unit, p and j range from 0 to $l_x - 1$, q and k range form 0 to $l_y - 1$.

[1] Depending on which of the two types of PS is used (two genes vs four states)

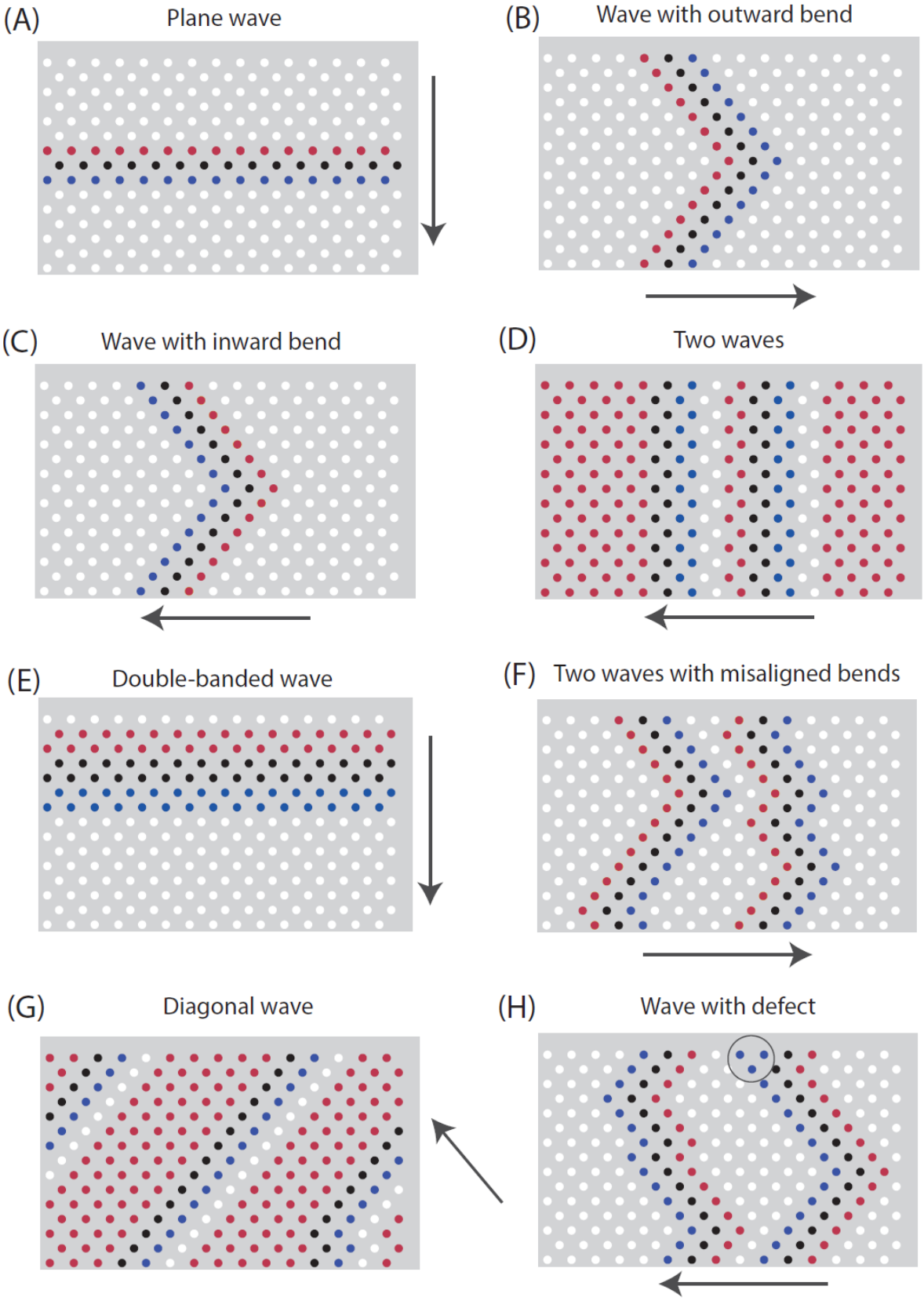


FIGURE 2.1: This figure was taken from Dang et al. (prep). Examples of different travelling waves. However, (H) is not considered a true wave, since it has no spatial symmetry. Colour indicates cell state (gene 1, gene 2): white (OFF,OFF), red (ON,OFF), blue (OFF,ON), and black (ON, ON). The arrow indicates the travelling direction of the wave.

Next, the PS is also normalised. This and the previous normalisation ensure the PS sums to 1:

$$PS^{(i)}(t) = \frac{|Y^{(i)}_{p+1,q+1}(t)|^2}{(N-1)\sqrt{N}} \tag{2.13}$$

, now the PS of gene/state i is obtained. A visual representation of the PS is shown in appendix B.

Travelling waves have the property that all bands (see figure 2.1) of the wave have the exact same spatial orientation. Therefore, the PS of each state (except the background state) will be identical. This spatial symmetry may be used to quantify how much any given state resembles a travelling wave. There are other patterns that also have this property, but they are more trivial and are less likely to spontaneously occur during the simulations, let alone persist multiple time steps. However, there is one other common situation where the PS of all states are identical: a homogeneous grid. When the system becomes homogeneous, however, the simulation stops. It can therefore be distinguished from a wave.

The spatial symmetry score (SSS) is based on the identical PS of the wave-band states. The SSS is the normalised sum of the pairwise difference of the Fourier modi of the three wave-band states:

$$SSS_{states} = \frac{\sum_{j=1}^{2} \sum_{\substack{k=2 \\ k \neq j}}^{3} |PS^{(j)} - PS^{(k)}|}{6} \tag{2.14}$$

, the division by 6 ensures the SSS upper bound is 1 (max difference PS $= 2$, three pairs). In case of a travelling wave, the three PS are identical. Therefore, a travelling wave will result in a SSS of 0. When all three PS are completely different, the SSS will be equal to 1. Appendix figure B.3 shows a visual representation of equation 2.14.

In case of non-binary cell states it is impossible to use PS of cell states. Instead, a PS is calculated for each gene. Then the SSS is calculated in a similar manner:

$$SSS_{genes} = \frac{|PS^{(j)} - PS^{(k)}|}{2} \tag{2.15}$$

In appendix B some additional variations of the SSS are considered.

2.4.1 Quantifying dominant spatial orientation

Power Spectra (PS) can also be used to determine the dominant spatial orientation of a given grid of cells. The PS are obtained as described above, but instead of calculating the SSS they are used to establish what the most prevalent orientation of a given grid is. This can be done using specific rows and columns of the PS, which match a specific spatial orientation, as can be seen in appendix figure B.2.

The dominant spatial orientation is calculated using the three PS corresponding to the three cell states that occur in the TW. This because the dominant orientation is compared to the orientation of the TW that will eventually form. More specifically, for each orientation (horizontal,vertical, diagonal) the specific column/row is summed across the three PS. For instance, for all three waves states the central PS columns are summed to obtain the horizontal orientation value. The maximum value of the three resulting orientation values is the dominant orientation of the grid. For vertical orientation the central row can be used instead of the column. For diagonal orientation a different coordinate system is used. This ensures that the central row of the PS corresponds to the diagonal orientation.

2.5 Statistical Methods

In order to determine whether variables contribute to travelling wave formation, generalised linear models (linear and logistic models) are used. These models are fitted using the statistical programming language R. They are explained in more detail together with the results in appendix C.2.

Chapter 3

Results

3.1 Only 5 genetic circuits with identical core-structure generate travelling waves

There are 44 distinct possible genetic circuits for the two gene system that is used here (Dang et al., prep). The genetic circuit specifies how cells respond to a signal. It determines whether a signal can affect the cell, and if it can, it determines whether this signal promotes or inhibits the secretion of signalling molecules, as was already shown in figure 1.1. The connections (arrows) in a genetic circuit do not assume a specific signal transduction mechanism. They can be interpreted as the summary of the signalling cascade. Travelling waves (TWs) likely require specific genetic circuits to be in place. To confirm this, simulations were run for all circuits. Only 5 genetic circuits turned out to produce TWs (fig 3.1). Interestingly, these network share a core structure: an activation-inhibition interaction between the two signals with an additional self-activation on one of the two genes. Furthermore, there is also a consistent pattern in the strength these interactions should have in order to produce TWs. The strength of one type of interaction deviates from all the others. Positive self-interactions are always very sensitive to signals compared to all other interactions. In other words, the positive self-regulation needs to be relatively strong in order for TWs to form. Both of these aspects, the core structure and the strong positive self-regulations, are no coincidence. Yiteng Dang developed a (heuristic) theoretical argument for the necessity of these features (Dang et al., prep).

Together, these two aspects (activation-inhibition between genes and strong positive self-regulation) give the required cellular properties for a system capable of producing TWs. However, simulations show that a system that meets both conditions does not always produce TWs. Therefore, the next question is what properties within such a system affect the formation of TWs, if any.

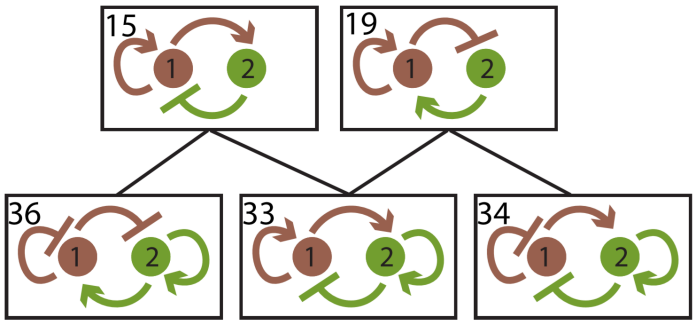


FIGURE 3.1: Adapted from Dang et al. (prep). The only genetic circuits that can produce travelling waves. Arrow indicates activation, flat arrow inhibition. Colour indicates from what source signal the effect comes (brown effects of signal 1, green effects of signal 2). The numbers are the unique identifiers of the circuit (out of the possible 44 circuits).

3.2 Initial fractions of active genes impact travelling wave formation

Once suitable system parameters (cellular properties), as described in the previous subsection, have been picked, TW formation in principle only depends on the initial state of the system, because the model is deterministic. Therefore, any initial condition of the system (cell states of all cells at t = 0 given the model parameters) can be assigned one of two labels: 1) initial condition that will yield a TW or 2) initial condition that will not yield a TW. This raises the question whether initial conditions have certain properties that can be used to obtain the probability that an initial condition will yield a TW. Properties with high predictive value do not necessarily exist. It could be that system behaves in a chaotic manner. Slightly different initial conditions would then lead to entirely different time trajectories and end results. If that is the case predicting the formation of a travelling wave may be impossible. Here, the effect of two important properties of the initial state on TW formation will be investigated: the fraction of active genes and the spatial order.

The fraction of active genes is simply the fraction of genes that is turned on (for both gene 1 and gene 2). Figure 3.2 shows how the fraction of formed TWs depends on the initial fraction of active genes (p_1 and $p2$). The figure shows that no TWs form for extreme values of p. This makes intuitive sense. For instance, very low initial p values cannot produce a TW. Both p's will go to zero, all genes will be OFF and the system will stay in the homogeneous state.

Given the clear effect of the initial fraction of ON-genes on TW formation, the consequent dynamics of both p values was further investigated. This to establish whether an initial difference in active gene fractions would persist throughout the simulation. Figures 3.3 and 3.4 shows this is not the case. All simulations that continue for longer than about 8 time steps have trajectories trough p-space that usually spend the majority of simulation time in a subspace where it seems the trajectories behave chaotically. Even simulations with initial p values close to the termination point of the simulation usually do, for example figure 3.3b. Simulations that are shorter than 8 time steps often terminate when the system is homogeneous, for example 3.4a.

Another indication that simulations with identical initial values for p can have chaos-like trajectories is shown in figure C.1 (appendix). It shows that the time required to form a TW does not depend on the initial p values.

The termination points in p-space make sense. All p-space trajectories of waves terminate at p value combinations that correspond to the correct number of cells for a travelling wave. All p-space trajectories of simulations that did not yield travelling waves either end up with in a stable homogeneous system ($p_1 = p_2 = 0$ or $p_1 = 1, p_2 = 0$ for this circuit and parameters), or reach the maximum number of simulated time steps and therefore are still in the chaotic subspace, such as figure 3.4c.

3.3 Indication that initial spatial clustering affects travelling wave formation

Setting the fractions of initially active genes can largely determine whether TWs will form. However, fixing only these initial fractions (and the system parameters) does not specify a specific spatial system (except for homogeneous systems). There are multiple spatial arrangements of cell states that will result in the same fraction of active genes. Some of these spatial arrangements result in a TW, other do not. There may be common properties in the spatial arrangements that yield a TW. Here, I will investigate the effect of two aspects of the initial spatial order on TW formation. First, the effect of clustering of cells with the same genes switched on. Second, whether the initial spatial orientation is associated with the orientation of the TW that forms.

The spatial clustering is measured separately for each gene, using a modified Moran's I (Moran, 1950; Olimpio et al., 2018):

$$I^{(g)} = \frac{N}{\sum_{i,j\neq i} f(r_{ij})} \frac{\sum_{i,j\neq i} f(r_{ij})(X_i^{(g)} - < X^{(g)} >)(X_j^{(g)} - < X^{(g)} >)}{\sum_i (X_i^{(g)} - < X^{(g)} >)^2} \tag{3.1}$$

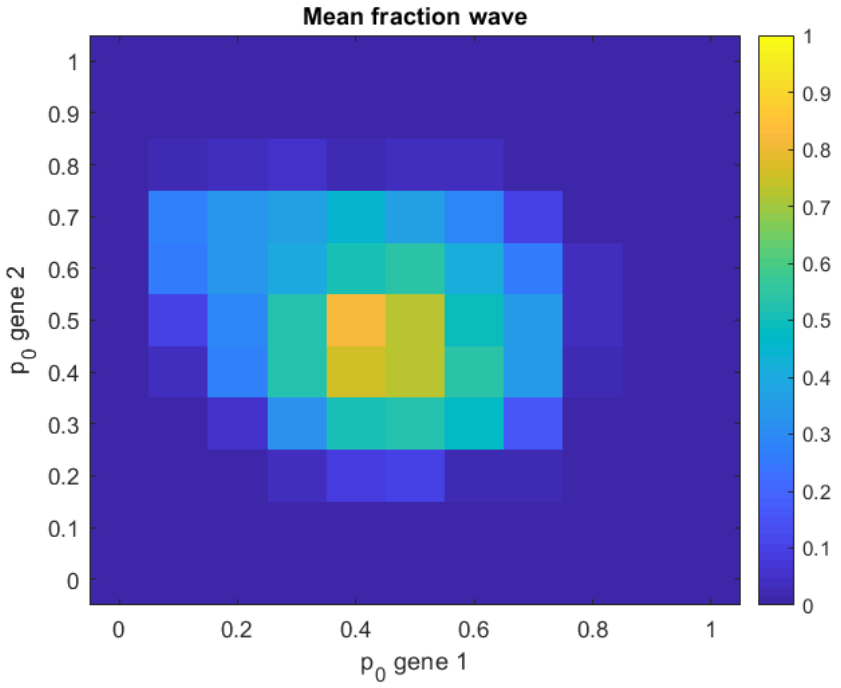


FIGURE 3.2: Heatmap showing the fraction of travelling waves (indicated by colour) for different fractions of initially active genes (p_1 and p_2 for gene 1 and 2 respectively). The spatial order is kept constant ($I^{(1)} = I^{(2)} = 0$). In total 18150 simulations were run, 150 for every unique p value combination.

where $I^{(g)}$ denotes the spatial clustering of gene g, N is the number of cells, $X_i^{(g)}$ the state of gene g in cell i, and $f(r_{ij})$ as in equation 2.3. The value of I ranges from: $-1 < I < 1$ [1]. $I = 0$ indicates a randomly ordered system, a large $|I|$ indicates an ordered system. For $I < 0$, cells with the same gene on avoid each other and for $I > 0$ cells with the same gene on are clustered together (fig C.2). For homogeneous systems I is set to 0.

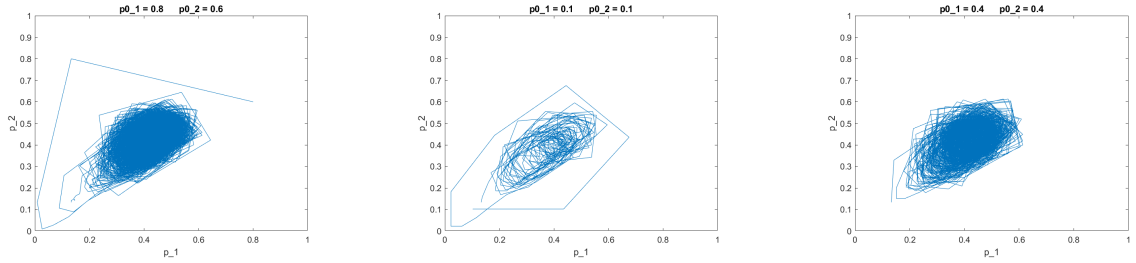
The effect of the initial amount of spatial clustering on TW formation was investigated by running simulations where the initial $I^{(1)}$ and $I^{(2)}$ were varied. Each simulation was run until it terminated, because a static or periodic pattern emerged, or it reached the maximal number of time steps (10000).

Figure 3.5 shows that the mean fraction of formed waves is highly similar, between 0.6 and 0.7, across most combinations of initial I values. This indicates that the initial spatial clustering of the cell states has little impact on whether the system will end up producing a travelling wave. However, it should be noted that for the lowest I values the fraction of travelling waves appears to be consistently lower.

In order to draw more statistically rigorous conclusion about the results displayed in figure 3.5, a logistic regression model was fitted to the simulation data using the statistical programming language R. A logistic regression model uses one or multiple predictor variables to calculate a probability that a sample belongs to one out of two possible classes. In this case we can use the initial values of $I^{(1)}$ and $I^{(2)}$ to predict if a TW forms (classes: TW, no TW).

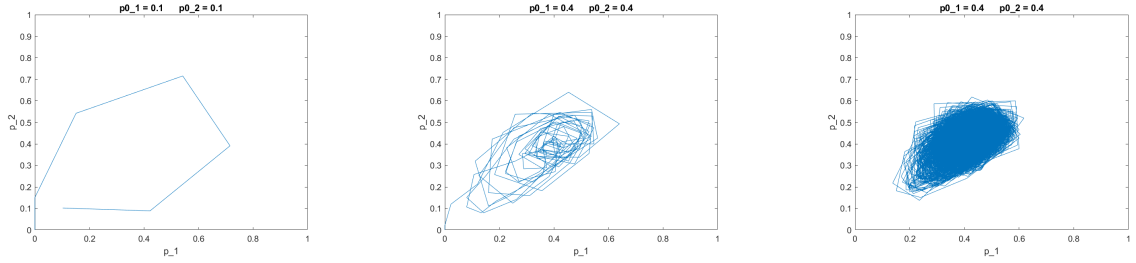
First it was determined that the logistic regression model using the information of both I values (the proposed model) was significantly better than the null model. The null model only uses the mean fraction of simulations that became a wave calculated using all simulations (0.5435), without considering specific I values. The fact that the proposed model was significantly better than the null model indicates that there is information in the I values. However, the residual deviance (indication goodness of fit model based on the loglikelihood of the data given the model) hardly drops going from the simple null model to the more complicated proposed model (fig C.3). This means that the proposed model only significantly explains a small part of data better than the null model. This corresponds with the fact that the trend that lower initial I-values yield fewer TWs is not very strong (fig 3.5).

[1] In practice I does not get smaller than about -0.1, however. As mentioned in (Olimpio et al., 2018)



(A) (B) (C)

FIGURE 3.3: p-space plots of three simulations that ended up yielding a travelling wave. The simulations had different initial p values: a) $p_1 = 0.8, p_2 = 0.6$, b) $p_1 = 0.1, p_2 = 0.1$, c) $p_1 = 0.4, p_2 = 0.4$



(A) (B) (C)

FIGURE 3.4: p-space plots of three simulations that did not result in a travelling wave. The simulations had the following initial p values: a) $p_1 = 0.1, p_2 = 0.1$, b) $p_1 = 0.4, p_2 = 0.4$, c) $p_1 = 0.4, p_2 = 0.4$

Next, the proposed model is used as a classifier. For each of the simulations the logistic model uses the initial I-values to predict whether the simulation will result in a TW or not. More specifically, the logistic model gives a probability that initial I values ($I^{(1)}$ and $I^{(2)}$) will result in a TW (fig C.4). If this probability exceeds 0.5 a TW is predicted.

The classifier always predicted a wave, except for the relatively low I values (fig C.5) that indeed seem to correspond to the lower wave fractions in figure 3.5. The overall performance of the classifier was further assessed and the results are in table 3.1. The accuracy and AUC give an indication of the general quality of the predictions, the recall indicates how well TWs were predicted, and the specificity how well non-TW were predicted. A random performance in this case correspons to 0.50, because there are only two classes. The performance metrics are explained in more detail in appendix C.2.1.

The performance metrics indicate that the logistic classifier performed slightly better than random (accuracy > 0.5 and AUC > 0.5). Most TW were correctly predicted (high recall), but this is only because the classifier mostly predicts waves. Many simulations that did not yield a TW were incorrectly labelled as TW (low specificity). It should be noted that the classifier was used to predict the data it was trained on. The reported performance scores are too high in such cases, thus the model performance score would likely be somewhat lower when the classifier is confronted with new data. In case of such a drop, the overall performance of the classifier would likely be random.

A similar argument can be made about the effect of the initial spatial clustering on the time required to form a TW. Simulations with a higher initial spatial order required less time to form a wave. However, only a very small part of the variation in formation time can be explained by the initial spatial order. This is shown in appendix C.2.1.

TABLE 3.1: Performance metrics of the logistic model fitted to the data shown in figure 3.5. The logistic model is used as classifier on the same data as it was fitted on.

Metric	Accuracy	AUC	Precision	Recall	Specificity
Value	0.562	0.555	0.565	0.835	0.235

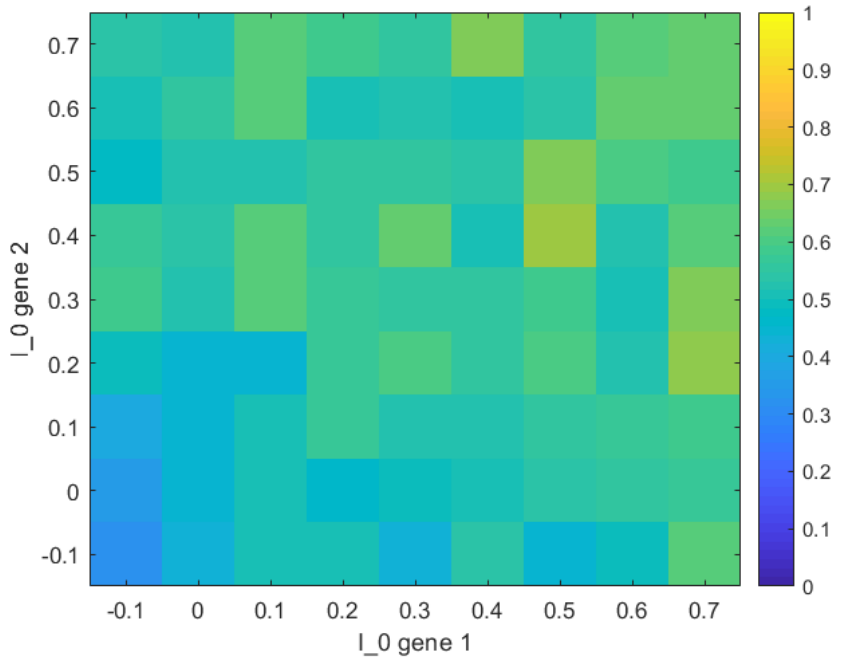


FIGURE 3.5: Heatmap showing the fraction of travelling waves (indicated by colour) for different initial spatial clustering ($I^{(1)}$ and $I^{(2)}$, for gene 1 and 2 respectively). The fraction of active genes is kept constant ($p_1 = p_2 = 0.5$). In total 8100 simulations were performed, 100 simulations per unique I combination

3.4 Initial dominant spatial orientation does not match the travelling wave orientation

Since TWs are oriented in a specific spatial direction (section 2.2), the spatial clustering investigated above may not be specific enough. The I-values do not provide information in which direction the system is ordered, only about the amount of clustering. Therefore, the power spectra obtained by a spatial Fourier Transform are used to determine in what direction the initial cellular grid is predominately oriented (section 2.4.1). For all TWs that appeared in the two simulationsets mentioned above (varying I and p) it was checked whether the dominant initial orientation matched the orientation of the TW. The detailed results of this analysis are shown in appendix C.2.2.

In the first simulation set, where p is varied, this is clearly not the case. The performance metrics correspond to a random situation.

In the second simulation set, where I is varied, it is a bit more complicated. Horizontal was are predicted quite well, but vertical waves not at all. Diagonal waves are absent is this simulation set. However, given that not all wave orientations are predicted reliably, I would argue the dominant initial orientation also does not match the TW orientation in this case.

Therefore, initial spatial orientation of both cases does not directly match the orientation of the TW that eventually forms.

3.5 Similar initial state suggests similar outcome, however trajectories may differ

The obtained results on the importance of the initial condition suggest that only altering the fraction of active genes will have a large impact on the probability of producing a TW. The spatial order of the system seems to be less important, but clustering still has some effect. If spatial order would not be important at all, the behaviour of the systems would likely be chaos-like and the initial states leading to TWs would be randomly distributed in the space of all possible initial states. If spatial order is

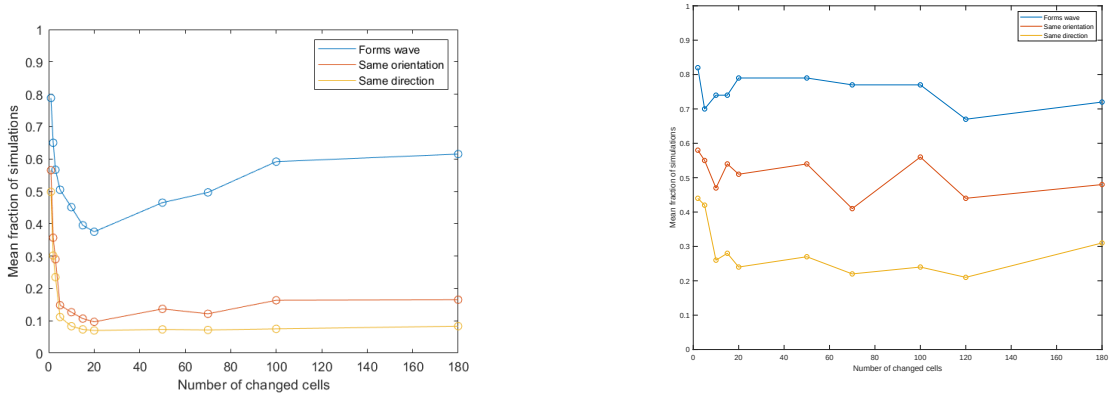
important to some extent, however, it is likely that an initial state that only differs slightly from an initial state that will yield a TW will also result in a TW.

To investigate this an initial state that results in a TW was perturbed. More specifically, the state of an increasing number of randomly chosen cell was changed whilst checking whether the simulation still yields a TW after this alteration (fig 3.6). The key difference between figure 3.6a and 3.6b is the fraction of simulations that is expected to result in a wave by chance. Based on earlier simulations we expect that about 40% and 75% of of randomly started simulations in figure 3.6a and 3.6b would result in a TW respectively.

Figure 3.6 shows that changing only a few cells results in a higher fraction of TWs than the base line. When more than 5 cells are changed the observed fraction of TWs is around the baseline. Furthermore, the fraction of waves in the same direction and orientation as the original wave drops quite fast, especially for fig 3.6a, indicating that in many cases a different TW forms.

The time required to form a TW seems independent of the number of cells that is changed (fig C.8). Furthermore, there is a large variance in time required to form a wave. There is also no difference in time required to reform a wave in the same direction and orientation or a new wave. Together, this suggest that the trajectories (of the full system in state space) differ after perturbation.

This perturbation experiment suggest that there are several initial states close to each other in state space (5 or fewer cells change) that have a higher chance of generating a TW. However, even if the a TW in the same direction and orientation forms in such a closely related initial space, the dynamics of the system through time can be very different.



(A) (B)

FIGURE 3.6: Effect of perturbing an initial condition that yielded a wave by changing an increasing number of randomly chosen cells. The total number of cells is 225. The blue line gives the fraction of simulations that still forms a wave after the perturbations, the red line shows the fraction of simulations that yield a wave with the same spatial orientation as the undisturbed original simulation, the orange line shows the fraction of waves with both the same orientation and travelling direction. a) 600 simulations per number of changed cells in genetic circuit 15. b) 94 simulations per number of changed cells in genetic circuit 33.

In order to investigate the different trajectories for similar initial conditions in more detail, I ran an additional perturbation experiment. Here a single random cell of an initial condition is again changed to a random new state. The trajectories of the unperturbed and perturbed conditions are compared at every time step (fig 3.7). 12% of perturbed initial conditions had identical trajectories to the unperturbed situation. The dissimilar trajectories all reached a baseline distance of about 0.75, indicating that at this baseline a random fraction of cells was in the same state as the unperturbed system at each time step. Before the baseline was reached, trajectories were more similar. As an indication how fast the baseline was reached the time step at which half of the mean maximum distance (of all non-identical trajectories) was determined. This was at 38 time steps. It was also checked whether trajectories were in fact similar, but just shifted in time. This was not the case (appendix C.3.2).

Slightly different initial conditions can therefore have different trajectories in time. This property is also observed in chaotic systems.

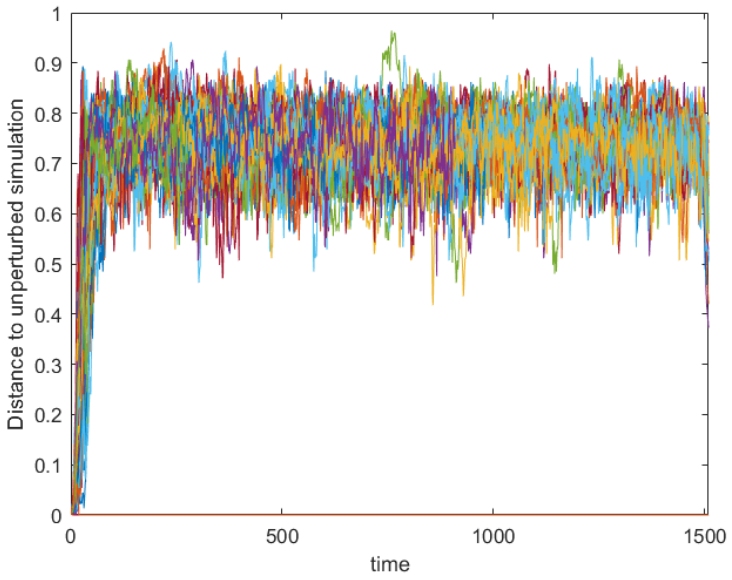


FIGURE 3.7: Perturbation experiment where random single cell was changed to a different state at t = 0. Each line represents a simulation trajectory initiated with such a change. The altered trajectory at time t is compared to the unperturbed trajectory at time t and the normalised distance between the trajectories is plotted on the y-axis. 1 indicating that all cells states are different between the trajectories, and 0 indicating that all cell states are identical. 600 perturbations were performed.

3.6 Travelling waves are stable against perturbation of up to 9% cells

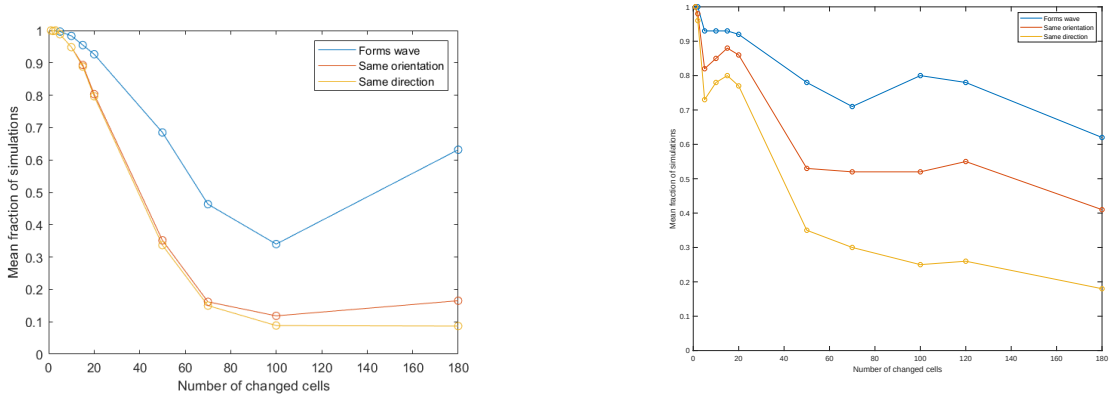
In the previous sections the probability of obtaining a TW was investigated. In this section the focus is on the stability of a TW that has already formed. Here, this stability is investigated by perturbing an increasing number of cells of an already existing TW. Not only the cells of the wave itself can be changed, but also cells somewhere else on the grid.

The results of such wave perturbations are shown in figure 3.8. Just like in the previous section, the key difference between figure 3.8a and 3.8b is the fraction of simulations that is expected to result in a wave by chance. Again, 40% and 75% of of randomly started simulations in figure 3.8a and 3.8b would result in a wave respectively.

In both figure 3.8a and 3.8b it can be seen that changing a single cell will never disrupt the travelling wave and from two cells onward the fraction of waves starts to drop slowly. When 20 cells or more are changed, the fraction of waves drops faster. Finally, between changing 70 - 100 cells the fraction of waves equals the fraction you would expect at random. However, for 3.8a it increases again after. Nevertheless, it appears like travelling waves are quite stable up until more than about 20 cells (9%) are changed. Furthermore, the fraction of waves with the same orientation and travelling direction also starts to diminish faster at this point in figure 3.8a and slightly later in figure 3.8b.

Figure C.9 shows that the time required to (re)form the travelling wave is lower when 20 cells or less are changed. Furthermore, reforming the same travelling wave requires less time than forming a wave with another spatial orientation or travelling direction, but this time advantage diminish as the number of cells that is perturbed increases. However, it should be noted that the variation in wave formation time is quite large and it is likely that not all these trends are statistically significant.

In conclusion, travelling waves are quite stable up until the point where 20 cells (9%) are perturbed.



(A) (B)

FIGURE 3.8: Effect of perturbing an existing wave by changing an increasing number of randomly chosen cells. The blue line gives the fraction of simulations that still forms a wave after the perturbations, the red line shows the fraction of simulations that yield a wave with the same spatial orientation as the undisturbed original simulation, the orange line shows the fraction of waves with both the same orientation and travelling direction. a) 600 simulations per number of changed cells in genetic circuit 15. b) 100 simulations per number of changed cells in genetic circuit 33.

3.7 Only patterns strongly resembling a travelling wave identified by Spatial Symmetry Score

In the sections above TWs were identified once they appeared. However, quantifying TW resemblance for each time step of a simulation could provide more insight. It is possible that trajectories that result in a TW have more TW-like states, for instance. The Spatial Symmetry Score (SSS, section 2.4) is designed for this quantification purpose. Before applying the SSS, its reliability will be investigated.

The most basic requirement of the SSS is that any travelling wave is recognised and will result in a SSS of 0. The SSS based on the four power spectra (PS), one for each state, meets this criterion. However, the SSS based on the two PS, one for each gene, fails in some cases. This is unfortunate, because the SSS version based on two genes is the only one that can be applied to situations where cell states are non-binary (such situations have not been discussed yet).

The reason for the non-zero score of the two genes SSS is that the gene states in TWs are not necessarily spatial symmetrical (unlike the cell-states of the TW). This is for instance the case in figure 3.9. The PS for gene 1 will be based on the yellow and black row, two non-adjacent rows. The PS of gene 2 will be based on two adjacent rows: blue and black. Therefore, the PS of gene 1 and the PS of gene 2 will be different and resulting in $SSS > 0$. It should be noted that this problem does not always arise in TWs where the (ON,ON) state is not in the middle. For example figure 2.1 D also has this property. However, in this case the PS of gene 1 is based on all columns, except the adjacent blue and white column. The PS of gene 2 is based on only the blue and black column, which are also adjacent. Both PS are based on two columns, therefore the absolute signal is identical for this TW and the absolute difference of the PS will be equal to 0.

Given these results, only the SSS based on the four states will be discussed in the remainder of this section. In addition to the basic requirement that TWs end up with an SSS of 0, this version of the SSS also assigns scores close to 0 when structures closely resemble a wave, as expected. Figure 3.10 shows that the SSS gradually drops as the pattern becomes a TW.

However, on patterns that do not closely resemble a TW, the SSS does not distinguish between random patterns and patterns with a more organised structure (fig 3.11). Therefore, the SSS can only be used to identify system states that closely resemble TWs.

Figure 3.12 is an example of applying the SSS to identify TW-like states. The figure shows the SSS at every time step of a simulation. The majority of the system states is around some kind of baseline SSS of around 0.42. This score could correspond to random states (figures 3.11a and C.10), but from observation it is clear these state are more structured, like fig 3.11b. Compared to the simulated time

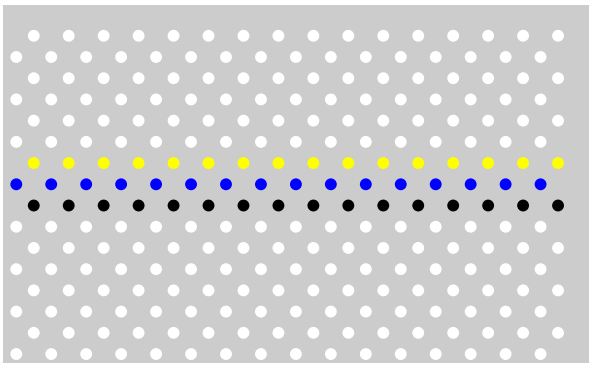


FIGURE 3.9: Example of a travelling wave that does not have a Spatial Symmetry Score (SSS) of 0 when the SSS is based on two genes. Colour indicates cell state (gene 1, gene 2): white (OFF,OFF), yellow (ON,OFF), blue (OFF,ON), black (ON,ON).

the drop in SSS just before the TW has formed is quite sudden. In principle more subtle changes that the SSS cannot pickup could happen before this drop, but this is not supported by observation of the simulation. The SSS starts to drop about 7 time steps before the TW is formed (fig 3.12b). This is not the always the case. In most simulations clear TW-like patterns that can be picked up by the SSS are present only one or two time steps before the TW is formed (fig 3.13). This again indicates that TW appear quite sudden. In some cases the pattern a single time step before the TW did not closely resemble a TW, indicated by 0 in the histogram of figure 3.13. The attached movie clips show some examples of TWs that become apparent at different amounts of time before TW formation (folder: Examples_lengths_clear_TW_formation).

Another interesting feature of figure 3.12a is the drop in SSS that occurs t = 678. This drop indeed corresponds with a transient pattern that closely resembles a wave (fig C.11).

Many simulations show qualitative similar behaviour to that of figure 3.12a. The SSS is around a baseline value, although sometimes the baseline slightly changes, with occasional drops to lower values. TW onset is relatively uncertain. In appendix C.4 (fig C.12) there are examples of other SSS trajectories. These also include the other SSS variants (appendix B). The qualitative behaviour of the different SSS variants is highly similar.

In summary, only the SSS based on the four cell-states can be used reliably. However, even this score is not entirely reliable. It is only sensitive to patterns that already closely resemble a TW. Therefore, it can be used for detection of such patterns. For instance to detect transient TWs or to estimate how many time steps before the TW forms structures closely resembling a wave exist, which appears to be relatively short.

3.8 Travelling waves propagate at lower Hill coefficients, but no spontaneous formation was observed

In order to check how important binary cell states are for the formation and propagation of TW, simulations with continues cell states - where cells respond to signal gradually with a finite Hill coefficient, instead of a step function - were run. Some details are discussed in appendix C.5.

In the first set of simulations it was checked if already existing TWs could propagate for a large set of parameters (fig 3.14). The lowest Hill coefficient where TWs propagated was 4.

In the second set of simulations it was checked whether TWs would form spontaneously for a single set of parameters that has a large fraction of TWs for an infinite Hill coefficient (fig 3.14). No TWs formed spontaneously for Hill coefficients ≤ 12, higher values were not tried.

In both cases the behaviour of the end-state of each simulation was classified in one of four categories. 1) Static homogeneous: all simulations that ended with a static homogeneous system (the fraction of static end-states that were not homogeneous was negligible). 2) Homogeneous oscillations: systems for which the entire homogeneous grid oscillates with a certain period (grids need not be entirely homogeneous, small deviations allowed). 3) Pure wave: a travelling wave. 4) Infinite dynamics: the remaining simulations, containing both periodic patterns that are neither TW nor simple oscillation and simulations that have not finished before the maximum simulation time had passed.

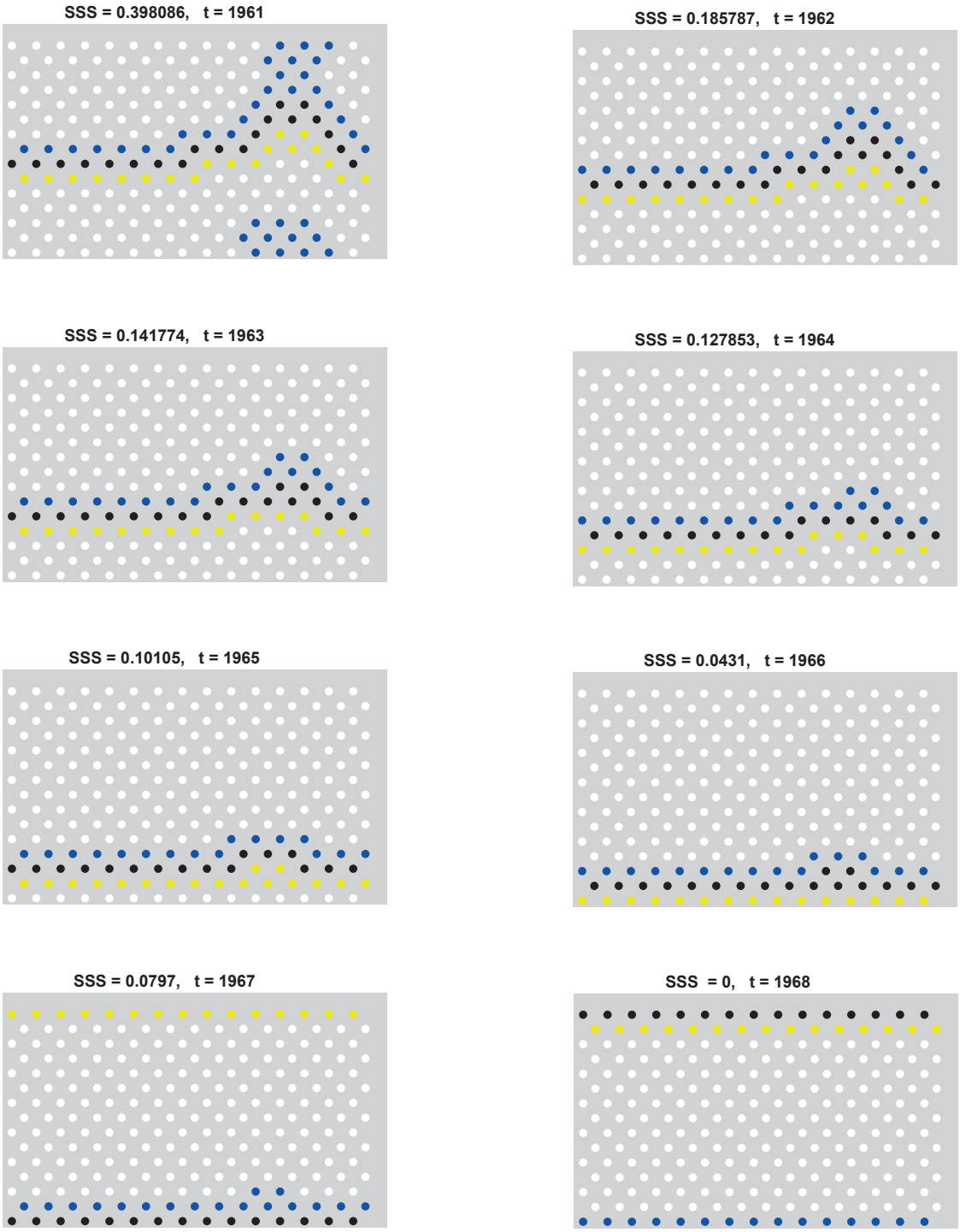
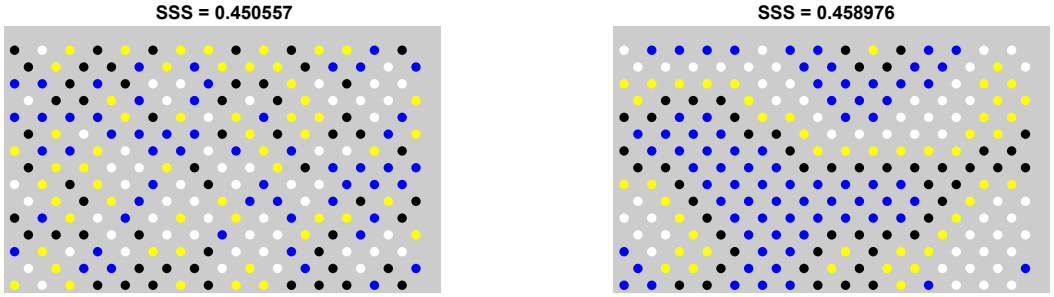
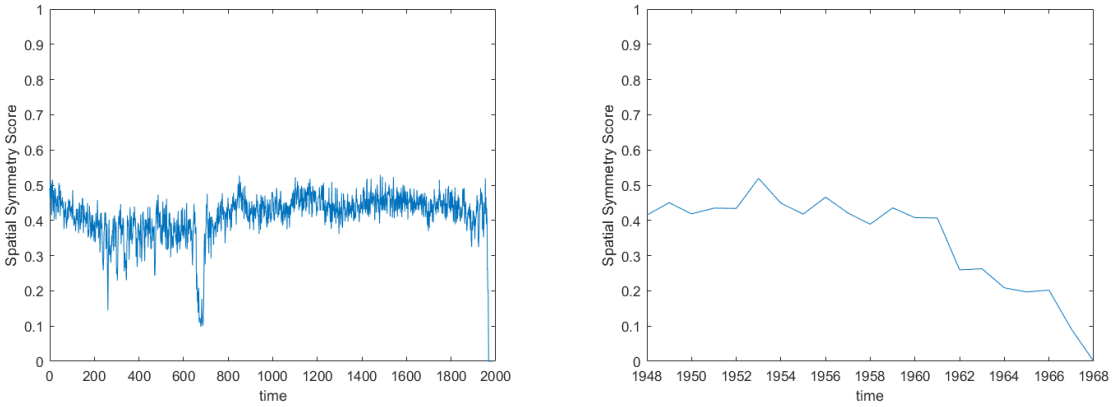


FIGURE 3.10: Example of the final time steps (from left to right) of a simulation where a TW forms. For each time step (t) the Spatial Symmetry Score (SSS) based on four cell states is given.



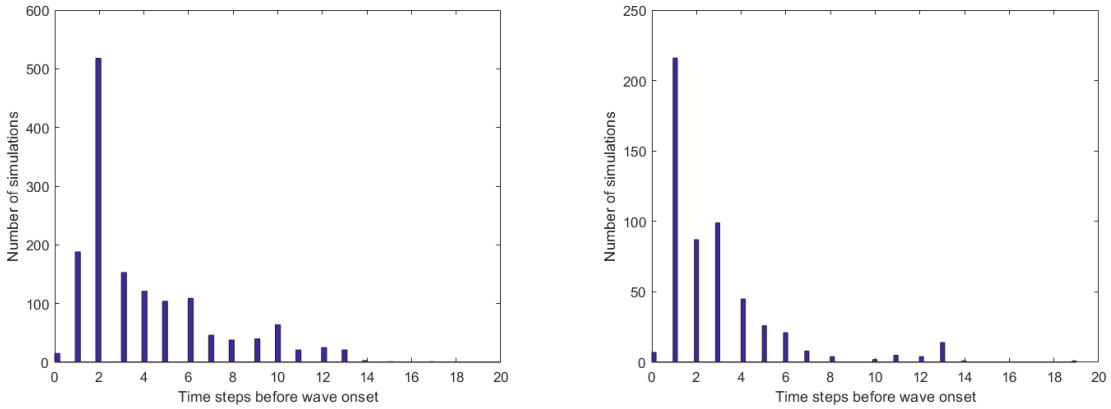
(A) (B)

FIGURE 3.11: Example of two grids with a highly similar Spatial Symmetry Score (SSS) based on the four states, but different patterns. A) Random pattern. B) Pattern with more structure. There are some organised bands present.



(A) (B)

FIGURE 3.12: The Spatial Symmetry Score calculated for all time steps of an example simulation. A travelling wave forms at t = 1968. A) Shows the entire simulation, B) is zoomed in on the last 20 time steps.



(A) (B)

FIGURE 3.13: Histogram number of time steps before the onset of the travelling wave where the Spatial Symmetry Score (SSS) was smaller than 0.3. A SSS < 0.3 roughly indicates a structure that closely resembles a wave. This was done for the simulation in which p was varied. A) for a horizontal TW. B) For a vertical TW.

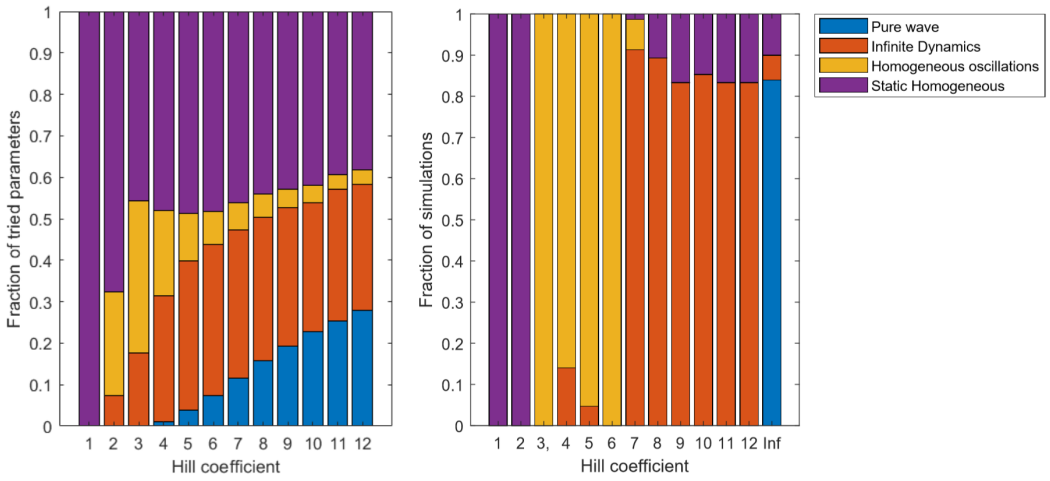


FIGURE 3.14: Showing the fraction of each of four categories, indicated by colour, of distinct behaviour for each Hill coefficient. Both simulations sets were performed in genetic circuit 15. On the left: simulations started with a travelling wave for 2534 different parameter sets (different K and Con) for each Hill coefficient. On the right: simulations started with a random condition (150 different initial conditions) for a single parameter set.

Based on inspection they usually show more complex patterns than simple oscillations. However, this fourth category is not flawless.

In the first simulation set (initiated with TW) the category infinite dynamics mostly consist of periodic patterns. In the second simulation set (random initial grid) the category infinite dynamics mostly consisted of simulations that had not finished for the Hill-coefficients 7-12. For lower Hill-coefficients they consisted of periodic patterns.

Among the movies clips attached to this thesis there are several examples of patterns of different categories at different Hill-coefficients (Examples_Finite_Hill_patterns).

Chapter 4

Discussion & Conclusion

In multicellular systems cells coordinate their behaviour in both space and time by sensing and secreting signalling molecules, thereby creating dynamic patterns. These patterns are traditionally explained by reaction-diffusion models. However, a disadvantage of these reaction-diffusion models is that they do not take into account cells explicitly. Therefore, a new model, a cellular automaton that takes into account diffusion, was introduced to address this problem (Maire and Youk, 2015; Olimpio et al., 2018; Dang et al., prep). In this thesis I used this model to investigate a specific dynamic pattern: the travelling wave (TW).

4.1 Required cellular properties for TWs

First it was shown that TWs only form and propagate with specific cellular properties (genetic circuits) in place. There needs to be a negative feedback mechanism between the two genes that encode for the two signalling molecules. One signal promotes the synthesis of the second signal, and this second signal in turn inhibits synthesis of the first signal. In addition to this activation-inhibition motive a strong self-activation on one of the genes is required. This structure makes intuitive sense: it ensures that fast activation and slow inhibition is in place. This enables oscillatory behaviour of gene states (on or off), which is also required for TW. Moreover, the identified core structure corresponds to two key reaction-diffusion systems: activation-inhibition and substrate depletion (Gierer and Meinhardt, 1972; Kondo and Miura, 2010) (which of the two it resembles depends on where the self-activation is placed).

When cells with continues gene states are used instead of cells with binary genes states, cell-states change gradually. As a consequence the strong self-activation may be less important. However, further research is necessary to establish this.

4.2 Effects initial conditions on TWs

A suitable genetic circuit alone does not guarantee that TWs will from. Initial conditions of the systems also affect TW formation. .

The fraction of initially active genes can largely determine whether TWs will form. For very high or very low initial fractions of active genes TWs usually do not form. For values in between these extremes the fraction of simulations resulting in a TW varies more gradually. The fraction of TW reached also depends on the parameter set (not shown).

Although in some cases fixing the initial fractions of active genes can already result in the majority of simulations producing TWs, there still remains variation, not all simulations will yield a TW. Whether a simulation results in a TW depends on the initial placement of cells. Therefore, I also investigated the importance of the amount of initial spatial order for TW formation.

A small, but statistically significant, trend was found between TW formation and the initial amount of spatial clustering. Initial systems where cells in the same state are not clustered produced fewer TWs than initial systems where cells in the same state are more clustered. However, the fraction of TWs that forms is fairly similar for most values of spatial clustering, except for very high and very low values. Therefore it is not that surprising that a classifier that predicted whether an initial system would result in a TW based on the initial clustering hardly performed better than random. This indicates that initial spatial clustering may not be important, except for extreme values of clustering. However, further research is needed to be sure. An appropriate next step to investigate the importance of initial spatial clustering is running more simulations in which both the initial

fraction of active genes and the initial amount spatial clustering are varied. I did not do this in this thesis, but it is important because the amount of spatial order is constrained by the fraction of active genes. It could be that the importance of spatial clustering depends on the fraction of active genes.

Clustering is a rough measure of spatial order, more subtle spatial structures could be important. In this thesis one such possibility was examined. The dominant initial spatial orientation direction was calculated for simulations that yielded a TW. However, this initial orientation did not mach the orientation of the TW. It is possible that initial spatial order does not matter and initial spatial distribution leading to TWs are more or less randomly distributed in the space of all possible spatial distributions. Nevertheless, it might be worth investigating the subtle spatial patterns in more detail. The dominant spatial orientation used in this thesis does for instance not consider how strong the initial orientation is, it simply picks the direction in which orientation is strongest. Taking into account the strength of orientation may be important. Furthermore, the initial dominant orientation is determined by the summation of particular Fourier modi of all three cell states that occur in the wave. It may be that looking at individual cell states, instead of summing them, provides more insight.

It may not be surprising if TW formation indeed turns out to be relatively unaffected by initial spatial order. After all, the self-organised patterns that emerge from most multicellular (developmental) systems should not be too sensitive to initial variations. However, the shape of patterns in traditional reaction-diffusion models is sensitive to initial spatial order (Page et al., 2005). At least, when there is a relatively strong initial pattern in place, such as a gradient of signalling molecule concentration.

In reality the importance of initial spatial order varies from system to system, as illustrated by the following experiments done in Drosophila embryo's. In the first experiment (Lucchetta et al., 2005) the bicoid gradient (essential for antero-posterior axis development) was perturbed in early embryo's. Despite the absence of the bicoid gradient, the embryo's developed normally. In the second experiment (Johnson et al., 2017) the ERK concentration was perturbed at different spatial location of early embryo's. This resulted in a halt of normal development.

4.3 Predicting timing of TW formation

In addition to the effort of predicting whether or not certain conditions will eventually result in a TW, potential signals of when the TW would form were investigated. Unfortunately, the Spatial Symmetry Score (SSS) can only be used to identify patterns closely resembling a TW. However, this still means it was possible to determine how many time steps before the TW forms patterns closely resembling a TW were present. This was often only a few time steps, thus TWs appear to form rather suddenly. Furthermore, there are also transient TW-like patterns. This means that when the SSS drops you cannot be sure whether a true TW will form or a temporary structure will briefly appear.

Other metrics can also be used to show behaviour that is qualitatively highly similar to the SSS. Instead of plotting the SSS the fraction of active genes, the amount of spatial clustering, or the cell transitions per time step can be plotted to show similar behaviour (Dang et al., prep). These other metrics therefore likely have the same pitfalls as the SSS score. However, investigating the cell transitions more closely might be valuable. There may, for instance, exist subtle differences in what transitions are most abundant just before a transient wave and a true TW.

However, it could also be possible that TW formation is a chaos-like process and it is impossible to find a signal that can be used to predict the timing of TW formations. This is for instance supported by the results of the perturbation experiments. Perturbing a single cell in many cases results in different system trajectories.

When looking at the simulations by eye it is also difficult to judge when a TW will form. During the simulations there are always some sub-patterns that are ordered and TWs form quite sudden. Two movie clips of entire simulations have been attached as an example (folder: Examples_entire_trajectories).

4.4 TWs with finite Hill-coefficients (continues cell-states)

All main results were obtained for cells with binary genes (either on or off). Cells responded to signal concentration with a step function (infinite Hill-coefficient). In order to establish how important this is for TW formation, simulations in which cells respond more gradually (finite Hill-coefficient) to signal concentrations were run.

TWs could propagate for Hill-coefficients of 4 and higher. However, TWs did not spontaneously form in systems with random initial conditions and a Hill-coefficient of 12 or lower (higher finite coefficients were not tried). For these random initialised systems simulations, a high fraction of simulations did not stop before the maximum simulation time had passed, indicating that TWs may have formed if the simulations were allowed to run longer.

It should be noted that the Hill-coefficient of this model is not a Hill-coefficient of a receptor-ligand reaction. Instead, it is a 'lumped' coefficient, a summary of the entire signalling cascade leading to ultrasensitivity.

Such ultrasensitive responses can have high Hill-coefficients. For instance, Cdc25C activation by Cdk1 activity in the Xenopus egg can have a Hill-coefficient as high as 32 (Trunnell et al., 2011).

An example of a developmental system with two signalling molecules is the signal response in lung development in mice (controlled by SSH and FGF), which has a Hill coefficient around 2 (Menshykau et al., 2012).

An example of a signalling cascade with a Hill coefficient between 4-5 (lowest coefficient in which TWs were observed to propagate) is the MAPK cascade . This cascade consists of three kinase reactions and is involved in a variety of cellular processes (among others: differentiation, stress response, apoptosis) (Huang and Ferrell, 1996; Plotnikov et al., 2011).

The lowest Hill-coefficients for which TWs could propgate therefore falls well within the range of signalling cascades of real systems.

4.5 Conclusion

Essential cellular properties required for TW formation and propagation have been identified. Only genetic circuits with an activation-inhibition interaction and at least one strong self-activation can produce TWs. Furthermore, TW formation is also affected by the initial fraction of active genes. The initial spatial clustering seems to be important as well, but further research is necessary to establish the importance of initial spatial order. Lastly, TWs form rather suddenly. No clear signals to predict when this formation will happen were identified. There are some indications that TW formation is chaos-like, but this could be further investigated.

The research in this thesis is a sub-part of work that may help increase the understanding of dynamic patterns in multicellular systems.

Appendix A

Wave detection algorithms

A.1 Wave detection

Two algorithms were used to detect wave. In both algorithms the first step is the same. The simulation is checked for periodic behaviour. At each simulated time step it is checked whether this state has occurred before. If so, the simulation is stopped. The system is deterministic and if the same state reappears behaviour will repeat: a periodic pattern. If such a periodic pattern is detected, the wave algorithms will check if this pattern is a wave.

The first algorithm simply checks if the following values remain constant during the periodic pattern: 1) fractions ON-genes (both gene 1 and gene 2), and 2) amount of spatial order, measured by Moran's I (see Chapter 3). If this is the case, there is a travelling wave.

The second algorithm checks if the power spectrum, obtained by a discrete spatial Fourier Transform of the cellular grid (see Section 2.4), is identical for each time step of the periodic pattern. If so, there is a travelling wave.

The second method is more rigorous, but in practice both algorithms give identical output. The first algorithm is computational less intensive, therefore that is the one that was used in most cases.

A.2 Determine wave properties

Once a wave has been detected, this algorithm can be used to determine the properties of the wave in more detail. If this algorithm is used in the absence of a wave, it will still produce output, but it will make no sense. Another important remark is that this implementation only works for binary gene states.

The algorithm works as follows. The cellular grid at the time the wave first appeared is used to determine: 1) the orientation (horizontal, vertical, or diagonal) of the wave on the grid, 2) the number of waves (sometimes multiple waves follow each other), 3) the number of bands within each wave (this is not established for diagonal waves), 4) the number of times a diagonal wave wraps around each of the two axis (not established for horizontal and vertical waves), and 5) travelling direction of the wave (this also requires the cellular grid of the next time step) 6) Whether the wave is bended or not. Examples of waves with certain properties are shown in figure A.2.

Figure A.1 gives an overview of the algorithm. The first step is identifying what cells are the background state to this wave. These cells will be ignored, one of the non-background states will be selected. For the selected state the number of rows and the number of columns to required to cover all cells of the selected state (fig A.2) is determined. This can be used to determine the orientation of the wave (fig A.1). For instance, if the wave is horizontal, a single row would be required, but N columns. Once the orientation is known, the number of waves can easily be established by counting the bands of the same state that are not adjacent and the number bands within the same wave by the bands of the same state that are adjacent. For diagonal waves the number of wraps around an axis is calculated by counting the number of cells in the same state for a fixed point on the axis in question.

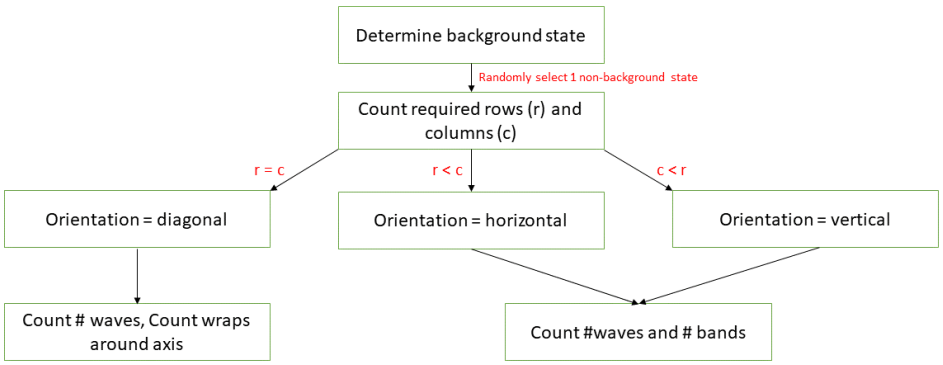


FIGURE A.1: Flowchart describing the algorithm used to determine wave properties. Described in the main text.

(A) (B)

FIGURE A.2: Example for a grid of 25 cells. A) shows a row of cells. A horizontal wave will be orientated in this direction, therefore few rows are required to cover all cells of a horizontal wave. B) show a column of cells. A vertical wave will be orientated in this direction, therefore few columns are required to cover all cells of a vertical wave.

Appendix B

Power Spectra and Spatial Symmetry Scores

B.1 Visualisations

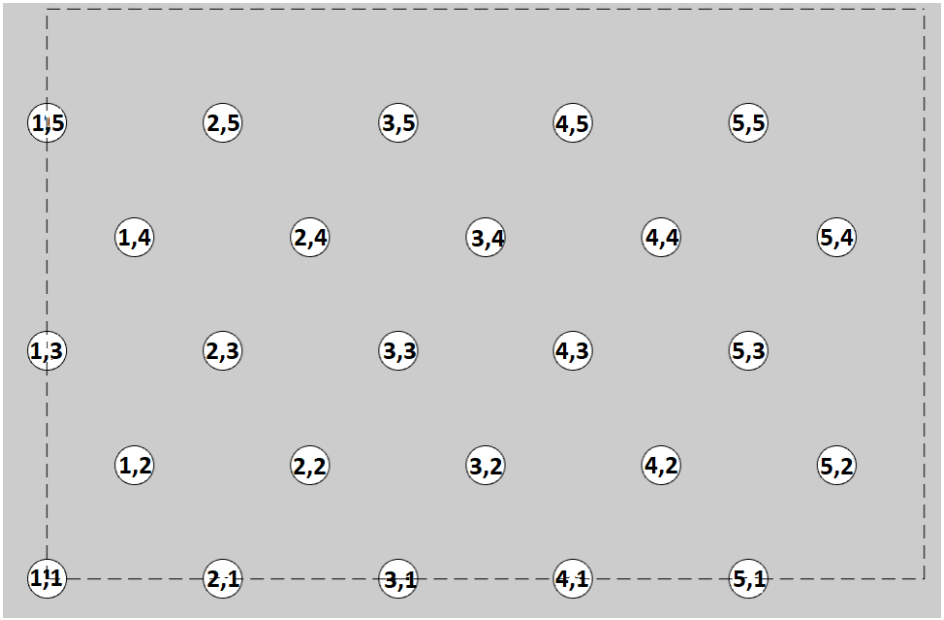


FIGURE B.1: Example of the coordinates (x,y) on a grid of 25 cells.

Figure B.1 shows an example of the coordinates each cell has on a cellular grid. The y-direction is defined in a zigzag manner. Another coordinate system was also used in case of diagonal waves, with the y-direction in a diagonal manner (not shown). Both methods gave similar results and both could be used.

Figure B.2 shows examples of Power Spectra (PS) for three different waves. There are four PS for each of the three waves, one PS for each possible cell state (binary genes: (ON,ON), (ON,OFF), (OFF,ON), (OFF,OFF)). For each wave, three PS are identical and one deviates. The three identical PS correspond to the three states that make up the wave, the other is the background state. For the first two waves (fig B.2a and B.2b) the background state is white, therefore the PS of (OFF,OFF) deviates from the other three. For the third wave (fig B.2c) the background state is yellow, therefore the PS of (ON, OFF) deviates from the others.

Furthermore, the PS of the horizontal wave (fig B.2a) all have a signal spread in the y-direction. This is because in the original figure, the variation is in that direction as well. If you would compare the different x-coordinates with each other, and keep y fixed, there would be no difference between the x-coordinates. In other words, no variation in the x-direction. If you would do the same for y, however, there would be a difference. Some rows contain only white cells, other blue, etc. The same type of reasoning can be applied to the PS of the other waves.

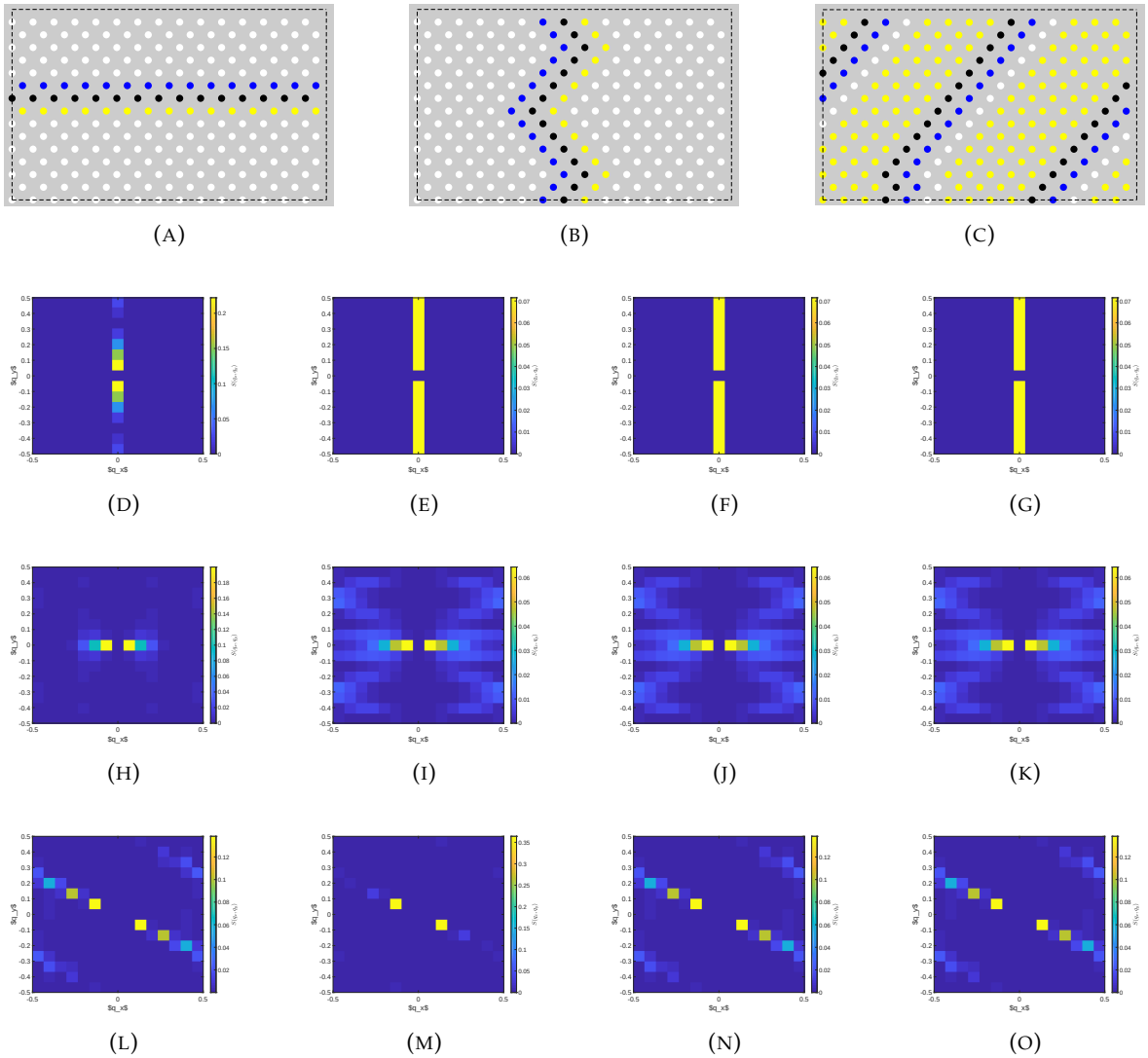


FIGURE B.2: Examples of the Power Spectra (PS) of each of the four states (gene 1, gene 2) for three different waves. A) Is an example of a horizontal wave. B example of a vertical wave. C) example of diagonal wave. The figures underneath the wave examples are visualisations of the PS. The first column (D,H,L) are the PS of state (OFF,OFF), the second column (E,I,M) are the PS of (ON,OFF), the third column (F,J,N) are the PS of (OFF,ON), and the fourth column (G,K,O) of (ON,ON). Each row belongs to one of the three example waves. The first row (D-G) are the PS of the horizontal wave (A), the second row (H-K) are the PS of the vertical wave (B), and the third row (L-O) are the PS of the diagonal wave (C). The cell colour indicates the cell state: white (OFF,OFF), yellow (ON,OFF), blue (OFF,ON), and black (ON,ON).

Figure B.3 again shows some PS. All PS can be understood as a visualisation of a matrix. Each element in this matrix is a Fourier Modus. When the the absolute pairwise difference is taken between two PS, the matrices are subtracted element-wise. When this is done for figure B.3a and B.3b it results in B.3c. All the Fourier modi in B.3c sum to 0.5437. The sum of pairwise difference of the Fourier Modi in the Spatial Symmetry Score (see main text) is calculated in this way.

B.2 Alternative versions Spatial Symmetry Score

Two main other variants of the Spatial Symmetry Score have been used. The first of these is only useful when the outcome of a simulation is known (it can be applied afterwards). If this simulations resulted in a travelling wave, one of the PS when the wave has formed can be used. For all time steps the pairwise difference to the PS of that time step to the PS of the wave can be calculated.

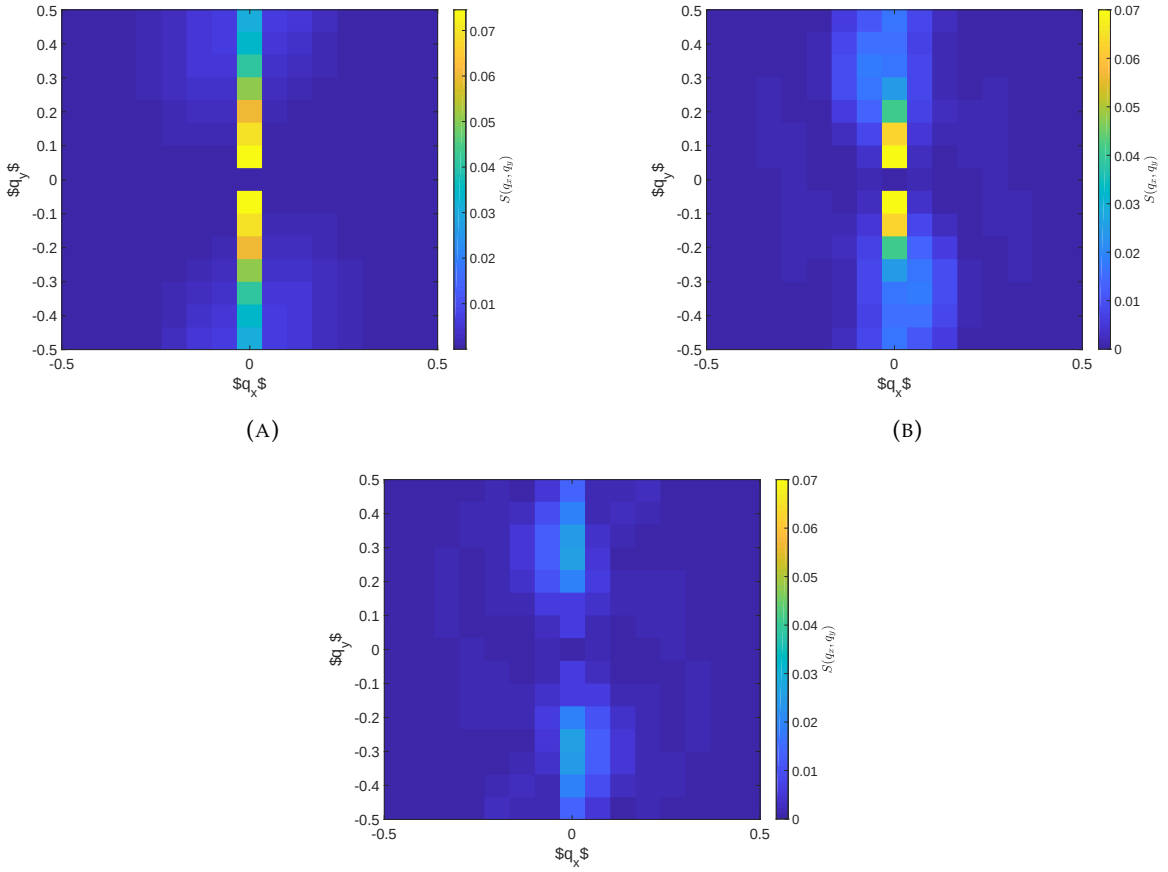


FIGURE B.3: When the absolute difference between the Fourier modi of the PS displayed in A and B is taken, the result is the PS shown in C.

The second way compares PS of genes across time steps instead of within time steps. More specifically the same gene is compared at different time steps, $|PS_{\text{gene i}}(t) - PS_{\text{gene i}}(t+1)|$.

Appendix C

Additional results

C.1 Results concerning initial fraction of active genes

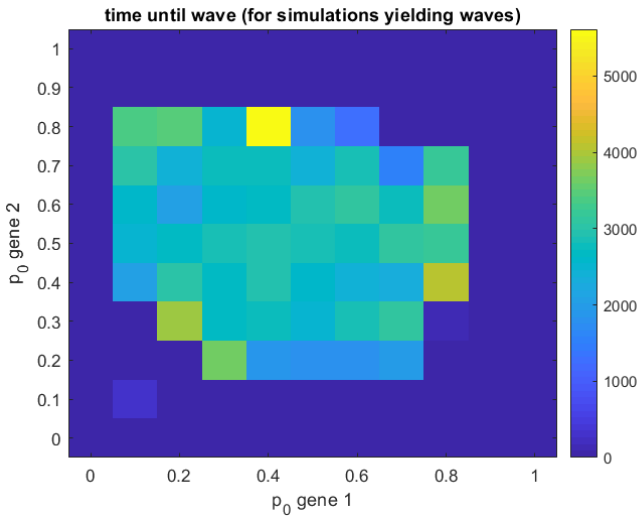


FIGURE C.1: Mean time required to form TW for specific initial p combinations, indicated by colour.

C.1.1 Parameters

Parameters used for the simulations: genetic circuit = 33, K = [63.8186 369.0979; 541.3526 19.8009], Con = [993.7731 790.9581],I0 = [0 0], N = 225.

C.2 Results concerning initial spatial order

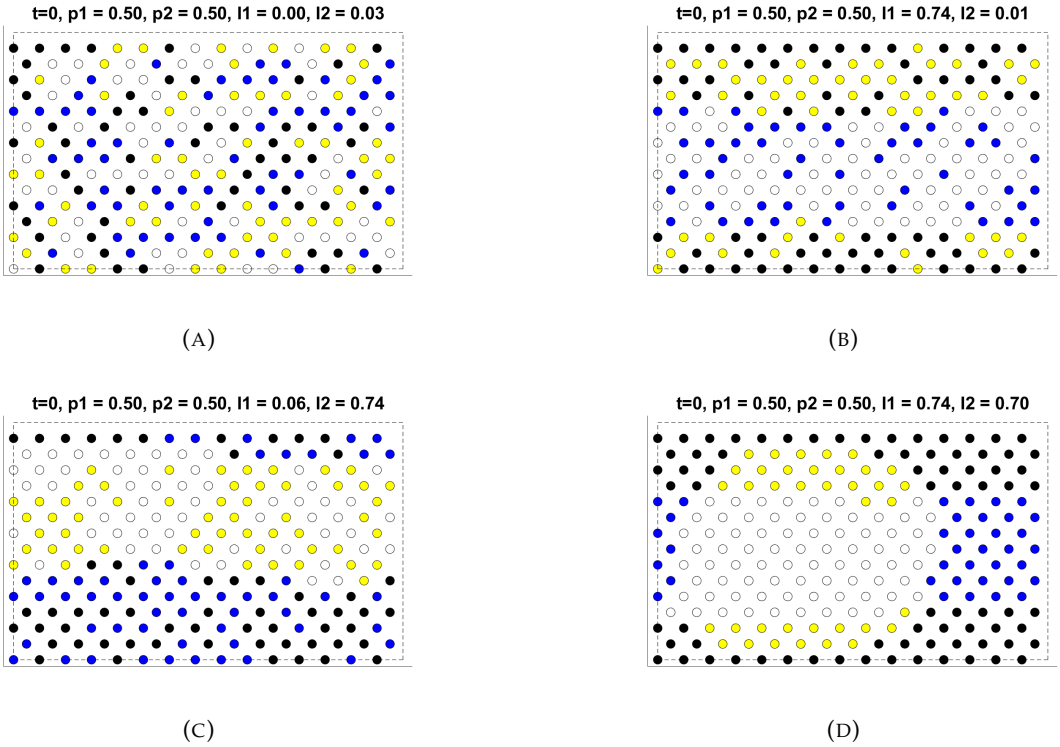
C.2.1 Spatial clustering as measured by I

Figure C.2 gives examples of systems with different values for I.

Figures C.3 shows that the logistic model taking into account both initial I values (and the interaction between them) is significantly better than the other models, as discussed in the main text (chapter 3.3).

For every unique initial I value combination, the obtained logistic model gives the probability that a TW will form (fig C.4). By setting a threshold this model can be used as a classifier. TW will be predicted when the model probability for a TW exceeds 0.5 (fig C.5). The predictions of the logistic classifier are summarised in table C.1, a so-called Confusion Matrix. This matrix displays the true positives (TP), false positive (FP), true negatives (TN), and false negatives (FN). From these values the performance metrics can be calculated:

$$\text{Accuracy} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{FP} + \text{TN} + \text{FN}} \tag{C.1}$$



(A) (B)

(C) (D)

FIGURE C.2: Examples of grids of cells with different values of spatial clustering coefficients I. I_1 and I_2 are the clustering coefficients for gene 1 and 2 respectively. Cell colour indicates cell state (gene 1, gene 2): white (OFF,OFF), yellow (ON,OFF), blue (OFF,ON), and black (ON, ON). A) Both I's are 0, a random grid. B) $I_1 = 0.74$, $I_2 = 0.01$, only cells with gene 1 switched on cluster together. C) $I_1 = 0.06$, $I_2 = 0.74$, only cells with gene 2 switched on cluster together. D) $I_1 = 0.74$, $I_2 = 0.70$ cells with gene 1 switched cluster together, and cells with gene 2 switched on cluster together.

$$\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}} \tag{C.2}$$

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}} \tag{C.3}$$

$$\text{Specificity} = \frac{\text{TN}}{\text{TN} + \text{FP}} \tag{C.4}$$

The AUC is the area under the curve of the receiver operating characteristic (ROC) plot (Fawcett, 2006).

TABLE C.1: Confusion matrix of the predictions of the logistic classifier. The predictions where done for the simulations the model was fitted to (the training set) The rows correspond to what the classifier predicts (TW or not), the columns the actual situation. The numbers indicate the number of simulations in each category.

	Actual Travelling Wave	Actual absence Travelling Wave
Predicted Travelling Wave	3677	2823
Predicted absence Travelling Wave	725	875

In addition to the logistic classifier above that predicts whether or not a TW will be formed a linear model was used to perform a similar task. Instead of predicting whether a given combination of initial I values results in a TW, however, the linear model outputs how much time it will take to form this TW. The linear model was fitted to the data obtained form the same simulations as the logistic model.

Fitting logistic regression models

```
Mlog0 = glm(Wave ~ 1, family = 'binomial',data = M_wave) #null model, mean fraction wave
Mlog = glm(Wave ~ I1_0 + I2_0, family = 'binomial',data = M_wave) # both I's
Mlog_inter = glm(Wave ~ I1_0 + I2_0 + I1_0:I2_0, family = 'binomial',data = M_wave) # including interaction

anova(Mlog0,Mlog,Mlog_inter,test = 'Chisq')
```

```
## Analysis of Deviance Table
##
## Model 1: Wave ~ 1
## Model 2: Wave ~ I1_0 + I2_0
## Model 3: Wave ~ I1_0 + I2_0 + I1_0:I2_0
##   Resid. Df Resid. Dev Df Deviance Pr(>Chi)
## 1      8099      11168
## 2      8097      11088  2   79.230  < 2e-16 ***
## 3      8096      11084  1    4.066  0.04376 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

FIGURE C.3: Output generated comparing the different logistic models as given by R.

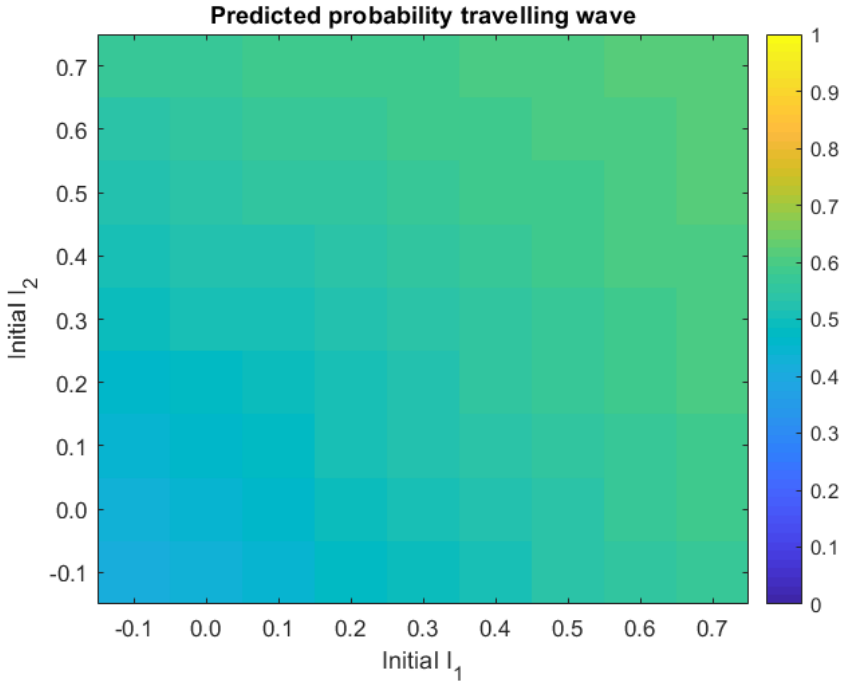


FIGURE C.4: Heatmap of the probability of forming a travelling wave (indicated by colour) given the initial I values as predicted by the logistic model that was fitted to the I data.

Figure C.6 shows that simulations with high initial I values require less time steps to form a travelling wave than simulations with lower initial I values. It should be noted that the figure only shows the mean and does not indicate the variance, which is large in all cases.

The results from the statistical analysis are similar to those of the logistic model. The linear model that includes both I terms was significantly better than the null model. However, the drop in the residual sum of squares (RSS) between the models is very small (fig C.7). This indicates that the I terms in the more complex model do not increase the amount of explained data much. This is also reflected in the (adjusted) R^2 value of the model: 0.006337. This means there is a lot of variation in the time required to form a travelling wave that cannot be explained by the I values alone.

Parameters

Parameters used for simulations: genetic circuit = 19, gridsize = 15, K = [50.4903 232.9002; 319.2420 0], Con = [767.5744 483.3084], p0 = [0.5 0.5], N = 225.

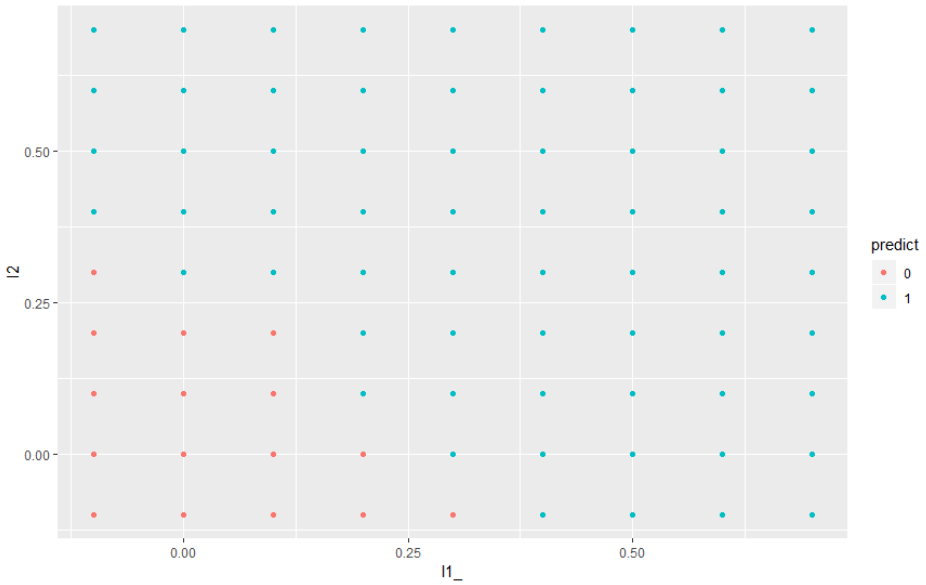


FIGURE C.5: Results of using the logistic model as a classifier. For each input ($[I^{(1)}, I^{(2)}]$) the model predicts whether an initial states with these I-values results in a travelling wave (1, turquoise) or not (0, orange).

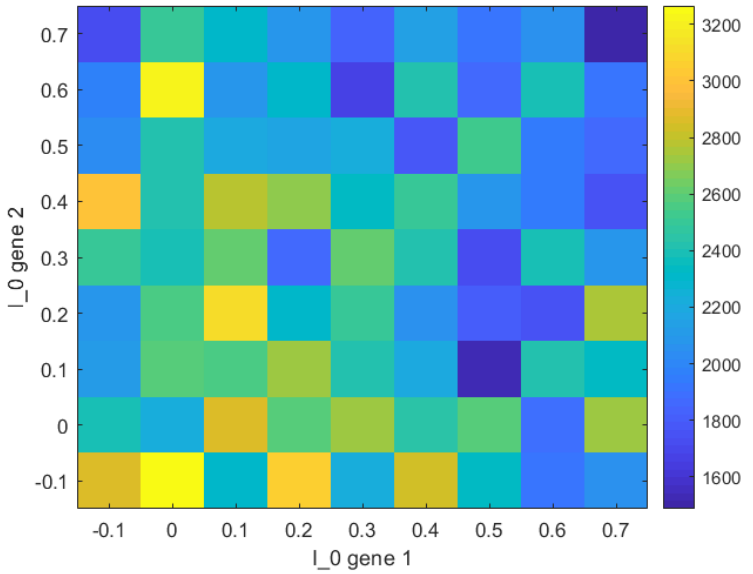


FIGURE C.6: Heatmap showing how the initial spatial order influences wave formation. Only simulations that yielded a wave (4402) are plotted. Colour indicates the mean number of time steps required to form a wave.

```
lm0 <- lm(t.wave ~ 1, data = t_wave)
lm1 <- lm(t.wave ~ I1_0, data = t_wave)
lm2 <- lm(t.wave ~ I2_0, data = t_wave)
lm3 <- lm(t.wave ~ I1_0 + I2_0, data = t_wave)
lm4 <- lm(t.wave ~ I1_0 + I2_0 + I1_0:I2_0, data = t_wave)

anova(lm0,lm1,lm3,lm4)
```

```
## Analysis of Variance Table
##
## Model 1: t.wave ~ 1
## Model 2: t.wave ~ I1_0
## Model 3: t.wave ~ I1_0 + I2_0
## Model 4: t.wave ~ I1_0 + I2_0 + I1_0:I2_0
##   Res.Df        RSS Df Sum of Sq       F    Pr(>F)
## 1   4401 2.6840e+10
## 2   4400 2.6745e+10  1  94758946 15.6334 7.809e-05 ***
## 3   4399 2.6658e+10  1  87436404 14.4253 0.0001478 ***
## 4   4398 2.6658e+10  1    245284  0.0405 0.8405794
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

FIGURE C.7: Output generated comparing the different linear models as given by R.

C.2.2 Dominant initial orientation

The initial dominant orientation was used to predict orientation of the TW. The results are displayed in tables C.2 and C.3. From these so called Confusion Matrices the recall, precision, and specificity are calculated, shown in tables C.4 and C.5, as an indication of how well the TW orientation is predicted from the dominant initial orientation.

The simulationset in which p was varied (table C.4) has recall values that correspond to a random prediction (0.333). The high precision value for the horizontal orientation is largely a consequence of the majority of waves consisting of horizontal ones, consequently the precision for the other orientations are lower. The specificity values are also around random, which is 0.666 in this case, because the True Negatives consist of the other two classes. The overall accuracy was 0.343262411, again implying random performance.

The simulationset in which I was varied (table C.5) has a high recall value for the prediction of horizontal waves, but a very low one for vertical waves. Again, the precision score are largely a consequence of the majority of the waves consisting of horizontal ones. Futhermore, the majority of the vertical TW in this simulation set also had a dominant initial horizontal orientation, this was also the case for the p simulations, but here the difference is far greater. The overall accuracy was 0.601544752. For reliable predictions all wave orientations need to be predicted well. This is not the case here, because vertical waves are not predicted correctly in the majority of cases. Furthermore, there are no diagonal waves detected whatsoever. The high amount of horizontal waves could have to do with the larger number of highly clustered initial states that are present in this simulationset, but this would need to be investigated further to draw that conclusion.

Interestingly, the orientation type that is most abundant, for both simulation sets and in both in the dominant initial orientation and the actual TW and in both simulations sets, is horizontal > vertical > diagonal. I do not know why this happens. Investigating this in more detail may reveal more about the conditions under which TWs form.

TABLE C.2: Confusion Matrix of simulations in which initial fraction of active genes (p) was varied. Each row represents an initial orientation (the prediction), each column the actual orientation of the travelling wave. The number indicate the number of cases for each instance. For example, there were 484 simulations that were predicted to have a vertical TW, but the TW was actually horizontal. The diagonal TWs found in this simulation are not true TW, because they are not spatial symmetrical, nevertheless they were included.

	Horizontal (TW)	Vertical (TW)	Diagonal (TW)
Horizontal (init)	505	193	34
Vertical (init)	484	176	27
Diagonal (init)	480	171	45

TABLE C.3: Confusion Matrix of simulations in which initial spatial clustering (I) was varied. Each row represents an initial orientation (the prediction), each column the actual orientation of the travelling wave. The number indicate the number of cases for each instance. For example, there were 342 simulations that were predicted to have a vertical TW, but the TW was actually horizontal. No diagonal waves occured in this simulationset.

	Horizontal (TW)	Vertical (TW)	Diagonal (TW)
Horizontal (init)	2404	1208	0
Vertical (init)	342	244	0
Diagonal (init)	123	81	0

Parameters

The initial dominant spatial orientation was calculated for all TWs of the simulations were p was varied and all simulations for which I was varied. Therefore, the parameters are in sections C.1.1 and C.2.1.

TABLE C.4: Recall,Precision, and specificity for the simulationsset in which the fraction of initially active genes (p) was varied.

	Horizontal	Vertical	Diagonal
Recall	0.343771	0.325926	0.424528
Precision	0.689891	0.256186	0.064655
Specificity	0.648606	0.675555	0.675958

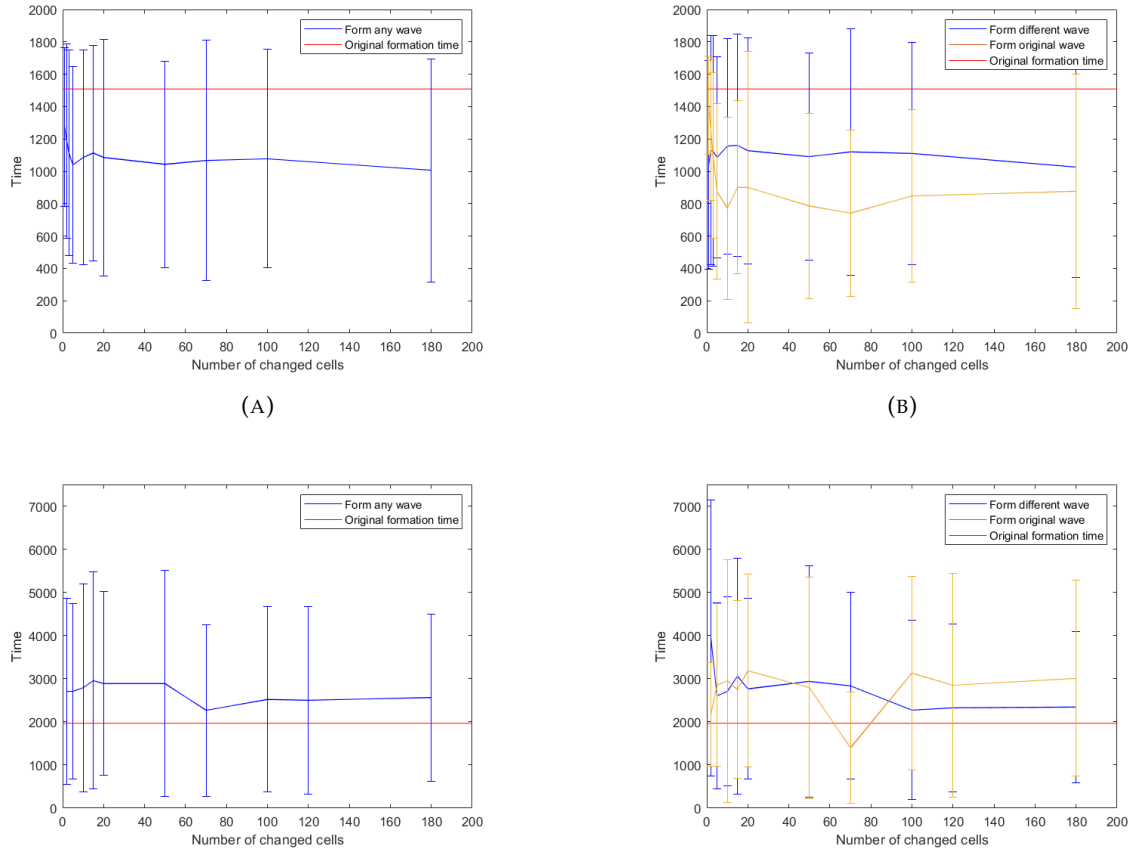
TABLE C.5: Recall,Precision, and specificity for the simulationsset in which the amount of initial spatial clustering (I) was varied. No diagonal waves were detected in this simulationset.

	Horizontal	Vertical	Diagonal
Recall	0.837923	0.159165	-
Precision	0.665559	0.4163823	-
Specificity	0.212002	0.880794	-

C.3 Results concerning perturbation experiments

C.3.1 Perturbation experiments investigating fraction TW

Figure C.8 and C.9 show the results on the time required to (re)form a wave after perturbation of the initial state and of a wave respectively.



(A) (B)

(C) (D)

FIGURE C.8: Time steps required to (re)form a wave after perturbation of an initial state that yielded a wave for an increasing number of changed cells in genetic circuit 15 (a,b) and circuit 33(c,d). a,c) The mean number of time step required to form a travelling wave of any orientation and direction (blue). b,d) The blue line shows the mean number of time steps required to form a wave with a different orientation than the wave that was originally formed. The orange line shows the same, but then for a wave with the same orientation and travel direction as the original wave. In all cases the red line indicates how many time steps were required to form the original wave from the initial condition. The error bars show the standard deviation.

Parameters

The parameters for circuit 33 are identical to section C.1.1 with the addition that p = [0.5 0.5]. For circuit 15: K = [0 22; 11 4], Con = [18 16], I0 = [0 0], p0 = [0.4 0.4], N = 225.

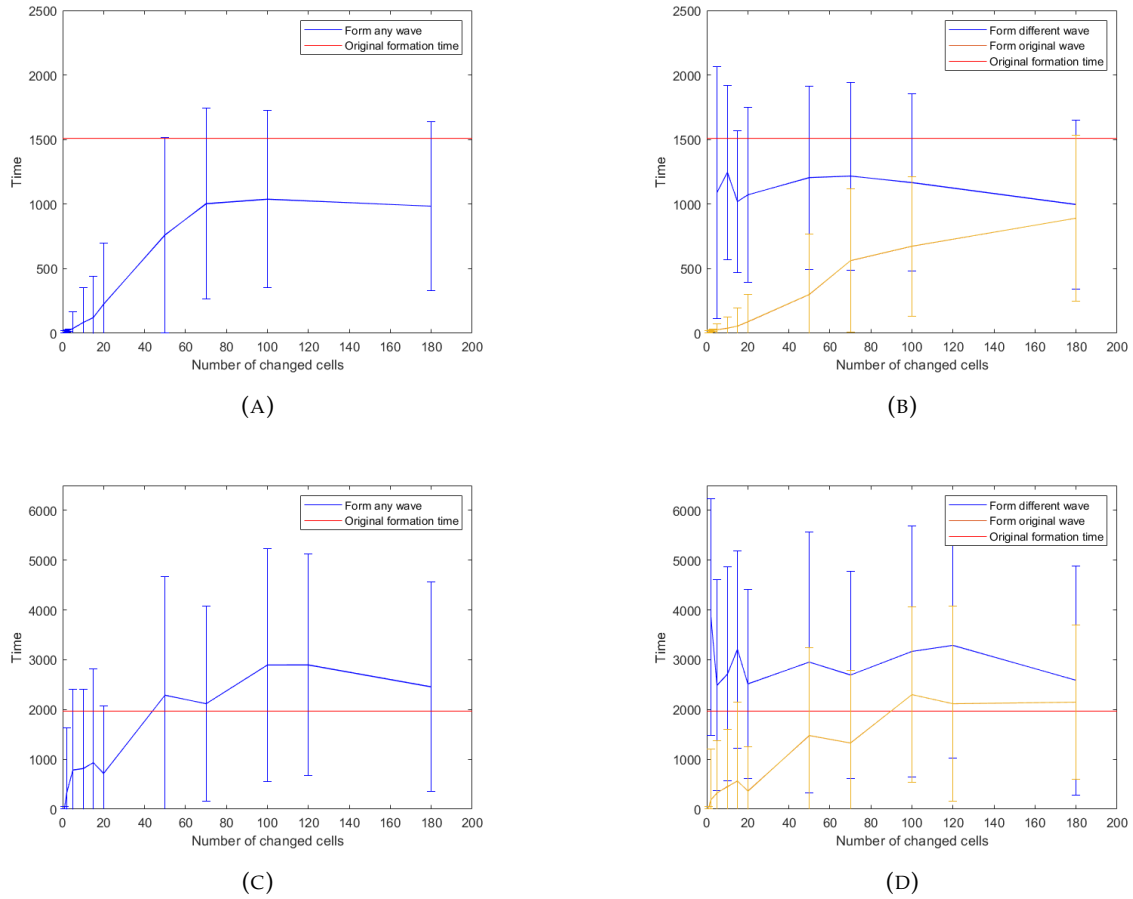


FIGURE C.9: Time steps required to (re)form a wave after perturbation of an already formed wave for an increasing number of changed cells in genetic circuit 15 (a,b) and circuit 33(c,d). a,c) The mean number of time step required to form a travelling wave of any orientation and direction (blue). b,d) Here the blue line shows the mean number of time steps required to form a wave with a different orientation than the wave that was perturbed. The orange line shows the same, but then for a wave with the same orientation as the perturbed wave. In all cases the red line indicates how many time steps were required to form the original wave from the initial condition. The error bars show the standard deviation.

C.3.2 Perturbation experiment comparing time trajectory

In order to establish whether there had been a shift in trajectories, the first 81 timesteps (or the the maximal number of time steps in case of shorter simulation) of the perturbed simulation were compared to the first 301 time steps of the unperturbed simulation. For each comparison a similarity score was computed by comparing all cells of both grids to each other: +1 if a corresponding cell (cell with same coordinates) was in the same state -1 if not. Hereby a scoring matrix with the dimensions (81,301) was obtained for each perturbed simulation. For each row in this matrix (corresponding to a time step in perturbed simulation) the maximum score is set to 1 if, and only if, this score > 0.5 N (number of cells), the other scores are set to 0. Now all values on the diagonal can be summed and divided by the sum of all matrix elements. If there has been a time shift, this fraction is expected to be low. If not, the fraction should be higher. The mean fraction across all 600 simulations was: 0.9963 ± 0.0190 ($mean \pm std$), indicating that no time shift occurred.

Parameters

Parameters used for simulations: genetic circuit = 15, gridsize = 15, K = [0 22;11 4], Con = [18 16], p0 = [0.4 0.4], N = 225.

C.4 Results concerning Spatial Symmetry Score

Figure C.10 shows that the Spatial Symmetry Score (SSS) for randomly ordered grids roughly has values between 0.4 and 0.6.

Figure C.11 shows that the system from figure 3.12a indeed resembles a wave at the local minimum SSS.

Figure C.12 shows that the different variants of the SSS show similar results qualitatively speaking. The variant that deviates most is the comparison of state (1,1) at time t with state (1,1) when the TW has formed. This score does for instance not always drop when the other do (fig C.12c), indicating that the transient wave like structure is differently orientated than the final TW.

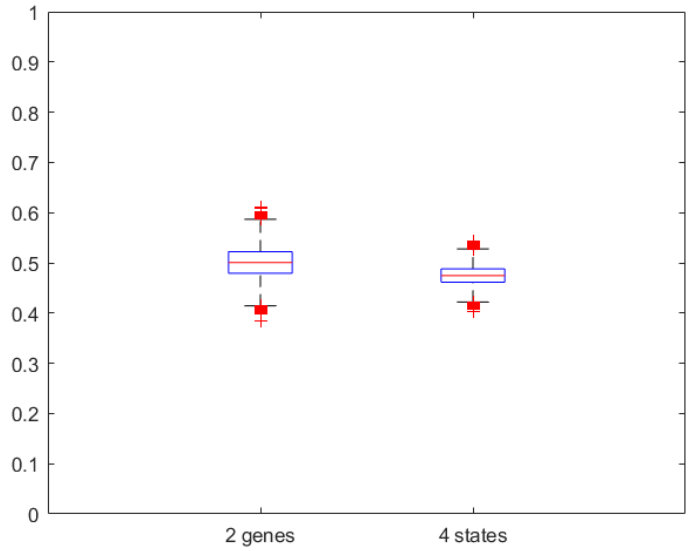


FIGURE C.10: Boxplots of the Spatial Symmetry Score, both the 2 genes and 4 states version, of 10000 randomly generated grids of cells.

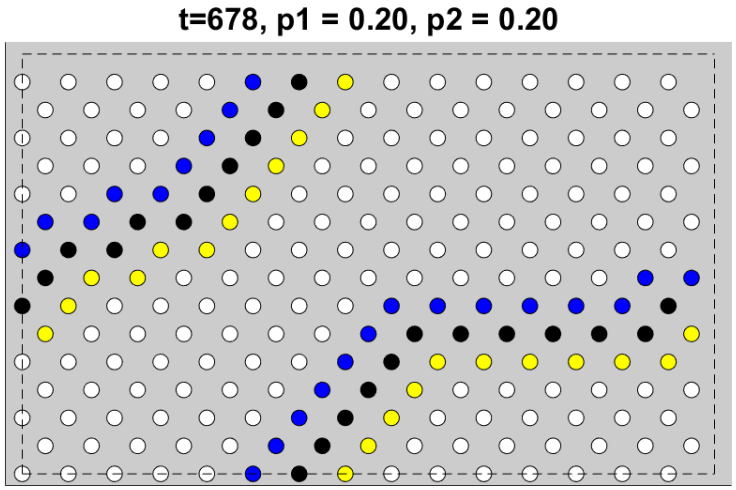
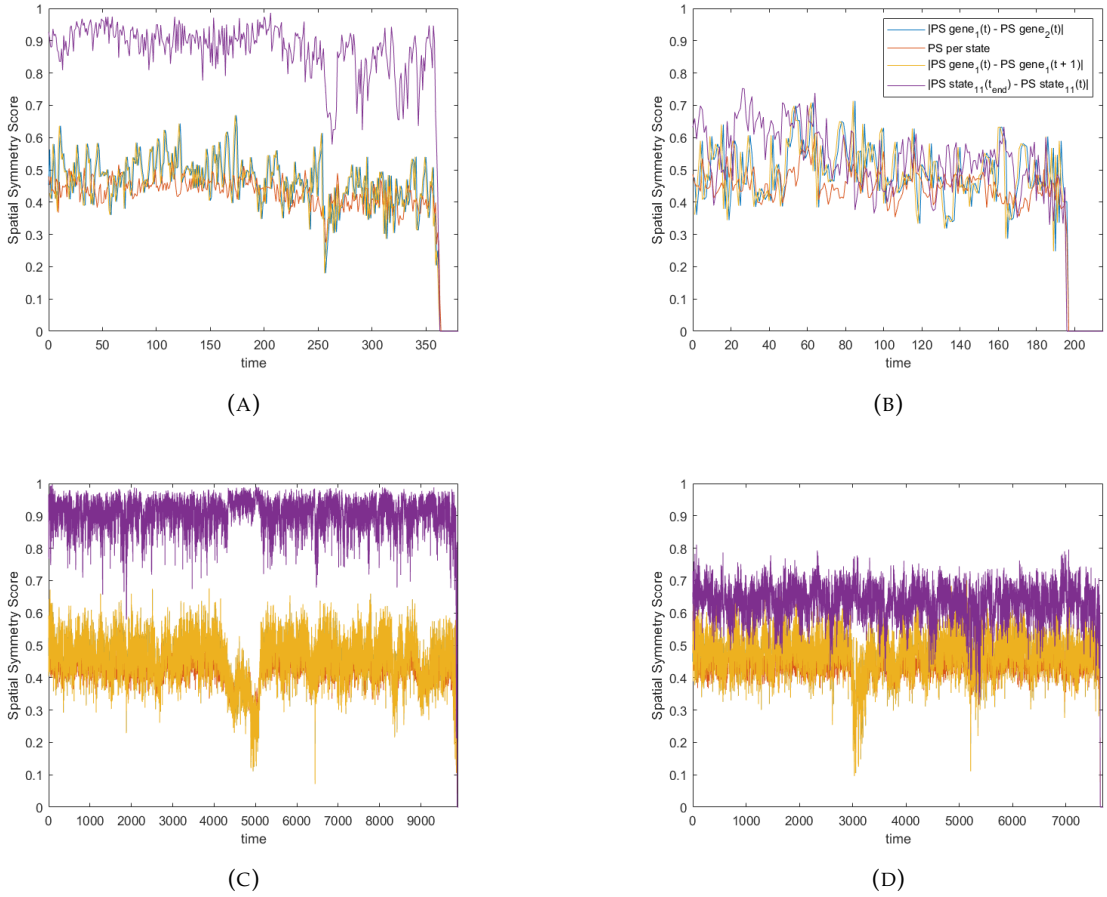


FIGURE C.11: System from figure 3.12a in the local minimun in SSS at time 678.



(A) (B)

(C) (D)

FIGURE C.12: Four different variants of the Spatial Symmetry Score (SSS) are calculated for four different simulations (A-D). Each colour indicates a variant of the SSS, as indicated in the legend in B.

C.5 Results concerning finite Hill

The simulations initiated with a TW were done for a large set of parameters. These parameters were a subset from a large set of parameters obtained over a large part of parameter space (using latin-hypercube sampling). The subset of 2534 parameter sets for which TWs could propagate for an infinite Hill coefficient (binary cell states). One simulation was run for each parameter set.

The simulations with random initial conditions were for a single set of specific parameters. 150 distinct initial conditions were used.

For all finite Hill simulations rounding of numeric values is very important. Here, cell states have been rounded to 6 decimal numbers. This because numerical errors were observed if this was not the case (breaking of symmetric structure due to accumulating of small numerical errors over time). Another example is when grids should be called homogeneous. I decide this based on the standard deviation. If it is small enough the grid is called homogeneous, even if the cells have slightly different values. Based on observations of the simulation the std border was set to 0.02 for a system of 225 cells.

C.5.1 Parameters

The parameter set used for the randomly initiated simulations: genetic circuit = 15, gridsize = 15, K = [0 9; 11 4], Con = [18 16], p0 = [0.5 0.5], N = 225.

Bibliography

Dang, Y., Grundel, D. A. J., and Youk, H. (in prep). Insert title. *Undetermined journal*.

Deneke, V. E. and Di Talia, S. (2018). Chemical waves in cell and developmental biology. *Journal of Cell Biology*, 217(4):1193–1204.

Deneke, V. E., Melbinger, A., Vergassola, M., and Di Talia, S. (2016). Waves of Cdk1 Activity in S Phase Synchronize the Cell Cycle in Drosophila Embryos. *Developmental Cell*, 38(4):399–412.

Economou, A. D., Ohazama, A., Porntaveetus, T., Sharpe, P. T., Kondo, S., Basson, M. A., Gritli-Linde, A., Cobourne, M. T., and Green, J. B. A. (2012). Periodic stripe formation by a Turing mechanism operating at growth zones in the mammalian palate. *Nature Genetics*, 44(3):348–351.

Fawcett, T. (2006). An introduction to roc analysis. *Pattern recognition letters*, 27(8):861–874.

Gierer, A. and Meinhardt, H. (1972). A Theory of Biological Pattern Formation. *Kybernetik*, 39:30–39.

Gomez, M. M. and Arcak, M. (2017). A Tug-of-War Mechanism for Pattern Formation in a Genetic Network. *ACS Synthetic Biology*, 6(11):2056–2066.

Green, J. B. A. and Sharpe, J. (2015). Positional information and reaction-diffusion: two big ideas in developmental biology combine. *Development*, 142(7):1203–1211.

Hemelrijk, C. K., van Zuidam, L., and Hildenbrandt, H. (2015). What underlies waves of agitation in starling flocks. *Behavioral Ecology and Sociobiology*, 69(5):755–764.

Huang, C.-Y. and Ferrell, J. E. (1996). Ultrasensitivity in the mitogen-activated protein kinase cascade. *Proceedings of the National Academy of Sciences*, 93(19):10078–10083.

Johnson, H. E., Goyal, Y., Pannucci, N. L., Schüpbach, T., Shvartsman, S. Y., and Toettcher, J. E. (2017). The spatiotemporal limits of developmental erk signaling. *Developmental cell*, 40(2):185–192.

Kondo, S. and Miura, T. (2010). Reaction-Diffusion Model as a framework for understanding biological pattern formation. *Science (New York, N.Y.)*, 329(September):1616–1620.

Loose, M., Kruse, K., and Schwille, P. (2011). Protein Self-Organization: Lessons from the Min System. *Proceedings of the Annual Conference of the International Speech Communication Association, INTERSPEECH*, pages 901–904.

Lucchetta, E. M., Lee, J. H., Fu, L. A., Patel, N. H., and Ismagilov, R. F. (2005). Dynamics of drosophila embryonic patterning network perturbed in space and time using microfluidics. *Nature*, 434(7037):1134.

Maire, T. and Youk, H. (2015). Molecular-Level Tuning of Cellular Autonomy Controls the Collective Behaviors of Cell Populations. *Cell Systems*, 1(5):349–360.

Marcon, L. and Sharpe, J. (2012). Turing patterns in development: What about the horse part? *Current Opinion in Genetics and Development*, 22(6):578–584.

Menshykau, D., Kraemer, C., and Iber, D. (2012). Branch mode selection during early lung development. *PLoS Computational Biology*, 8(2).

Moran, P. A. (1950). Notes on continuous stochastic phenomena. *Biometrika*, 37(1/2):17–23.

Olimpio, E. P., Dang, Y., and Youk, H. (2018). Statistical dynamics of spatial-order formation by communicating cells. *iScience*, 2:27–40.

Page, K. M., Maini, P. K., and Monk, N. A. (2005). Complex pattern formation in reaction–diffusion systems with spatially varying parameters. *Physica D: Nonlinear Phenomena*, 202(1-2):95–115.

Plotnikov, A., Zehorai, E., Procaccia, S., and Seger, R. (2011). The mapk cascades: signaling components, nuclear roles and mechanisms of nuclear translocation. *Biochimica et Biophysica Acta (BBA)-Molecular Cell Research*, 1813(9):1619–1633.

Raj, A. and van Oudenaarden, A. (2008). Nature, Nurture, or Chance: Stochastic Gene Expression and Its Consequences. *Cell*, 135(2):216–226.

Trunnell, N. B., Poon, A. C., Kim, S. Y., and Ferrell, J. E. (2011). Ultrasensitivity in the Regulation of Cdc25C by Cdk1. *Molecular Cell*, 41(3):263–274.

Turing, A. M. (1952). The chemical basis of morphogenesis. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, 237(641):33–72.

Weiss, J. N. (1997). The Hill equation revisited: uses and misuses. *FASEB Journal*, 11(11):835–841.