D.S. Laman Trip

Simulation of tissue folding with the Cellular Potts Model

Bachelor Thesis, 10 July 2013

Supervisors: dr. R.M.H. Merks, dr. S.S. Urdy

Mathematical Institute, Leiden University

Abstract

Villi formation in the intestine is an example of tissue folding and a poorly understood mechanism. The Cellular Potts Model is a computational model to simulate cellular structures. A mathematical model for epithelial folding using the Cellular Potts Model was implemented to analyze the behaviour of a three layered tissue for different cell properties. The centers of mass and the forces on the cells were used to analyse the simulation results. The forces on the cells are higher during simulations than at the end, and higher growth ratios result in larger forces. Various relations are found between homotypic bond energies, resistance to compression or growth increments of the cells and the amplitude, frequency or size of the tissue. Resistance of the medium and the behaviour of the boundary cells is responsible for epithelial folding and only a small part of the parameter space yields tissues that have a wavelike shape and unbroken cell layers.

Contents

1 Introduction

The mammelian digestive system is a complex system of organs including the stomach and intestines. The large and small intestine contain cells that are responsible for absorbtion and secretion processes of the gut. Villi and crypts form small finger-like shapes and invaginations in the intestinal wall which increases the surface area of the intestines and provides more surface area for cells to absorb or secrete water and nutrients.

Formation of the small intestine is a complex mechanism that requires strict genetic regulation. The intestine forms during fetal development starting with gastrulation. The blastula forms a three layered structure, one of which is the endoderm that in human, chick and mouse consists of a single sheet of epithelium cells. The endoderm forms a gut tube that gives rise to the lungs, liver and intestines by tissue rearrangements and patterning after gastrulation. The endoderm proliferates after the tissue reorganizations but before villi formation. Villi are small projections from the surface of the intestine formed by epithelial folding that together with crypts, which are tubular invaginations of the epithelium, form a repeating pattern in the intestine. Villi formation with increased proliferation at the base and crypt development, where crypt-villus junctions migrate towards the intestinal cavity, creates characteristic patterning in the intestines, after which proliferation becomes limited to the crypts. (Spence et al., 2011).

Villi and crypt formation in the developing intestine is an example of tissue folding. Tissue folding is a change in the structure of tissues resulting from morphogenesis, which is the development of the shapes of organisms caused by biomechanical forces that lead to tissue deformations (Wyczalkowski et al., 2012). The biomechanical forces are the result of for example gene-expression, chemical gradients, material properties, the cytoskeleton, stress and adhesion molecules (Mammoto et al., 2010).

Various quantitative theories based on biomechanical processes exist for morphogenetic mechanisms. Turing (1952) proposed reactions between morphogens as a theory for morphogenesis, which has limited applicability. Another theory, the Differential Adhesion Hypothesis (DAH) introduced by Steinberg (1963), explains morphogenesis using differences in cellular adhesion, supported and validated by experimental results (Wyczalkowski et al., 2012).

Increasing computing power and extensive amounts of available experimental data provide oppertunities to investigate morphogenesis with computational models in a new level of detail. Computational models are never perfect, and cannot include all biological and mechanical systems in an organism, but can help to give insight and understanding in biological mechanisms (Wyczalkowski et al., 2012).

Formation of villi is essential for a functional intestine, and a lot of signaling pathways are involved in the formation of villi during the development of the intestine, but the process is pooly understood (Spence et al., 2011). Computer simulations of a mathematical model help to understand the mechanisms of folding and curving of tissues, and can lead to insight of the complex process of villi formation in the intestine. A model for epithelial folding and villi formation can eventually be relevant for research of colon cancers and biological tissue shapes. In this thesis the behaviour of tissue layers is investigated with a mathematical model based on the DAH, and the influence of various cell properties on the curving and folding of tissue layers is explored.

Various models exist for epithelial folding, such as a biological model from Leptin and Grunewald (1990) for invagination of the mesoderm in *Drosphila* with regulated gene activity. The experiments do not support previous theories of epithelial folding caused by cell growth and devision but suggests it is arranged by more than one mechanism, including multiple transcription factors.

A physical model for behaviour of tissues from Forgacs et al. (1998) is based on mechanical properties. The experimental data provides evidence that long time behaviour of embryonic tissue has liquid-like properties and supports the DAH.

Hannezo et al. (2012) introduced a mathematical model for epithelial folding based on the buckling instability of a three layered tissue. The tissue consists of a single layer of dividing epithelial cells with a continuous basement membrane and elastic stroma. Cell proliferation and death lead to stresses on the environment and buckling occurs at a critical surface tension. The model predicts many structures observed in the intestines and suggests a relation between the tissue shape and stresses caused by cell devision.

Drasdo et al. (2000) introduced a model based on the physical properties and interactions of cells. The orientation and shape of cells depends on the energy of a cell configuration based on the interaction, bending and rotational energy. This bending energy is equivalent to the basement membrane force introduced in the vertex model from Dunn et al. (2012). That model proposes growing epithelial cells, with the basement membrane (ECM) being the layer between stroma and a monolayer of epithelial cells. A basement membrane force that stabilizes the epithelial layer characterises the strength of adhesion between different layers and the stiffness of the membrane as shown in Fig. 1.

(a) ***(b)***

Figure 1: Simulation results for different basement membrane forces (adopted from Dunn et al., 2012). *Epithelial cells (yellow) are connected to non-proliferating stroma (green). (a) shows the configuration for a low basement membrane force, and (b) for a high membrane force. A higher force results in a stiffer basement membrane and a flat epithelial layer.*

In this thesis the Cellular Potts Model (CPM) is used to simulate the folding of a tissue layer. The CPM is an extension of the large-Q Potts model based on the Differential Adhesion Hypothesis, and was introduced by J. Glazier and F. Graner (1993). It is a mathmatical model to simulate cell rearangements, and can be used to simulate behaviour of tissues. Properties of cells, such as growth, adhesion and resistance to compression can be given an explicit value, and easily be changed.

A cluster cell is a cell devided into multiple subcells that each form a characteristic part of the cluster. The model of epithelial folding with the CPM simulates a layer of cluster cells. Each cluster cell consists of three individual cells, and the tissue can be interpreted as three layers of cells equivalent to the epithelial cells, basement membrane and stroma cells from Dunn et al. (2012). The assumption for folding of the tissue layer is differential growth of different cell layers opposed to the experimental results from Leptin and Grunewald (1990) but in agreement with the models from Hannezo et al.

(2012) or Dunn et al. (2012). Differential growth results in different cell volumes and creates stress and pressure in the tissue layer, which leads to buckling in the tissue.

This implementation differs from existing models because it decouples adhesion and cell elasticity. These parameters are contained in a single spring constant in the models used by Drasdo et al. (2000) and Dunn et al. (2012). The adhesion parameters compose the basement membrane force instead of a direct implementation by a parameter in the Cellular Potts Model. There are no elastic links between the cells or overdamped springs as in Drasdo et al. (2000) or Dunn et al. (2012), and the cell configurations are lattice based opposed to the spatial approach of other models.

2 Methods

The CPM simulates a squared lattice and discretizes continuous cells on it. Each of the N cells has a spin number $\sigma \in \{1, .., N\}$ and a cell type $\tau(\sigma)$. Each lattice position $\vec{x}$ has a spin $\sigma(\vec{x})$ representing cell σ, giving each cell multiple lattice sites. Every configuration of the cell latice has a corresponding total energy represented by the Hamiltonian (Glazier et al, 1993). In a step of the simulation a random lattice site $\vec{x}$ is selected and given a copy attempt of its spin number $\sigma(\vec{x})$ to a randomly selected neighbor $\vec{x}'$. The copy attempt is accepted or rejected depending on the change of the Hamiltonian. A Monte Carlo Step (MCS) is the number of copy attempts that equals the number of lattice sites, and can be interpreted as a unit of time. Various parameters of the model, such as the target volume of the cells, can be changed using additional conditions at the end of every MCS.

The total energy of a configuration is represented by the Hamiltonian. The Hamiltonian H is for neighboring lattice sites $\vec{x}$ and $\vec{x}'$ given by

$$H = \sum_{\vec{x},\vec{x}'} J_{\tau(\sigma(\vec{x})),\tau(\sigma(\vec{x}'))}(1 - \delta_{\sigma(\vec{x}),\sigma(\vec{x}')}) + \sum_{\sigma} \lambda_{\sigma}(a_{\sigma} - A_{\tau(\sigma)}(t))^2 \tag{1}$$

where $J_{\tau,\tau'}$ represents the bond energy between two cell types, λ_{σ} the resistance to compression for cell σ and a_{σ} and $A_{\tau(\sigma)}(t)$ are the current and target cell volume at monte carlo step t. The Kronecker delta is given by

$$\delta_{\sigma(\vec{x}),\sigma(\vec{x}')} = \begin{cases} 1 & \sigma(\vec{x}) = \sigma(\vec{x}') \\ 0 & \sigma(\vec{x}) \neq \sigma(\vec{x}') \end{cases} \tag{2}$$

setting the bond energy for the same spin zero and giving cells only a surface energy at their boundaries. A spin copy attempt $\sigma(\vec{x}) \to \sigma(\vec{x}')$ at temperature $T > 0$ is accepted with probability

$$P(\Delta H) = \begin{cases} e^{-\Delta H/T} & \Delta H > 0 \\ 1 & \Delta H \leq 0 \end{cases} \tag{3}$$

Calculating the change in Hamiltonian is easy and does not require the total Hamiltonian. It only involves the energy terms for the selected lattice site and its neighbors. Whether homotypic or heterotypic bonds are favored can be determined with the surface tensions, defined by

$$\gamma_{\tau,\tau'} = J_{\tau,\tau'} - \frac{J_{\tau,\tau} + J_{\tau',\tau'}}{2} \tag{4}$$

The cells in the tissue are part of cluster cells. A cluster cell w is defined as $C_w(\sigma_1, \sigma_2, \sigma_3)$ with subcells σ_1, σ_2 and σ_3 and cell types $\tau(C_{w,1}), \tau(C_{w,2})$ and $\tau(C_{w,3})$. Bond energies are defined inside and between cluster cells, where

$$J_{\tau(C_{w,i}),\tau(C_{w,i})} \qquad\qquad J_{\tau(C_{w,i}),\tau(C_{w,j})} \tag{5}$$

are the homotypic and heterotypic bond energies inside a cluster w with subcells i, j and

$$J_{\tau(C_{w,i}),\tau(C_{v,i})} \qquad\qquad J_{\tau(C_{w,i}),\tau(C_{v,j})} \tag{6}$$

are the homotypic and heterotypic bond energies between cluster cells v and w with subcells i and j. Different bond energies between cells inside a cluster and between

cluster cells simulates different behaviour between cells and clusters, resulting in compartmentalized cells.

Increasing the target volume $A_{\tau(\sigma)}(t)$ changes the Hamiltonian depending on a_σ and lets the cells grow by making a bigger cell volume energetically favorable depending on λ_σ. Changing the target volume with a different amount for each cell type results in different growth speeds and differential growth between cell types.
λ_σ represents the strength of the volume constraint for celltype τ as mentioned earlier. It can be interpreted as the elasticity or resistance to compression of the cells, and determines how cells adapt to their target volume. The mobility of the cells is also influenced by λ_σ, because a strong volume constraint makes it less energetically favorable for cells to perform a succesful copy attempt that changes the cell volume.
The bond energies $J_{\tau,\tau'}$ determine how strong connections between cells are and has an influence on the cell mobility as well. A high bond energy results in a large Hamiltonian and makes a lower bond energy energetically favorable with a copy attempt. This makes the contact between the cells fragile and results in a weak connection and motile cells. Lower contact energies result in more stationary cells. Different contact energies between cell types result in cells sorting to a configuration with contacts that minimalize the Hamiltonian.

The cells in simulations of the CPM grow and move depending on the change of the Hamiltonian, and cells in biological tissues grow and move according to the biomechanical forces on the cells. The change in energy can be used to express this biomechanical force on the cells in the tissue and relate the simulated cell layer to biological tissues. Work is defined as $W = F \cdot \vec{s}$, the force F on a point that moves distance $|\vec{s}|$ in direction $\frac{\vec{s}}{|\vec{s}|}$. It is also equal to the change in kinetic energy ΔE. Combining both equations expresses the force on a point as the change in energy devided by its displacement: $|F| = \frac{\Delta E}{|\vec{s}|}$. The change in energy for a copy attempt for a lattice site is represented by the change of the Hamiltonian. The copy attempt is accepted if the change of the Hamiltonian is negative, and it represents a positive force in that direction $\Delta E = -\Delta H$. If the change of the Hamiltonian is positive it can be interpreted as a force in the opposite direction $-\frac{\vec{s}}{|\vec{s}|}$. Then the force on a lattice site $\vec{x}$ is given by

$$\vec{F}(\vec{x}) = \sum_{\sigma(\vec{x}) \to \sigma(\vec{x}')} -\frac{\Delta H}{|\vec{s}|} \cdot \frac{\vec{s}}{|\vec{s}|} \tag{7}$$

Only the change of cell type on a lattice position results in a change in Hamiltonian. As a consequence the forces work on the borders of the cells. The forces on every point in the tissue gives insight in the stresses in the layer and help to explain the behaviour of the tissue not only for different parameter sets, but also the change of the tissue layer through time.

2.1 Model initialisation

The introduced model of epithelial folding considers a layer of cluster cells, each of which consists of three individual cells. This gives three connected cell layers surrounded by medium simulated by the CPM, which is implemented in the open source program CompuCell3D [6], a modelling environment for simulations of the CPM. The computational model in CompuCell3D has a XML file, Python scripts and a piff file that contain the description of the model for epithelial folding.

Figure 2: Initial cell lattice. *The tissue contains three cell layers or 33 cluster cells. The top layer (green, t) can be interpreted as epithelial cells, the center layer (blue, c) as the basement membrane, and the bottom cells (red, b) as stroma cells. Other lattice sites are considered medium.*

The initial configuration of the cell lattice is shown in Fig. 2. The layer has free boundary conditions with all cells free to move. The boundary cells exert pressure to the stationary lattice boundaries which simulates the counterpressure of other tissue. The length of the tissue layer is 33 clusters with a total of 99 cells, and the lattice dimensions are 297×297 pixels. The volume of the cells is given by the number of pixels on the cell lattice. The initial cell size is 9×3 pixels with cell volume 27. One simulation of epithelial folding has 100.000 MCS with each 297×297 copy-attempts, with the lattice sites for a copy attempt randomly selected. The theoretical growth ratio for a cluster cell with growth increments $\Delta A_t, \Delta A_c$ and ΔA_b is given by

$$growth(\Delta A_t, \Delta A_c, \Delta A_b) = \frac{3 \cdot 27 + 100000 \cdot (\Delta A_t + \Delta A_c + \Delta A_b)}{3 \cdot 27} \tag{8}$$

The initial random seed on which the outcome of the simulation and the selection of the next lattice sites depends can be used later to reproduce the results. The neigbor order is equal to 2, making spin copy attempt possible to nearest and second-nearest neighbors with a total of eight neigbors for each lattice site.

The temperature of the simulation is equal to $T = 10$ based on Glazier et al. (1993). A parameter set is a point from the parameter space containing the values of the constant and varied parameters. The parameters varied with simulations are the targetvolumes $A_{\tau(\sigma)}(t)$ of the three cell layers depending on celltype and simulation time (three parameters), the resistance to compression λ_σ depending on the cell surface of the cells (one parameter) and the homotypic bond energies $J_{\tau(C_{w,i}),\tau(C_{v,i})}$ of the cell layers (three parameters) depending on the cell types.

Resistance to compression of the tissue is defined as λ_σ and is a property of all cell types. The resistance to compression remains the same for different cell sizes if assumed to be constant. However, a cell with a smaller surface area has less copy attempts each MCS, which results in a lower probability to do a succesful copy attempt to resist compression and makes the cell more likely to lose volume. Then the actual resistance to compression becomes dependent on cell size.

A solution is to use a different λ_σ for different cell sizes. If two cells have surface areas $S_1 > S_2$ they theoretically have S_1, S_2 copy attempts each MCS, with ratio $\frac{S_2}{S_1}$. Define $\lambda_\sigma = \frac{\lambda}{S_i}$. Then the energy change for the volume constraint of one copy attempt

is equal to $\frac{\lambda}{S_i}((A_i - 1) - A_i)^2 = \frac{\lambda}{S_i}$ with ratio $\frac{S_1}{S_2}$. Then the ratio of a succesful copy attempt is approximately $\frac{S_1}{S_2}$ for the two cells. The larger cell has $\frac{S_2}{S_1}$ more copy attempts, but only $\frac{S_1}{S_2}$ as much are succesful which result in an equal total amount of succesful copy attempts for both cell sizes. For the model of epithelial folding the resistance to compression is defined as

$$\lambda_\sigma = \frac{\lambda}{S(\sigma)} \tag{9}$$

with $S(\sigma)$ the surface area of the cells, and λ the parameter used to change the resistance to compression.

The standard parameter values for the parameters varied with the simulations are shown in Table 1. All simulations are assumed to have these standard parameter values with differential growth and the same bond energies, unless stated otherwise. All cell types have bond energies with the medium and other bond energies that are also assumed to be constant during simulations (Details in section 5.1).

Table 1: *Standard parameter values of homotypic bond energies for top (t), center (c) and bottom (b) cells, resistance to compression and growth increments.*

Parameter	Value
$J_{t,t}$	40
$J_{c,c}$	40
$J_{b,b}$	40
λ	400
ΔA_t	0.0008
ΔA_c	0.0006
ΔA_b	0.0004

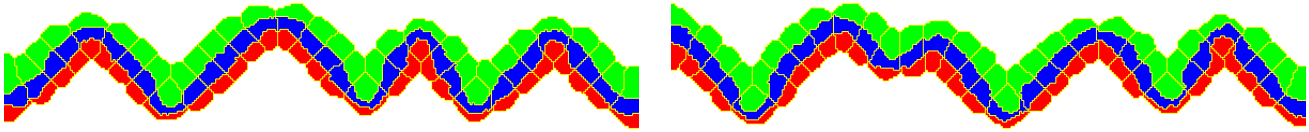
2.2 Parameter studies

Investigating the influence of cell properties on the tissue layer requires a vast amount of simulations and each parameter set must be run several times with different random seeds. Python scripts were written that automatically generate and run simulations using a template simulation (XML, Python and piff files, details in section 5.2) to generate and run all the simulations for selected parameter sets.

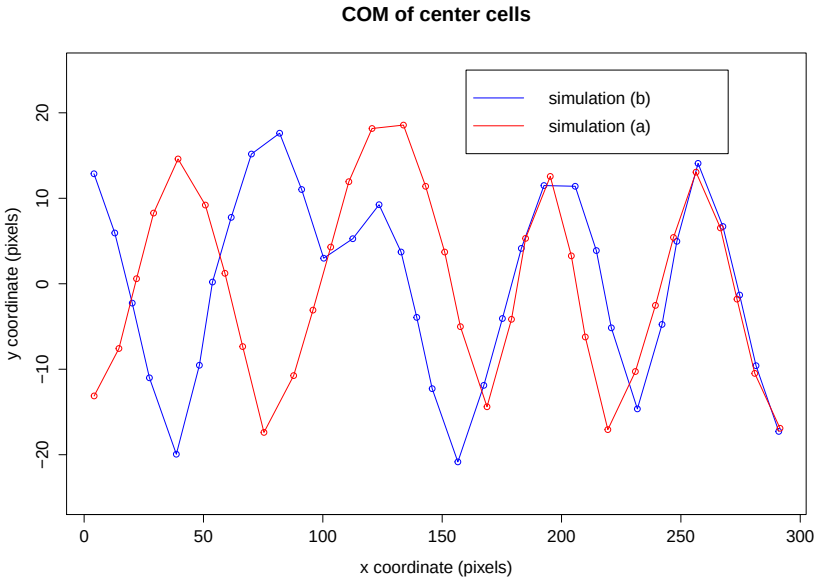
There are multiple variables available for analyzation after the simulations. The data that has been stored during the project are the piff files for the cell lattice that can be used to reconstruct the tissue layer at the end of the simulation. A Python script was made that converts the piff file of the cell lattice to an image. Also the Center Of Mass (COM) of all the cells, their average per cluster cell, the growth of the cells, the integrity of the tissue and the forces on the cell lattice are stored. All data is stored in a format suited to use with the statistical package R. The analysis of a combination of different data leads to insight in the behaviour of the tissue layer.

2.3 Preliminary results

An example of a cell lattice at the end of a simulation is shown in Fig. 3(a). Fig. 3(b) shows the cell lattice for the same parameter set but a different random seed. Because the parameters of the simulations are identical, the behaviour of the tissue is expected to be similar as well. Fig. 2(c) shows the COM of the center cells for both tissue layers. The positions of the COM are cosiderably different for x-coordinates $x < 150$, leading to difficulties in directly analyzing the COM. A Fourier transform relates data from a time domain to the frequency domain, which can be applied to the COM because the shape of the tissue at the end of a simulation is generally a wave.



(a) *(b)*



(c)

Figure 3: Cell lattice at 100.000 MCS for simulations with different random seeds. *The parameter set of the simulations is the same but the shape of the tissue is different. (a) and (b) show the cell lattice, (c) shows the center of mass of the center cells for both simulations.*

2.4 Morphometry

The Discrete Fourier Transform (DFT) transforms an equally spaced finite data sample to the coefficients of complex sinusoids. However, the COM of the cells are planar, not equally spaced measurements. A COM (x_i, y_i) can be interpreted as a measurement y_i at time x_i, or as unequally spaced measurements. These measurements are succesive in time, because the continuity of the tissue is lost in the case that $x_i > x_{i+1}$. The transform suited for this data is the Non-Uniform Discrete Fourier Transform (NDFT), that accepts unequally spaced data samples and returns the coefficients of the complex sinusoids just as the DFT.

The NDFT returns the amplitude and phase-shift of the fundamental frequency f_{fund} and its harmonics. The fundamental frequency is the lowest frequency sinusoid, and has frequency $f_{\text{fund}} = 1/N$ for $N = 33$ cells in the model of epithelial folding. The harmonics are integer multiples m of the fundamental frequency, with frequency m/N. The NDFT for data $\{(x'_j, f_j) : j \in \{0, .., N-1\}\}$ is defined as

$$F(k) = \frac{1}{N} \sum_{j=0}^{N-1} f_j e^{-ikx'_j} \tag{10}$$

with $k = -\frac{N}{2}, ..., \frac{N}{2} - 1$ and $x'_j \in [0, 2\pi]$ (Greengard et al., 2004). Let (x_j, y_j) be the center of mass of a cell in the tissue layer with x_j and y_j in pixels, and define

$$(x'_j, f_j) = (2\pi \cdot \frac{x_j}{x_{max}}, y_j) \tag{11}$$

with $x_{max} = 297$ pixels the length of the cell lattice. Then the NDFT for given centers of mass $\{(x_j, y_j) : j \in \{0, .., N-1\}\}$ of the tissue layer is

$$F(k) = \frac{1}{N} \sum_{j=0}^{N-1} y_j \cdot e^{-2\pi ik \cdot \frac{x_j}{x_{max}}} \qquad k \in \{0, ..., N-1\} \tag{12}$$

The amplitude and phase-shift of the fundamental frequency and its harmonics can be derived from the coefficients $F(k)$ of the complex sinusoids with frequencies k. The amplitude of the frequencies is used to analyze the tissue. The frequencies of the Fourier transform eliminate phase-shifts, translations and reflections from the data. Phase-shifts, translations and reflections in tissue can be caused by different random seeds, even with the same parameter set, and are not relevant for analysis. The amplitude of the frequencies is calculated with the modulus of the complex coefficients $|F(k)|$. The first coefficient is the mean of the y-coordinates, which is moved to zero before the Fourier transform.

Three centers of mass and their avarage are available for analysis from each cell cluster. Therefore there are four different Fourier transforms when analyzing the COM of cells in the tissue layer. Ten simulations with different random seed have been generated and run, and the average amplitude of the frequencies of the Fourier transform of all the four possible centers of mass is shown in Fig. 4.

The Fourier transform of the low frequencies is roughly the same for all four centers of mass, and the transform for the center cells and the average is roughly the same for all

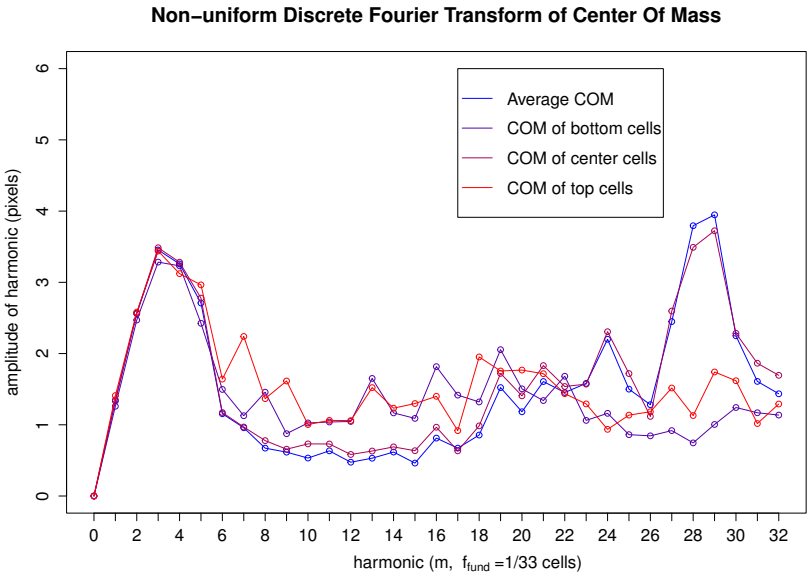


Figure 4: NDFT of COM of ten simulations at 100.000 MCS. The harmonics of the fundamental frequency are plotted against the amplitude of the harmonics in pixels. *The average amplitude of the frequencies of the COM and that of the center cells almost overlap, with minimal p-value 0.876 for the used frequencies. The top and bottom layer are different from the center layer and the average, but cosiderably alike for relevant frequencies.*

frequencies. The middle frequencies of the top and bottom cells have higher amplitude than the middle frequencies of the center cells, which makes it harder to determine the frequencies relevant for the shape of the tissue. The average COM is used for analysis because the position of all three cells in a cluster are relevant for the average center of mass and the Fourier transform of the center cells and average COM yield insignificantly different results.

The change of the Hamiltonian is needed for each copy attempt to calculate the forces on each lattice site. The source code of the CompuCell3D class `potts3D` has been modified (Details in section 5.3) to register the forces onto each lattice site during a simulation, and the `point3D` class has been altered to be able to handle floating point numbers for non-integer forces. The settings for the force field can be set in the simulation files for epithelial folding.

A hypothesis test is used to test if two simulations have significant different results, based on measurements such as growth of cells and the amplitude of the frequencies of the Fourier transform. The null hypothesis states that the means of the measured variables are equal. Two different simulations are independent samples of two parameter sets. Whether the variance of the samples are equal is unknown, but the samples have an equal sample size in all cases. It is assumed that the measurements of the samples

are normally distributed, and Shapiro-Wilk normality tests with average p-value 0.450 give no reason to assume otherwise. The conditions for Welch's t-test (Welch, 1947) are met. Welch's t-test is an adapted Student's t-test for samples with unequal variances. The test statistic t is defined as

$$t = \frac{\bar{X}_1 - \bar{X}_2}{\sqrt{\frac{s_1^2}{N_1} + \frac{s_2^2}{N_2}}} \tag{13}$$

with $\bar{X}_1$ and $\bar{X}_2$ the sample means, s_1, s_2 the sample variances and N_1, N_2 the sample sizes. The degrees of freedom are defined as

$$v = \frac{\left(\frac{s_1^2}{N_1} + \frac{S_2^2}{N_2}\right)^2}{\frac{s_1^4}{N_1^2(N_1-1)} + \frac{s_2^4}{N_2^2(N_2-1)}} \tag{14}$$

which is rounded down to the nearest integer. The null hypothesis that states that the means are equal is tested against the alternative that the means are different. The p-value is defined as $p = 2 \cdot P(T \geq |t|)$, and the null hypothesis is rejected if the p-value is lower than the significance level.

The parameter space is huge, and the possible different parameter sets are endless. A general sweep of the parameter space as well as parameter sweeps for most of the individual parameters have been generated and run using the Python scripts. The simulation results have been analyzed using a R script (Details in section 5.4).

3 Results

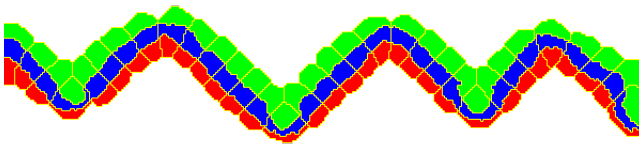
First the growth increment of all cell layers is kept equal, to verify the assumption of differential growth for epithelial folding. Next parameter sweeps are run for the homotypic bond energies of the bottom and middle cell layers to investigate the relations with the behaviour of the tissue. The bond energy of the top layer is not varied because of the symmetry of the tissue. Next the resistance to compression is varied and a parameter sweep for the growth increment of the middle cell layer is run to find the influence on the tissue layer.

The random seed makes every simulation different, even with the same parameter set. However, the general results remain the same. Fig. 5 shows the results of one single simulation for the model of epithelial folding.

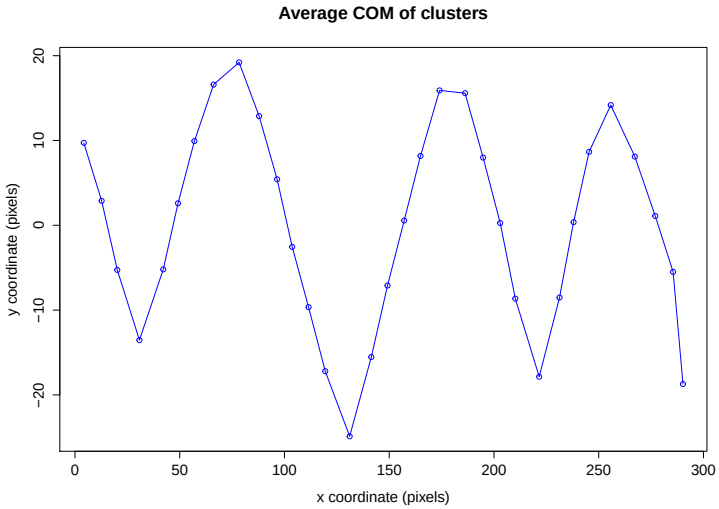
Fig. 5(a) shows the cell lattice at 100.000 MCS. The top (green) cell layer with $\Delta A_t = 0.0008$ grows faster than the center and bottom layers with $\Delta A_c = 0.0006$ and $\Delta A_b = 0.0004$, which is consistent with the cell volumes at the end of the simulation. The x-coordinate of each cell lies between the x-coordinates of its neighbors, which justifies the use of the Fourier transform. The tissue has a wavelike shape with 3.5 peaks, and a amplitude of roughly 18 pixels as shown in figure (b). Fig. 5(c) shows the Fourier transform. The low frequencies (harmonics) have a large wavelength and represent the behaviour of the layer at tissue level. The first frequency is the average y-coordinate, the equilibrium, of the tissue layer and translated to zero as shown in Fig. 5(b) and (c). The low frequencies form a peak and frequency 3 has the highest amplitde. This is in accordance with the 3.5 'folds' in the tissue. The higher frequencies have a small wavelength and represent fluctuations around the cell positions and shapes. These frequencies are not relevant for the shape of the tissue layer, and are most likely affected by the orientations of the cells. The middle frequencies have low amplitudes, which is to be expected because the tissue has a small wavelength and the middle frequencies will give it a higher frequency.

The target volume of a cell cluster increases 0.0006 every MCS, resulting in an expected 3.22 growth ratio using eq. 8. The growth of the cells varies between 3.17 and 3.36 times the original volume with an average growth ratio of 3.24, higher than expected. The fluctuations in growth are most likely caused by pressure and stress in the tissue. However, there is not an obvious relation between the location of the cell in the tissue and its growth ratio, since the peaks of the folds do not directly correspond with higher or lower growth ratios.

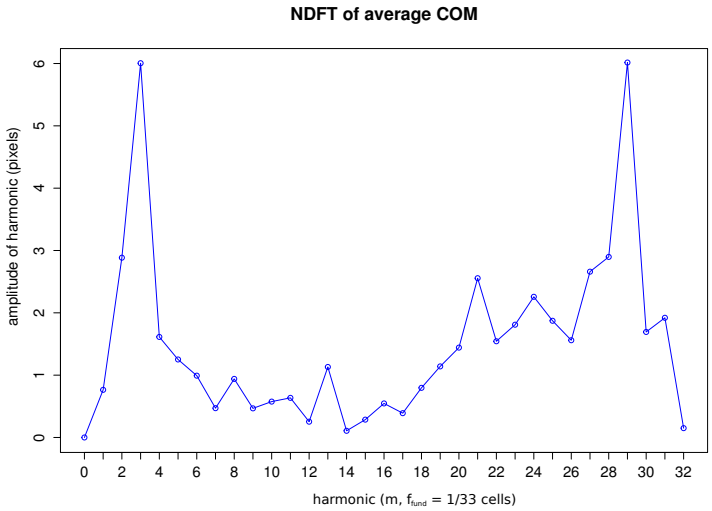
Fig. 5(e) and (f) show the forces on the cell lattice. The forces are higher during the simulation than at the end, and the average forces downwards from the equilibrium are lower compared to the forces upwards, which is probably caused by differential growth in the tissue layer. The opposite direction of these forces implies that the peaks of the waves in the tissue are moving apart, and the tissue at 50.000 MCS is moving faster compared to the end of the simulation. The forces at the end of the simulation are more organized, and forces on neighboring lattice sites are roughly opposite. This is as expected, because the medium and cells exert pressure on each other which leads to forces outwards from the cells and towards the cells from the medium and neighboring cells. This is most likely causeb by neighboring cells that push against each other to be able to grow.



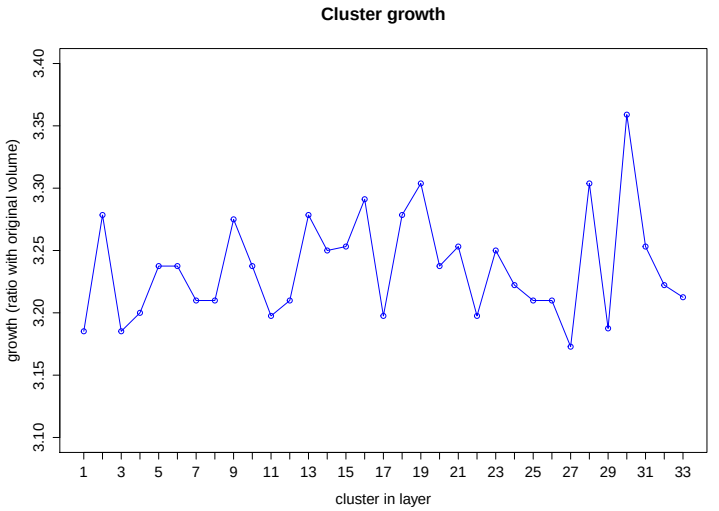
(a)



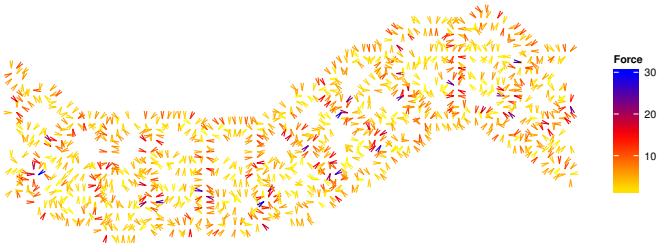
(b)



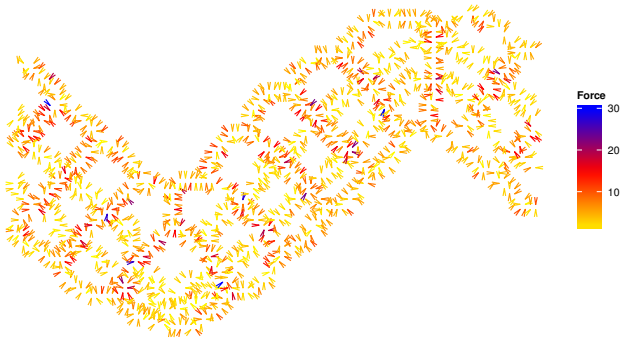
(c)



(d)



(e)



(f)

Figure 5: Simulation analysis at 100.000 MCS of one simulation. *(a) shows the cell lattice at the end of the simulation, (b) shows the the center of mass of every cluster, (c) shows the amplitude of the frequencies from the Fourier transform of the COM of the clusters shown in (b), and (d) shows the growth ratio of the cluster cells. (e) and (f) show the forces for a part of the cell lattice for x-coordinate* $x \in [100, 200]$ *pixels after 50.000 MCS and 100.000 MCS. The average forces are* 5.46 *during and* 5.28 *at the end of the simulation, with* 1.07 *and* 0.79 *up and downwards from the equilibrium during the simulation and* 0.87 *and* 0.59 *up and downwards at the end.*

3.1 Necessity of differential growth for epithelial folding

The assumption for epithelial folding is differential growth between three cell layers in the tissue. Fig. 6(a) and (b) show the cell lattice for two simulations with equal growth between the three layers. The tissue folds for equal growth between the cell layers as well. The cause for buckling could be the boundaries of the cell lattice that act as tissue that exerts counterpressure to the tissue. However, the same results, except for the shape of the cells at the boundaries of the tissue, are found for a bigger lattice without the cells reaching the boundaries. The folding of the tissue is most likely caused by pressure that builds up as a result from the growth of cells. The cells at the boundary of the tissue cannot move fast enough to relieve the pressure of the tissue layer which leads to buckling of the tissue.

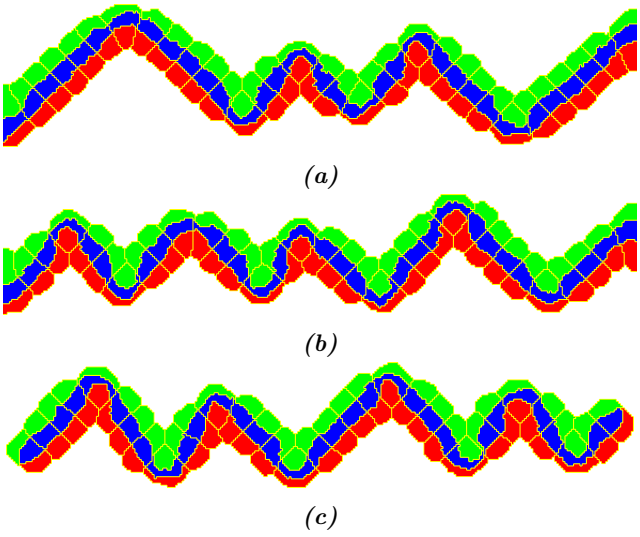
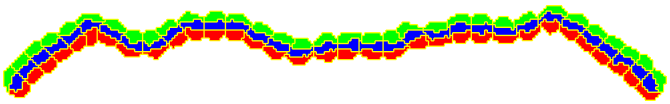


Figure 6: Cell lattice at 100.000 MCS for equal growth increments $\Delta A_t = \Delta A_c = \Delta A_b = 0.0006$***.*** *(a) and (b) show the cell lattice with the boundary cells pushing against a surface. (c) shows the simulation without boundaries for the same random parameter as (b).*

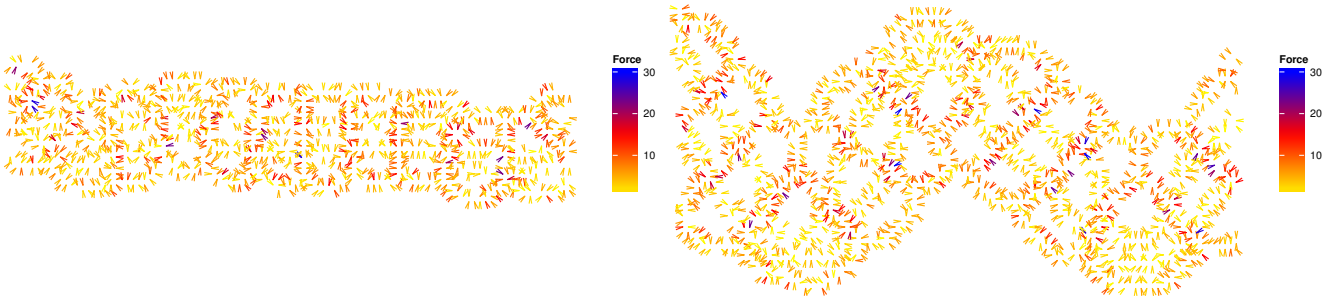
Fig. 7 shows the cell lattice for a simulation without growth. The tissue does not have the characteristic wave pattern from previous simulations, confirming that folding occurs as a consequence of pressure in the tissue because of growth and not because of differential growth between cell layers. This suggests that altough differential growth is not needed, a necessary element for epithelial folding is tissue growth.

Fig. 8 shows the forces on the cell lattice for the same simulation with and without lattice boundaries that exert counterpressure. The average forces are approximately the same, but the forces up from the equilibrium are not larger than the forces downwards. This suggests that the larger upward forces found in previous simulations are caused by differential growth. These larger forces lead to more pressure in the tissue, and makes the cell layer grow or move faster.

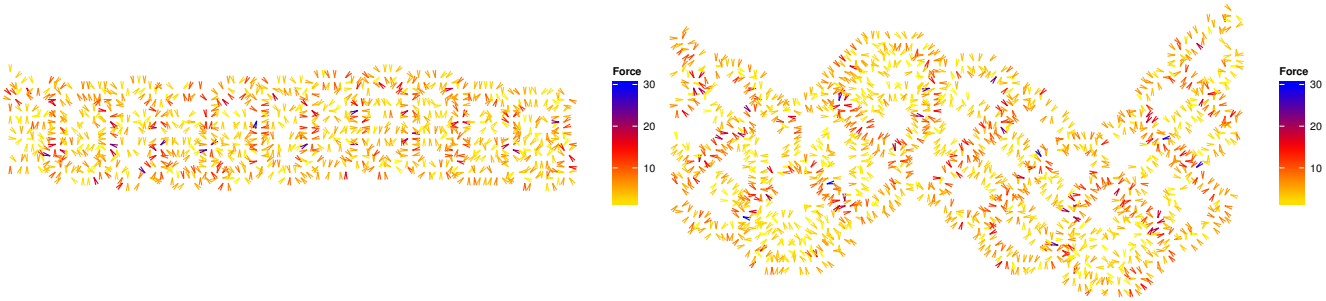


(a)

Figure 7: Cell lattice at 100.000 MCS without growth or boundaries. *The same random seed as in Fig. 6 is used.*



(a) *(b)*



(c) *(d)*

Figure 8: Part of the force field for x-coordinate $x \in [100, 200]$ ***pixels at 50.000 and 100.000 MCS for equal growth with and without boundaries.*** *(a) shows the force field with boundaries and average forces* 5.57 *with on average* 0.47 *up and* 0.10 *downwards from the equilibrium at 50.000 MCS. (b) shows the same result at 100.000 MCS with average force* 5.19 *and* 0.17 *and* 0.53 *up and downwards. (c) and (d) show the force field without boundaries with average forces* 5.28 *and* 5.07*. The forces upwards are* 0.33 *and* 0.83*, and the forces downwards from the equilibrium* 0.88 *and* 0.90*.*

3.2 Homotypic bond energy

3.2.1 Continuity of the tissue depends on all bond energies

Fig. 9 shows the cell lattice where all homotypic bond energies are varied. The tissue remains flat for bond energy $J = J_{b,b} = J_{c,c} = J_{t,t} = 5$, where the homotypic bonds are strong and cells maximizing their surface area with their neighbors which results in rectangular cells. When the bond energies increase, the amplitude and frequency of the wave seem to increase. The tissue layer folds for higher bond energies, but fractures as expected at multiple points for high surface energies because the contacts between the cells are weak and easy to break. Frequency analysis yields no significant results for low bond energies and is not justified for high bond energies because the tissue fractures. Growth ratio analysis yields no significant different results as well.

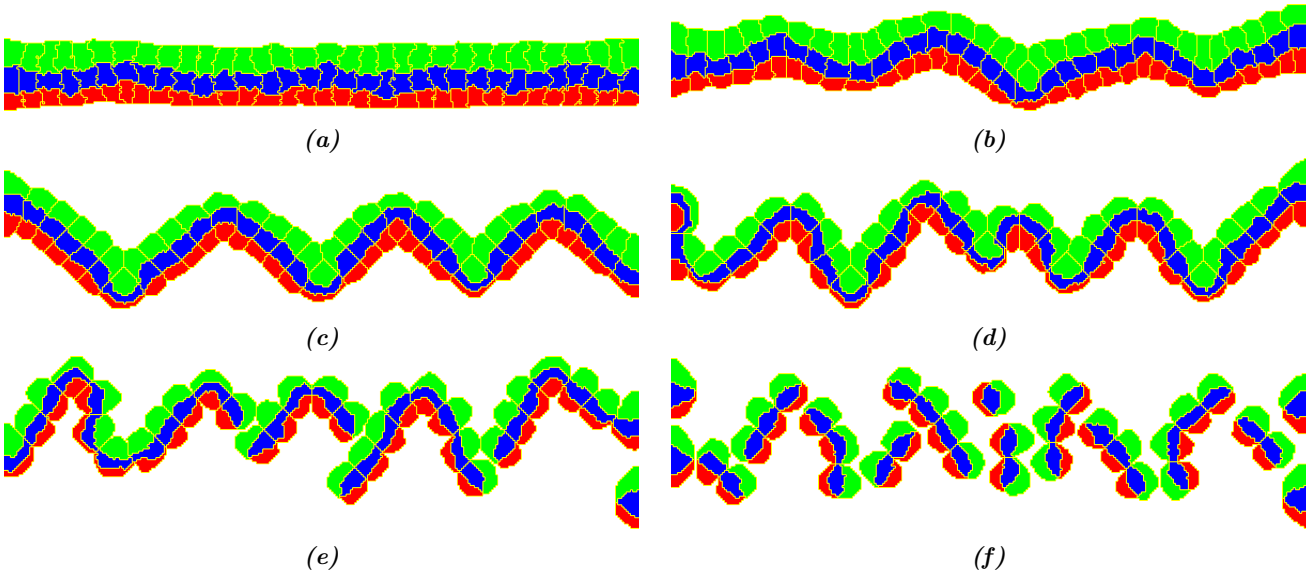


Figure 9: Cell lattice at 100.000 MCS for different homotypic bond energies for the three cell types. *(a) shows the cell lattice for homotypic contact energies* $J = J_{b,b} = J_{c,c} = J_{t,t} = 5$ *yielding* $\gamma_{c,t} = \gamma_{c,b} = 95$ *and* $\gamma_{c,b} = 195$. *(b) shows the cell lattice for surface energies* $J = 20$ *with* $\gamma_{c,t} = \gamma_{c,b} = 80$ *and* $\gamma_{c,b} = 80$, *(c) for* $J = 35$ *with* $\gamma_{c,t} = \gamma_{c,b} = 65$ *and* $\gamma_{c,b} = 165$, *(d) for* $J = 50$ *with* $\gamma_{c,t} = \gamma_{c,b} = 50$ *and* $\gamma_{c,b} = 150$, *(e) for* $J = 65$ *with* $\gamma_{c,t} = \gamma_{c,b} = 35$ *and* $\gamma_{c,b} = 135$ *and (f)* $J = 80$ *with* $\gamma_{c,t} = \gamma_{c,b} = 20$ *and* $\gamma_{c,b} = 120$.

Fig. 10 shows the forces on a part of the cell lattice of the simulation. The forces between cluster cells are low for a low bond energy. These contacts between cluster cells are lost for a high bond energy, resulting in a large surface area of the cluster cells that is in contact with the medium. The forces with medium are higher, resulting in a higher average force for high bond energies.

The forces are approximately perpendicular to the tissue for low bond energies, while they are directed to the center of the cluster cells for high bond energies. This is most likely caused by the loose contacts between the cells. The contacts are weak for high contact energies, which separates the cells in the cell layers and result in individual cluster cells. The formation of individual cluster cells is probably associated with the forces of the cells that are directed to the center of the cluster which results in round clusters.

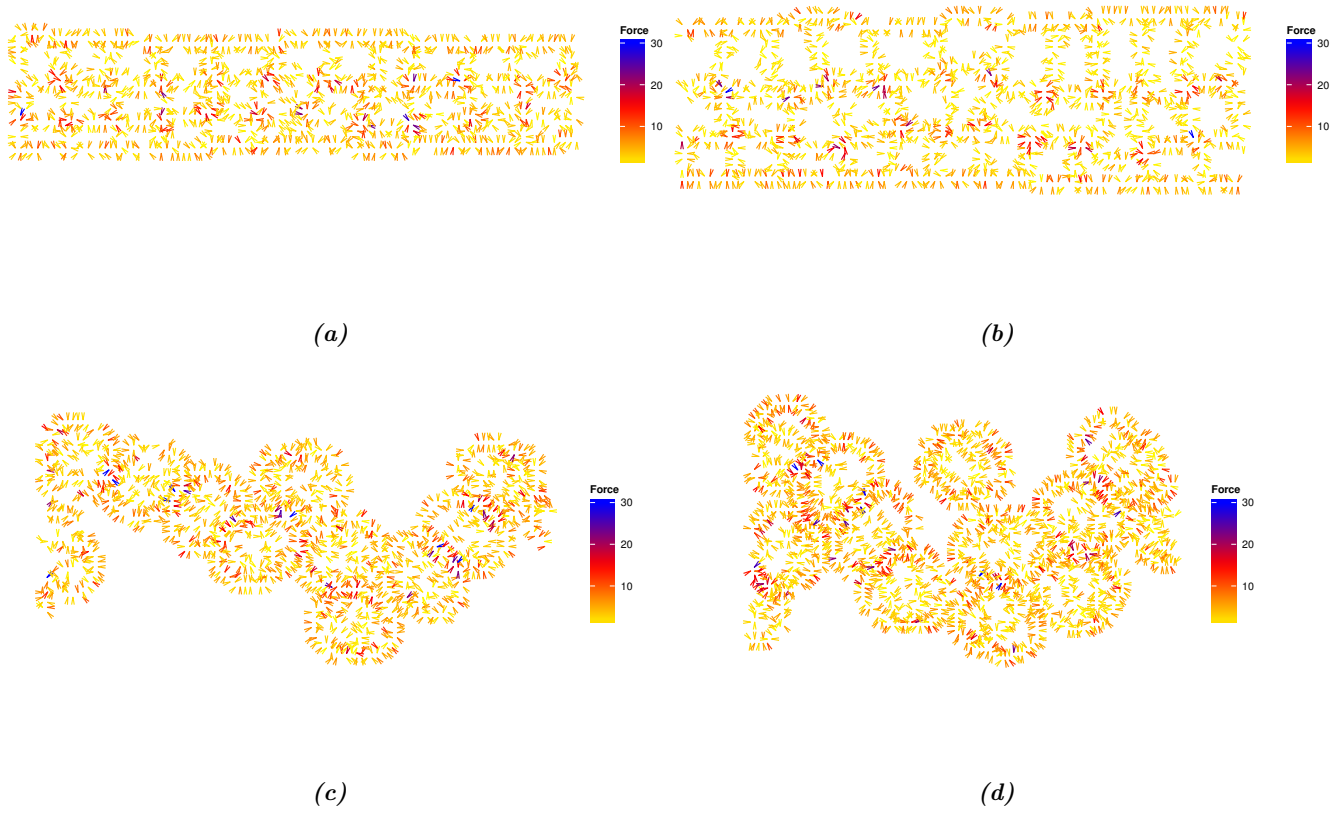


Figure 10: Part of the force field for x-coordinate $x \in [100, 200]$ ***pixels at 50.000 and 100.000 MCS for different homotypic contact energies.*** *(a) and (b) show the force field for bond energy* $J = 5$ *with average forces* 4.18 *and* 3.58 *at 50.000 and 100.000 MCS. (c) and (d) show bond energy* $J = 80$ *with average forces* 4.86 *and* 4.64 *at 50.000 and 100.000 MCS.*

3.2.2 Bottom homotypic bond energy determines the amplitude

The cell lattice for different bond energies for the bottom cell layer is shown in Fig. 11. The bonds energy for the bottom cells in the cell lattice in Fig. 11(a) is equal to $J_{b,b} = 10$ and is highly homotypic, which makes the bottom layer fold onto itself and break the other two cell layers at some points in the tissue. Strong cohesion of the bottom cells together with growth and pressure in the tissue make parabolic shapes appear at the top op the layer. The effect is less visible in Fig. 11(b) and 11(c) with higher surface energies.

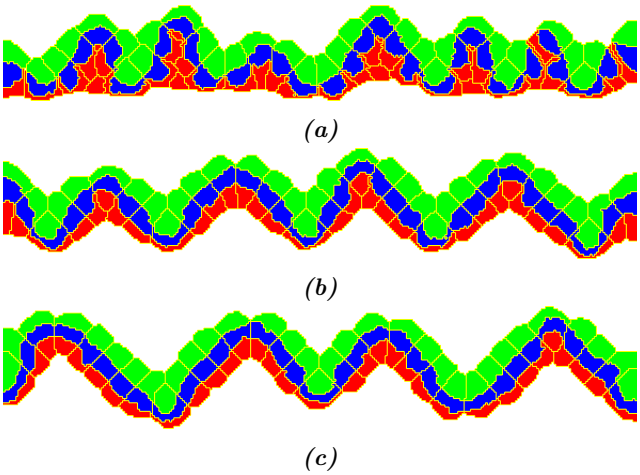


Figure 11: Cell lattice at 100.000 MCS for different homotypic contact energies of the bottom cell layer. *(a) shows the cell lattice for homotypic surface energy* $J_{b,b} = 10$ *with surface tensions* $\gamma_{b,t} = 175$ *and* $\gamma_{b,c} = 75$. *(b) shows the cell lattice for bond energy* $J_{b,b} = 24$ *with* $\gamma_{b,t} = 168$ *and* $\gamma_{b,c} = 68$ *and (c) for* $J_{b,b} = 36$ *with* $\gamma_{b,t} = 162$ *and* $\gamma_{c,t} = 62$.

Growth analysis yields no significant differences between simulations with different bond energy. A Fourier transform for frequency analysis is not justified because the tissue is not a wave for low surface energies and the tissue fractures at some points. The amplitude of the tissue is smaller for low bond energies.

Altough a Fourier transform is not justified, it yields that most of the simulations with low bond energies for the bottom cell layer have a broad band of frequencies that determine the shape of the tissue as shown in Fig. 12. The ninth harmonic has roughly the same amplitude for each parameter value which indicates the transition from shape frequencies to fluctuation frequencies. Low frequencies contribute more equally to the shape of the tissue for low bond energies, and the fourth harmonic varies strong between simulations with different bond energies. There is however no obvious relation between the surface energy and the amplitude of the frequencies.

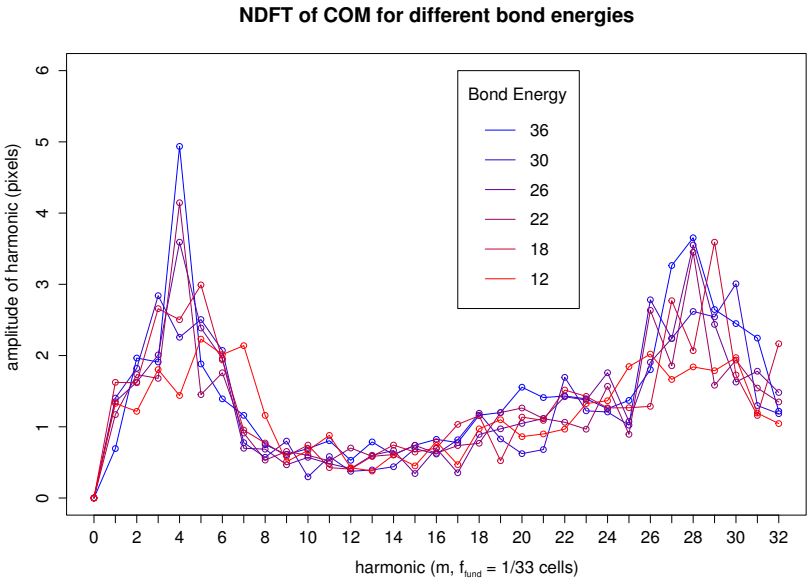


***Figure 12:** NDFT of COM of five simulations for each parameter set at 100.000 MCS for different homotypic surface energies of the bottom cell layer.*

An estimation of the amplitude of the wave is double the average distance of the centers of mass in the tissue to the middle of the layer. The average vertical displacement of the COM for different parameter values are shown in Fig. 13. The amplitude increases with a higher bond energy. The standard parameter set yields an average displacement of 9.41 with the linear equation. Then the real amplitude of the wave in the tissue should be approximately 18.8, which is in agreement with the tissue from Fig. 5(b). Table 2 shows the p-values for comparison of the average displacement between simulations with different bond energies. The values are lower than 0.05 for bond energies differing more than $\Delta J_{b,b} = 10$. The p-values show that it is unlikely that the average displacements for different bond energies is the same. This suggests that the amplitude of the tissue layer increases with a higher bond energy of the bottom cells.

***Table 2:** p-values for the comparison of the average displacement of the cluster cells and bond energies.*

	bond energy ($J_{b,b}$)				
bond energy ($J_{b,b}$)	10	16	22	28	34
10	-	**0.0077**	**0.022**	**0.0029**	**0.0018**
16		-	0.125	**0.0074**	**0.0008**
22			-	**0.0587**	**0.01**
28				-	0.440

The forces on the tissue layer at the end of the simulations are shown in Fig. 14. They are not significantly different for different parameter values. The average force for bond energy $J_{b,b} = 10$ is smaller compared to other simulations, but it has a larger force upwards from the equilibrium, indicating that the finger-like structures in the tissue are still moving, and move faster than the bottom of the tissue layer. The forces between

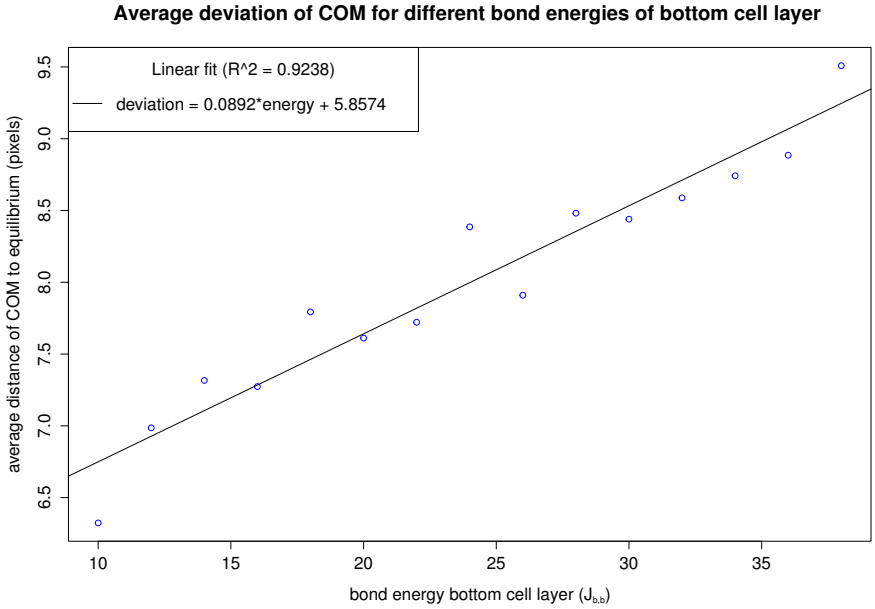
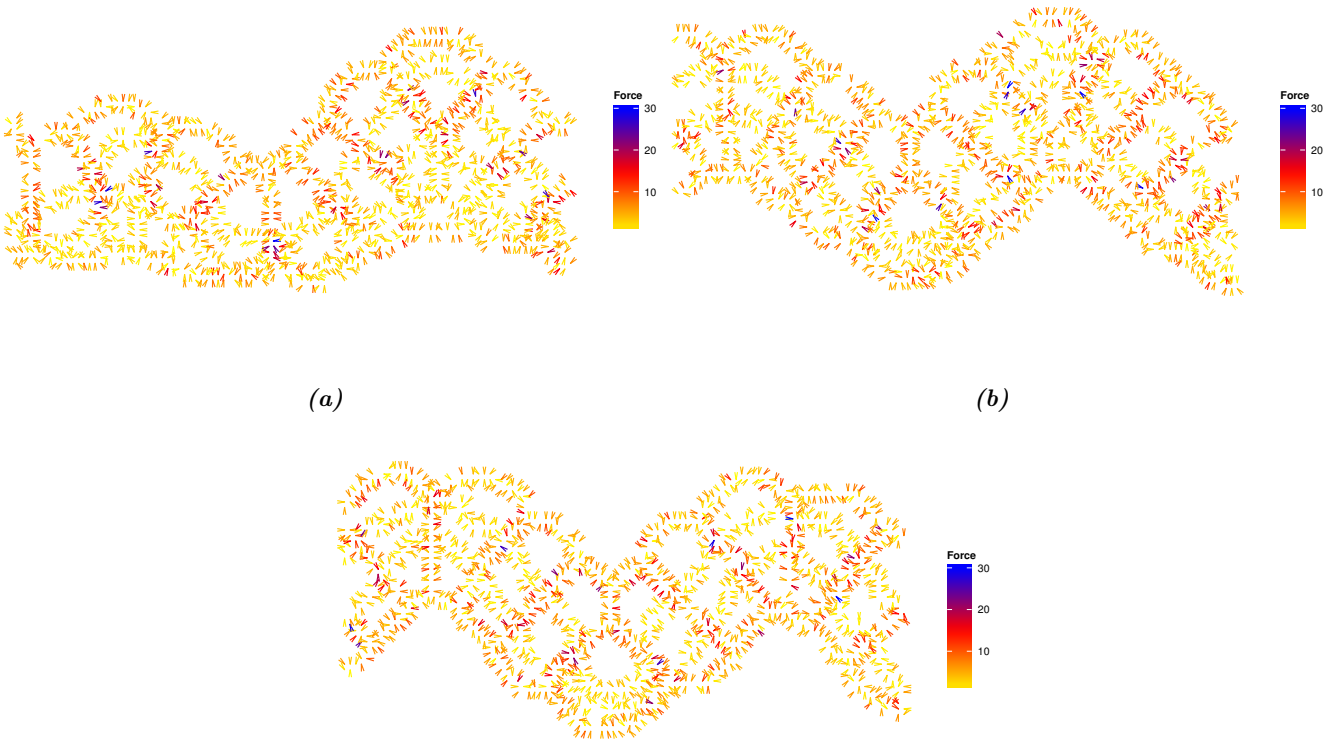


Figure 13: Linear fit for the bond energies of the bottom cell layer and the average vertical displacement from equilibrium. *Five simulations were ran for each surface energy, and the COM at 100.000 MCS are used.*

cluster cells of the bottom layer are smaller in comparison with the bottom layers for $J_{b,b} = 24$ or 36. The forces of neighboring lattice sites are roughly opposite to each other as expected for bond energies $J_{b,b} = 24$ and $J_{b,b} = 36$, but more disorganized for bond energy $J_{b,b} = 10$.

A low bond energy between cells of the bottom cell layer gives strong contacts between the cells and make the bottom cells maximize their contact area. This results in a stiffer cell layer which is harder to stretch. Growth of the cells leads to parabolic structures at the top of the tissue because the bottom cell layer is hard to bend. A higher bond energy makes the bottom layer easier to bend and lets the tissue fold in both directions.



(a) *(b)*

(c)

Figure 14: Part of the force field for x-coordinate $x \in [100, 200]$ ***pixels at 100.000 MCS for different homotypic bond energies of the middle cell layer.*** *(a) shows the force field for bond energy* $J_{b,b} = 10$ *with average force* 4.61*. The average forces above and under the equilibrium are upwards with* 1.17 *and downwards with* 0.56*. (b) shows the forces for surface energy* $J_{b,b} = 24$ *with average force* 5.10 *and* 0.93 *up and* 0.67 *downwards. (c) shows bond energy* $J_{b,b} = 36$ *with average force* 5.04 *and* 0.76 *up and* 0.44 *downwards.*

3.2.3 Center homotypic bond energy determines the frequency

The cell lattice for different surface energies for the middle cell layer is shown in Fig. 15. The contacts between the cells are strong if the surface energy is low, as shown in Fig. 15(a) where the center cells make as much contact as possible. A high bond energy results in weak connections and less contact area. Homotypic contacts are favored, but heterotypc bonds are energetically more favorable for bond energy $J_{c,c} = 52$ compared to $J_{c,c} = 16$. Growth analysis yields no significant different growth ratios for different bond energies, but the frequency of the tissue increases with higher surface energies.

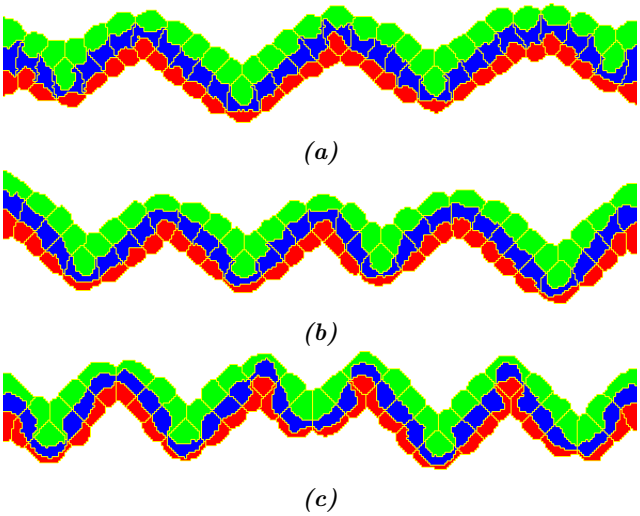


Figure 15: Cell lattice at 100.000 MCS for different homotypic bond energies of the middle cell layer. *(a) shows the cell lattice for surface energy $J_{c,c} = 16$ with surface tensions $\gamma_c = \gamma_{c,t} = \gamma_{c,b} = 72$. (b) shows the cell lattice for bond energy $J_{c,c} = 26$ and $\gamma_c = 67$ and (c) for $J_{c,c} = 52$ and $\gamma_c = 54$.*

The Fourier tranform of the tissue layers is shown in Fig. 16 for different bond energies. The lower frequencies corresponding with the shape of the tissue show that a higher bond energy yields a higher frequency peak with the Fourier transform, but the amplitude of the folds remains roughly the same. For lower surface energies, such as $J_{c,c} = 16$ and $J_{c,c} = 18$ the frequency with highest amplitude is 3, and for middle and higher contact energies the frequency peak is 4 and 5. The amplitude of frequency 7 is roughly the same for all contact energies, indicating that the frequencies larger than 7 are most likely not relevant for shape of the tissue layer. The middle frequencies have a higher amplitude for higher bond energies as anticipated. A higher surface energy yields a tissue with more folds, which makes the middle frequencies more important for the shape of the tissue. The high frequencies have a higher amplitude for low bond enrgies, which is most likely caused by the contact area of the center cells which causes fluctuations in the centers of mass.

The tissue layers from the simulations shown in Fig. 16 have sharp peaks in the Fourier transform at frequencies 3,4 and 5. The frequencies that determine the tissue shape have more evenly distributed amplitudes and broader peaks for other bond energies, which is probably caused by interference by transitions between different stable

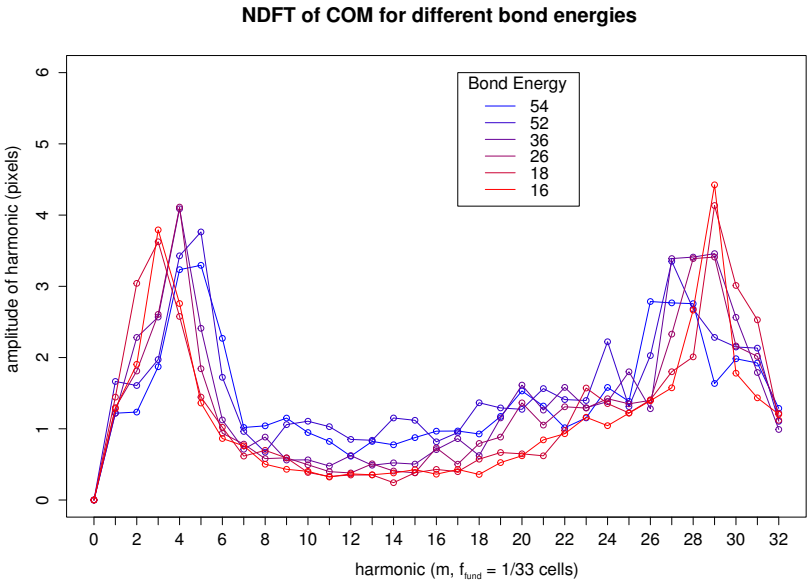


Figure 16: NDFT of COM of ten simulations for each parameter set at 100.000 MCS for different homotypic bond energies of the middle cell layer.

frequencies. This causes the frequency of the folded tissue to distribute over multiple frequencies. Another, more likely explanation is a different result between simulations with different random seeds. At some point the layer probably has to choose between multiple stable frequency configurations because of stress and pressure in the tissue. This splits the simulations into two or more stable frequencies which lead to a broader peak in the Fourier transform.

The p-values for the low frequencies are shown in Table 3. The p-values are on average lower than 0.1 for the simulations with the highest amplitude at frequency 3 or 4. The same holds for simulations with highest amplitude for frequencies 3 or 5 and 4 or 5. These p-values make it unlikely that the frequency of the tissue is not different for different surface energies, and indicate that the frequency of the tissue increases with higher bond energies.

Table 3: *p-values for the comparison of the frequencies for different bond energies. Only the p-values for the frequencies with high amplitudes that influence the shape of the tissue are shown.*

	frequency				
bond energies ($J_{c,c}$)	2	3	4	5	6
16 and 18	**0.074**	0.811	0.677	0.781	0.561
16 and 26	0.852	**0.068**	**0.059**	0.235	0.746
16 and 36	0.529	**0.039**	**0.078**	**0.039**	0.272
16 and 52	0.476	**0.013**	0.317	**0.000**	**0.022**
16 and 54	0.156	**0.002**	0.468	**0.003**	**0.004**
18 and 26	**0.058**	0.166	**0.021**	0.361	0.770
18 and 36	0.281	0.126	**0.032**	**0.065**	0.720
18 and 52	**0.019**	**0.041**	0.147	**0.000**	**0.075**
18 and 54	**0.006**	**0.013**	0.246	**0.004**	**0.010**
26 and 36	0.441	0.949	0.975	0.312	0.461
26 and 52	0.636	0.363	0.373	**0.002**	**0.038**
26 and 54	0.232	0.183	0.251	**0.026**	**0.006**
36 and 52	0.223	0.353	0.411	**0.030**	0.104
36 and 54	**0.082**	0.142	0.286	0.182	**0.014**
52 and 54	0.331	0.866	0.792	0.459	0.258

The forces on the cell lattice at the end of the simulations for some contact energies are shown in Fig. 17. The direction of the forces is disorganized and not always opposite to the forces of neigboring cells for surface energy $J_{c,c} = 16$. This suggests that the tissue is moving. The forces up and downwards from the equilibrium indicate that the peaks of the folds in the layer move in opposite directions, and the top half of the layer moves faster. The forces between cluster cells are smaller for the middle cell layer than the top and bottom cell layer. These forces are also smaller than the forces between the cells of the middel cell layer of other bond energies.

The forces for the other bond energies are approximately the same, but better organized and often in opposite directions for neighboring lattice sites, which indicates that the tissue layer is less motile. The forces outward from the equilibrium suggest that the vertical displacement of the peaks is increasing at the end of the simulation.

The relation between the bond energies for all simulations and the average frequency is shown in Fig. 18. The frequency increases with higher bond energies as a general trend. The fit has not a high accuracy, but that is to be expected because the bond energies with broader peaks have bigger fluctuations from their expected frequencies.

A low bond energy between cells of the middle cell layer results in stronger homotypic contacts and let the center cells adhere better between clusters. This makes the bonds between these cells stronger and the middle cell layer stiffer and harder to bend and stretch. This makes the tissue layer harder to fold because it requires larger forces to bend the middle cell layer, which results in a lower frequency at the end of a simulation.

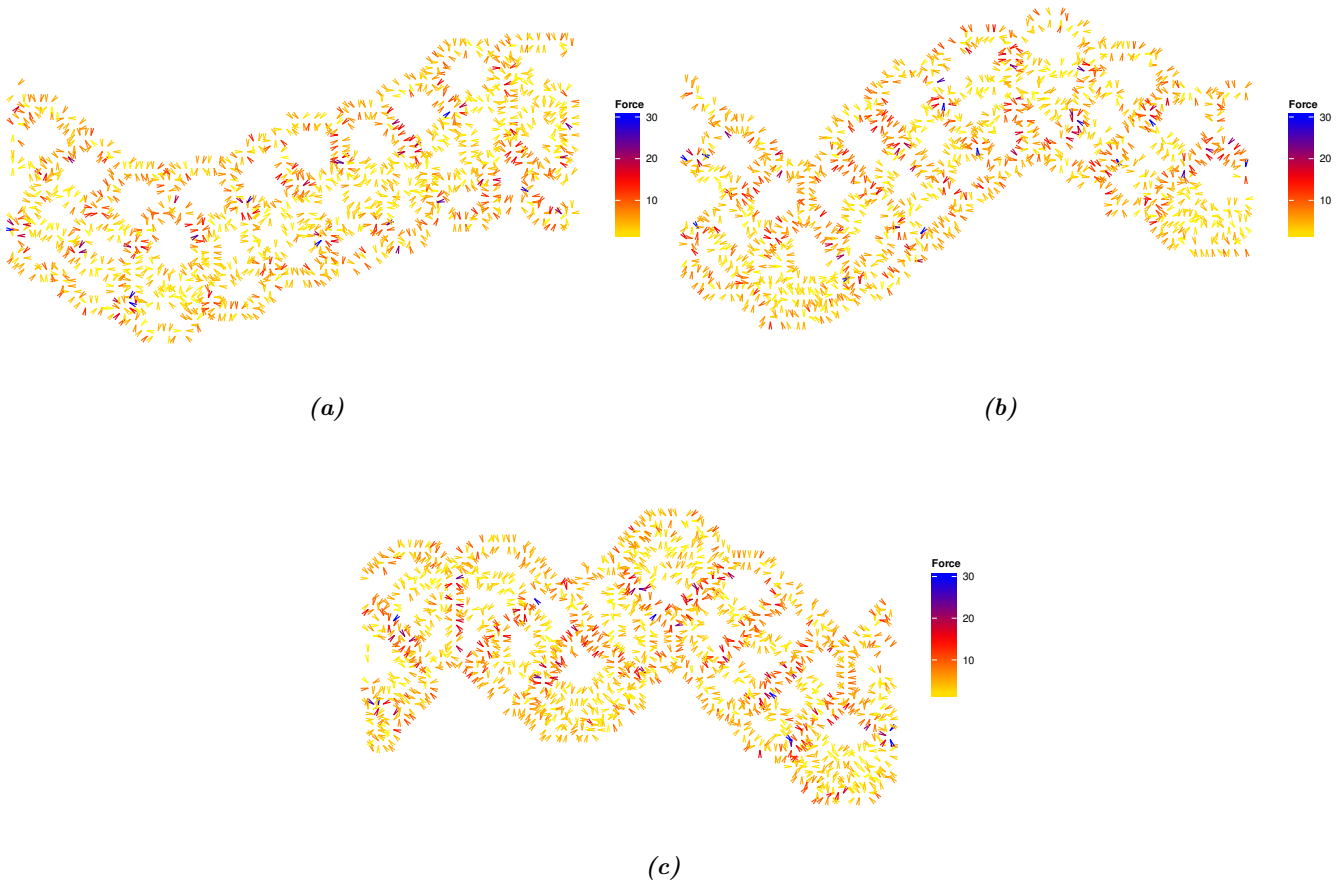


Figure 17: Part of the force field for x-coordinate $x \in [100, 200]$ ***pixels at 100.000 MCS for different homotypic surface energies of the middle cell layer.*** *(a) shows the force field for bond energy* $J_{c,c} = 16$ *with average force* 4.94*. The average forces above and under the equilibrium are upwards with* 1.19 *and downwards with* 0.44*. (b) shows the forces for surface energy* $J_{c,c} = 26$ *with average force* 5.15 *and* 0.62 *up and* 0.50 *downwards. (c) shows bond energy* $J_{c,c} = 52$ *with average force* 5.17 *and* 1.16 *up and* 0.55 *downwards.*

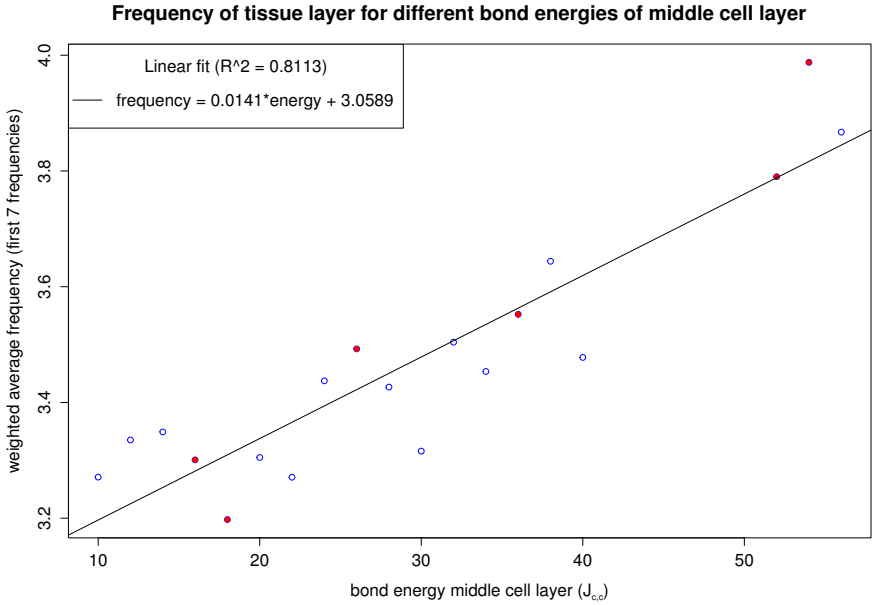
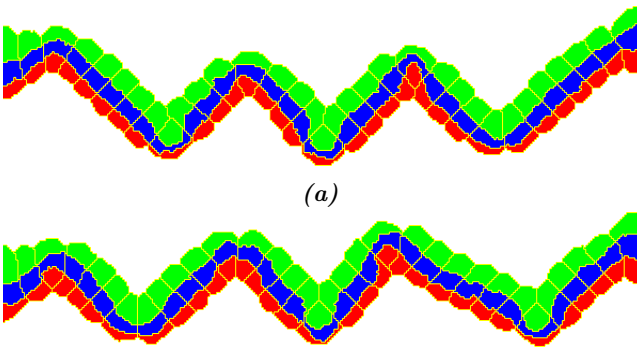


Figure 18: Linear fit for the bond energies of the middle cell layer and weighted average frequency of the first 7 ***low frequencies.*** *Ten simulations were ran for each surface energy, and the COM at 100.000 MCS are used for the NDFT.*

3.3 Resistance to compression specifies growth deviations

Fig. 19 shows the cell lattice for different resistance to compression λ. The resistance to compression determines the penalty for deviations from the target volume. It is anticipated that a higher penalty results in less motile cells, which would lead to smaller waves. However, the cell lattice shows no visible differences, and frequency and amplitude analysis yield no significant different results.



(a)

(b)

Figure 19: Cell lattice at 100.000 MCS for different resistance to compression. *(a) shows the cell lattice for a resistance to compression* $\lambda = 100$ *and (b) for* $\lambda = 800$.

The simulations have the same growth increments for all parameter sets. Fig. 20 shows the average growth ratio of the cluster cells compared with the resistance to compression, with theoretical growth ratio equal to 3.22 using eq. 8. The growth ratio is smaller than the theoretical value for low λ as expected, because the penalty for deviations is small which results in the cell volume lagging behind the target volume. The growth ratio increases further and above the theoretical value for higher λ. This is against expectation but probably caused by stress in the tissue which streches the cells. The growth ratio converges to the theoretical value as anticipated for high λ, because a higher penalty results in smaller deviations from the target volume.

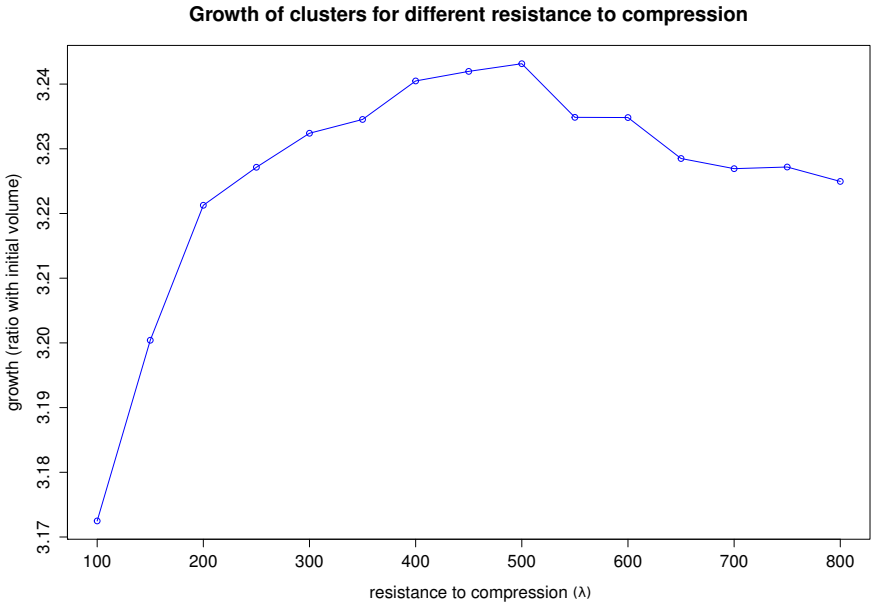
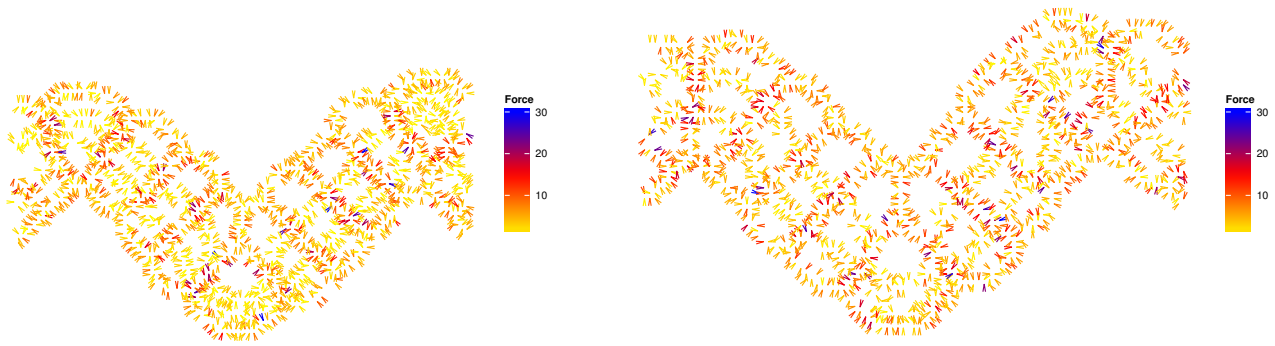


Figure 20: Average growth ratio for different resistance to compression at 100.000 MCS. *Five simulations have been run for each for each parameter set.*

Fig. 21 shows the forces on the cell lattice at the end of the simulations. The forces for $\lambda = 100$ are smaller than the forces for $\lambda = 800$. This is probably caused by pressure that cells exert on each other. All the cells have a high penalty for deviations from the target volume, which results in a bigger change of hamiltonian and bigger forces. The forces directed up and downwards from the equilibrium are approximately equal, indicating that the difference in average force is mainly caused by cells trying to retain their target volume. The forces above are larger than beneath the equilibrium, which suggests that the top peaks move faster than the bottom peaks in the tissue.

Fig. 22 shows the variance of the growth ratios of the simulations. The variance decreases for higher λ as expected. A higher penalty results in less deviations from the target volume and a smaller variance. The resistance to compression has, besides a small influence on the growth of the cluster cells for low values of λ, no influence on the behaviour of the tissue layer. The growth ratio from previous simulations cannot be linked to the location in the tissue because fluctuations in the growth ratio are too large. A large enough value for the resistance to compression results in small variance of the growth ratios, which makes the deviations in growth ratio because of the location in the tissue more significant.



(a) *(b)*

Figure 21: Part of the force field for x-coordinate $x \in [100, 200]$ ***pixels at 100.000 MCS for different resistance to compression.*** *(a) shows the force field for resistance to compression* $\lambda = 100$*. The average forces are* 4.69 *with* 0.47 *up and* 0.1 *downward from the equilibrium. (c) shows the force field for* $\lambda = 800$ *with average forces* 6.20 *and* 0.37 *and* 0.01 *up and downward from the equilibrium.*

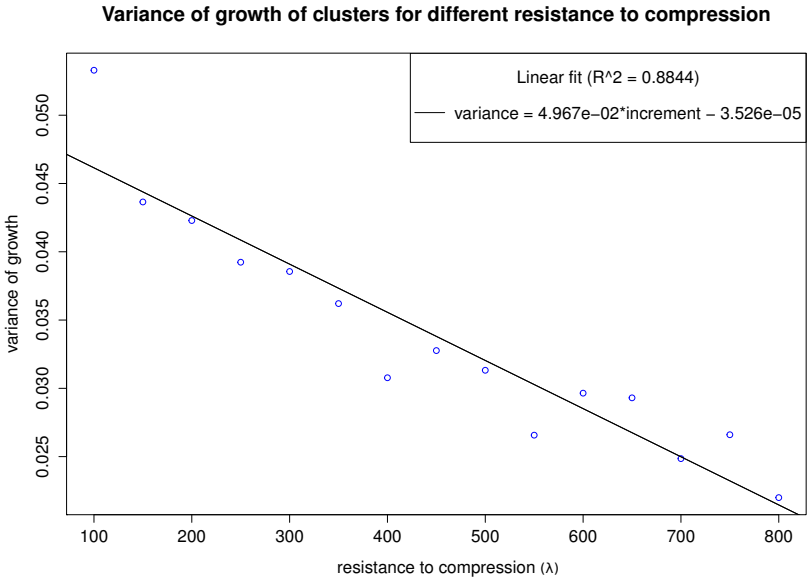


Figure 22: Linear fit of the variance of growth ratio at 100.000 MCS. *Five simulations were ran for each parameter set. The average variance of the growth ratio decreases for a higher resistance to compression as shown.*

3.4 Bigger growth increments speed up tissue folding

Fig. 23 shows the results for simulations with different growth speeds of the middle cell layer. The growth increment for each parameter set ranges from $\Delta A_c = 0.0002$ to $\Delta A_c = 0.001$.

A bigger increment of the targetvolume results in a higher growth of the center and cluster cells. This is to be expected, because deviations from the target volume of the cells increases the Hamiltonian which makes it energetically favorable for cells to expand their volume and results in bigger cell volumes. The center cells do not grow fast enough for the folding tissue for small growth increments, and the layer breaks as shown in Fig. 23(a).

The Fourier transform for frequency analysis yields no significant differences between different simulations, which all have the peak around frequency 4. Force field analysis yield no different results for different parameter values.

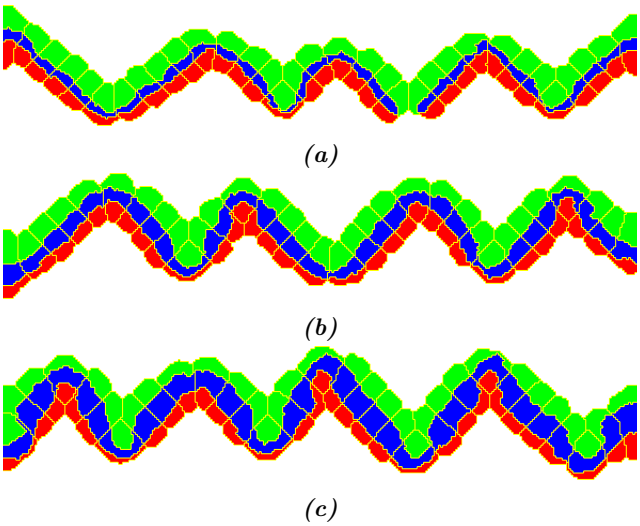


Figure 23: Cell lattice at 100.000 MCS for different growth increments of the center cell layer. *(a) shows the cell lattice for growth increment* $\Delta A_c = 0.0002$*. (b) shows the cell lattice for increment* $\Delta A_c = 0.0006$*, and (c) for* $\Delta A_c = 0.001$*.*

Fig. 24 shows the growth ratios of the cluster cells for different growth increments. A higher growth increment yields a higher growth ratio as anticipated, with all the cells in the tissue having roughly the same growth. The differences in average growth for different clusters in the tissue are most likely caused by the folds in the tissue, which lead to higher or lower pressure at certain cells and makes it harder for cells to grow or forces them to grow. There is no obvious relation between the growth ratios and the location of the cell in the tissue layer as stated earlier.

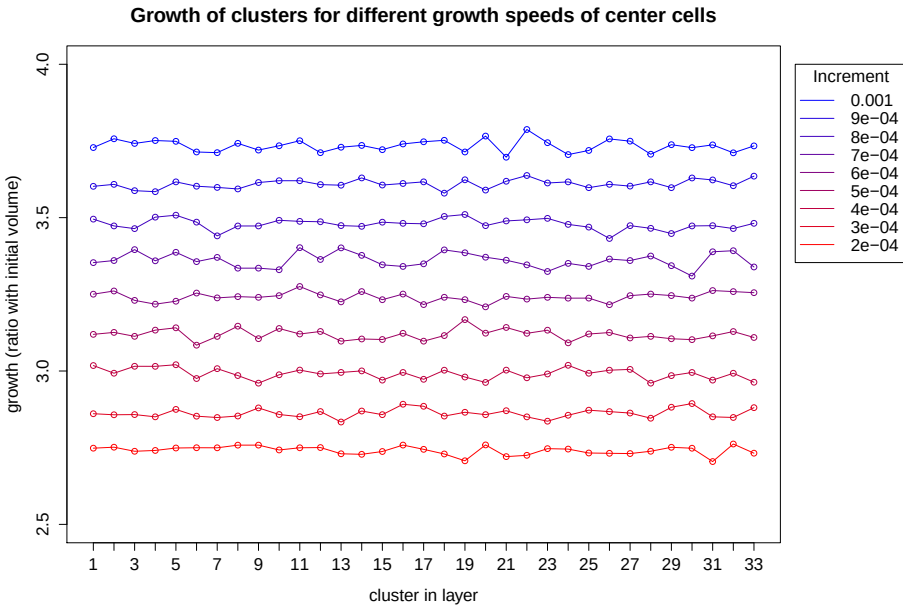


Figure 24: Growth ratio of five simulations for each parameter set at 100.000 MCS. *The average growth ratio increases for bigger increments of the targetvolume. The volume of the clusters at the end of the simulation varies between 2.74 and 3.73 times their original volume.*

Table 4: *p-values for the comparison of the growth of the cluster cells. The mean p-value for the clusters in the tissue layer is given. The comparison is between growth increments with a difference of 0.0001.*

growth increment	p-value
0.0002 and 0.0003	**0.0019**
0.0003 and 0.0004	**0.0021**
0.0004 and 0.0005	**0.0030**
0.0005 and 0.0006	**0.0031**
0.0006 and 0.0007	**0.0023**
0.0007 and 0.0008	**0.0054**
0.0008 and 0.0009	**0.0022**
0.0009 and 0.001	**0.0051**

The p-values for the comparison of the growth ratios are shown in Table 4. All p-values are notably small, indicating that the probability for results with at least as different average growth ratios if one assumes that the growth ratios are the same is extremely small. For growth increments that differ more than 0.0001 the p-values are smaller compared to those in Table 4.

A linear fit for the average growth ratio of the cluster cells and the growth increment of the center cells yields the relation $growth(x) = 1239.07 \cdot x + 2.494$ with $R^2 = 0.9999$. The average growth increment of the bottom and top cell layer is 0.0006, and the original cell volume is 27 pixels. Then the expected growth ratio, using eq. 8, for a cluster cell at the end of a simulation for growth increment x of the middle cell layer is

$$growth(x) = \frac{3 \cdot 27 + 100000 \cdot (0.0004 + x + 0.0008)}{3 \cdot 27} = 2.48 + 1234.6 \cdot x$$

which is considerably close to the experimentally found result.

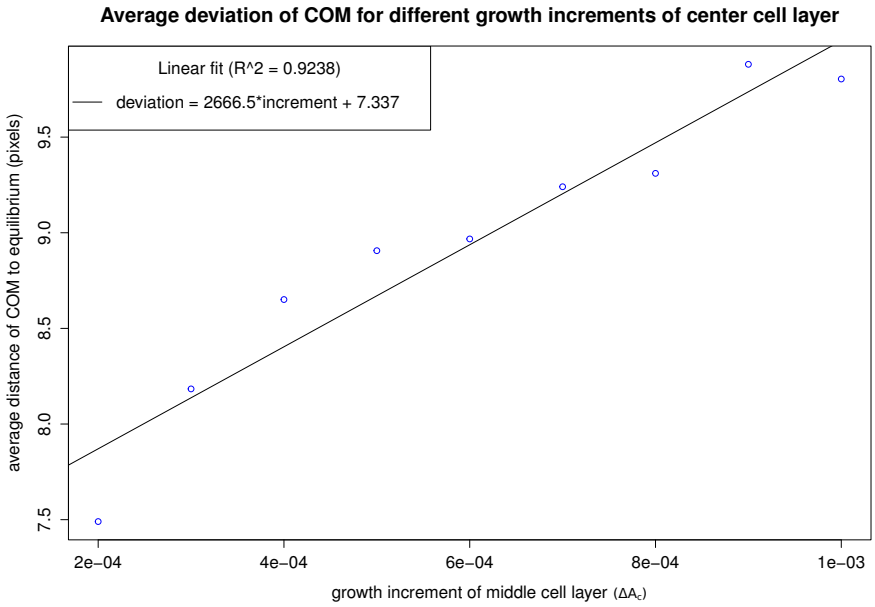


Figure 25: Linear fit for the growth increment and average growth ratio.

Fig. 25 shows the estimation of the average displacement of the COM to the equilibrium of the tissue. The average vertical displacement of the cells increases for a higher growth increment as anticipated, because the cells have a bigger cell volume at the end of a simulation when they grow faster. This increases the vertical displacements and makes the amplitude of the wave bigger. This suggests that growth increments of the middle cell layer influences the amplitude of the tissue, or the tissue folds faster with bigger growth increments.

3.5 Parameter sweep yields various results

A parameter sweep for the model for epithelial folding yields various results, only a few of which are a folding tissue as shown in Fig. 26. In most of the simulations one or more of the cell layers in the tissue break, resulting in individual cell clusters or a broken tissue. Therefore the data of these simulations cannot be used to analyze the properties of the tissue. The simulation results between simulations with the same parameter set are also considerably different, thereby making the analysis meaningless.

General trends resulting from the parameter sweep are already discussed in the previous chapters. Tissues with low growth increments tend to break when bond energies are middle or high. Other simulations result in islands of cluster cells, often with one cell type directed outwards. High growth increments result in large cells, often resulting in the tissue having a rectangular shape as in section 3.2.1. Cell sorting occurs for high bond energies resulting in mosaic cell configurations.

The tissue forms square cells in Fig. 26(a) for low bond energies, where the growth of the cells makes some of the cells of the same type stack onto each other, splitting cluster cells. The tissue folds slightly at the bottom, which is most likely caused by the distribution of the top cells in the tissue. The cells from Fig. (i) are square as well, but only the center cell layer has a high bond energy. High growth of the top cells separates the cells from the middle cell layer that maximize their contact by forming thin lines between the top cells, which results in square top and middle cells. The top cells closely resemble columnar epithelial cells, which is mostly found in the lining of the small intestine (Alberts et al., 2008), and the parameter values can be used to improve the model for epithelial folding.

The tissues from Fig. (b), (d) and (f) seem to fold, but the tissue layer breaks at multiple points, and the wave of the tissue is highly irregular. The tissues from Fig. (c) and (e) have some folds as well. These are probably caused by the top layer, which has high growth and is not uniformly distributed over the tissue. Fig. (c) has cone-like folds because the bond energy between cells of the top layer is low which makes the cell layer hard to bend. This parameter values can be used to model the formation of epithelial or neural tubes for the nervous system.

Figures (g) and (h) are examples of cells sorting into a mosaic. This is caused by the bond energies of the tissue. Some top cells in Fig. (h) are almost completely engulfed by cells of the middle cell layer.

The tissue from Fig. (j) is broken into several pieces. The low bond energy of the top cells along with their high growth makes the parts curve, where the bottom cell layer maximizes its contact area because the bottom homotypic bond energy is lower.

Combinations of somewhat more extreme parameter values do not lead to epithelial folding. This is to be expected because stresses and pressure in the tissue often lead to fractures in the cell layers, which disturbs the stability of the tissue layer.

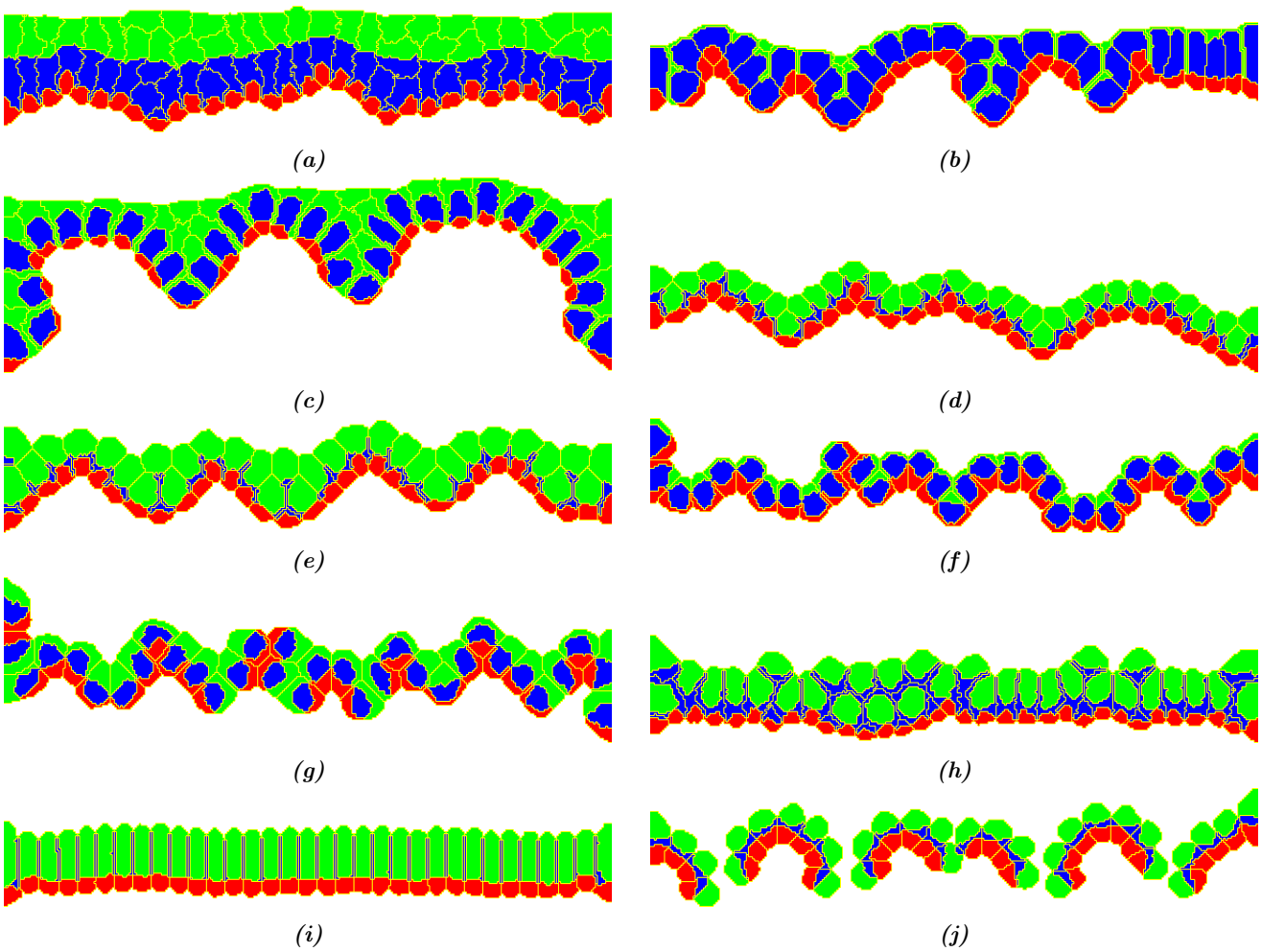


Figure 26: Cell lattice at 100.000 MCS for different parameter sets. *The standard parameter values are* $J_{t,t} = J_{c,c} = J_{b,b} = 40$, $\lambda = 400$ *and* $\Delta A_t = 0.0008$, $\Delta A_c = 0.0006$ *and* $\Delta A_b = 0.0004$. *The deviating parameters are:*
(a) $J_{t,t} = J_{c,c} = 5$, $\lambda = 310$ *and* $\Delta A_t = \Delta A_c = 0.0015$. *The top and center layers have strong bonds and grow fast.*
(b) $J_{t,t} = 5$, $\lambda = 150$, $\Delta A_t = 0.0001$ *and* $\Delta A_c = 0.0015$. *The top layer has strong bonds but grows slow. The middle cell layer grows fast.*
(c) $J_{t,t} = 5$, $J_{c,c} = 75$, $\lambda = 10$ *and* $\Delta A_t = \Delta A_c = 0.0015$. *The top layer has strong bonds, and the bonds of the middle layer are weak. Both layers grow fast, but the penalty for deviations from the target volume is very low.*
(d) $J_{c,c} = 5$, $\lambda = 150$ *and* $\Delta A_c = 0.0001$. *The center layer has strong bonds but grows slow with low resistance to compression.*
(e) $J_{c,c} = 5$, $\lambda = 610$, $\Delta A_t = 0.0015$ *and* $\Delta A_c = 0.0001$. *The simulation is the same as (d), but with high resistance to compression and a fast growing top cell layer.*
(f) $J_{c,c} = 75$, $\lambda = 610$, $\Delta A_t = 0.0001$ *and* $\delta A_c = 0.0008$. *The bonds between center cells are weak. The penalty for deviation from the target volume is high, and the top layer grows slow.*
(g) $J_{c,c} = 75$, $\lambda = 610$ *and* $\Delta A_c = 0.0008$. *The bonds of the middle cell layer are weak and the resistance to comression is high.*
(h) $J_{t,t} = 75$,$J_{c,c} = 5$,$\lambda = 10$, $\Delta A_t = 0.0015$ *and* $\Delta A_c = 0.0008$. *The bonds of the top cell layer are weak, and those of the middle layer are strong. The top cels grow fast, but have a low penalty for deviations from the target volume.*
(i) $J_{t,t} = 75$, $J_{c,c} = 5$,$\lambda = 150$, $\Delta A_t = 0.0015$ *and* $\Delta A_c = 0.0001$. *The bonds between top cells are weak, but the cells grow fast. The middle cell layer has strong bonds but grows slow. (j)* $J_{t,t} = 75$, $\lambda = 310$, $\Delta A_t = 0.0008$ *and* $\Delta A_c = 0.0001$. *The top cells have weak bonds, and the center cells grow slow.*

4 Discussion

The influence of cell properties on the folding and curving of tissue layers has been investigated in this project. Various relations between homotypic bond energies, resistance to compression or growth increments and the behaviour of the tissue were found. Most of the relations are expected and have a biological interpretation, but differential growth is not necessary for tissue folding.

The amplitude of the folds in the tissue is related to different bond energies of the bottom cell layer. This layer is stiff for low bond energies as anticipated, and remains flat resulting in parabolic structures at the top of the tissue. The bottom layer becomes softer for higher bond energies, resulting in a wavelike shape of the tissue as expected.

Stroma cells are connective tissue that supports and connects other tissues in which the epithelial cells are embedded (Alberts et al., 2008). The length of the villi could be explained by the characteristic strength of the stroma if the bottom layer is interpreted as stroma as in Dunn et al. (2012) or Hannezo et al. (2012).

Another factor that relates linearly to the amplitude of the tissue is the growth of the middle cell layer. A higher growth results in bigger cells than expected, that push the bottom and top cell layer as well as the neighboring cell clusters outwards increasing the amplitude of the wave.

The frequency of the tissue is linearly related to the surface energy of the middle cell layer. This relation is probably caused by symmetry of the tissue. The pressure on the layer must be evenly distributed between the top and bottom part of the tissue because the middle layer determines the stiffness of the layer. A low bond energy results in a stiff cell layer which is hard to bend and move, leading to a low frequency wave. High surface energies have a soft cell layer which is easy to bend and lets the tissue relieve pressure by a high frequency wave as anticipated.

Varying the homotypic bond energy of the middle cell layer yields the same results as the basement membrane force used by Dunn et al. (2012) but in opposite direction. A higher basement membrane force yields a stiffer epithelial layer that does not buckle, as shown in 1(b). This confirms the interpretation of the adhesion parameters as the basement membrane force.

The basement membrane is a thin mat of ExtraCellular Matrix (ECM) composed of polysaccharides and proteins that provide strength and anchorage for tissues and cells (Alberts et al., 2008). If the middle cell layer is interpreted as the ECM as in Dunn et al. (2012) or Hannezo et al. (2012), the frequency of epithelial folding could be explained by the strength of the ECM which is influenced by its composition of fiber-forming proteins and elastin molecules.

The tissue behaves as expected for different homotypic surface energies of all cell layers. The layer remains flat for low bond energies because the contacts are strong, and the tissue forms a wave with higher amplitude and frequency when the bond energies are high. The behaviour can be explained using both the relations for different bond energies of the bottom and center cell layer. When the bond energy increases, the frequency and amplitude of the wave increase resulting in epithelial folding. The tissue breaks as expected if the surface energies are very high because of the pressure and weak bonds in the tissue.

The resistance to compression determines the penalty for deviations from the target volume, and is related to the growth of the cells. The cells lag behind their target volume for a low penalty as anticipated, but unexpectedly have a larger growth ratio for intermidiate resistance to compression probably caused by stress in the tissue that stretches the cells. The growth ratio converges to the theoretical value for high λ. The variance of the growth of the cells from their growth ratio decreases for higher resistance to compression as expected. Cells try to fit their target volume for high λ which lowers the deviations from their average growth ratio and results in a lower variance.

The growth ratio can deviate from the theoretical value if the resistance to compression is not too high, giving cells the oppertunity to react to stress and pressure at their local positions. However, a lower resistance to compression results in a higher variance from the growth ratio. It is hard to determine if the deviations from the growth ratio are caused by fluctuations or position in the tissue if the variance is larger. The growth of cells can be linked to their position in the tissue layer if a balance is found between the oppertunity to deviate from the target volume and the corresponding variance of the growth ratio.

Differential growth is an assumption for the model of epithelial folding. This turns out to be unnecessary because the tissue forms a wavelike shape for equal growth between the cell layers as well. This is most likely caused by pressure resulting from growth of the cells in combination with the boundary cells not able to move fast enough to relieve the pressure. A tissue without growth results in a somewhat random shape confirms this. Previous studies by Dunn et al. (2012) and Hannezo et al. (2012) assume a growing epithelial layer with non-proliferating stroma, which results in differential growth. The experiments from Leptin and Grunewald (1990) show that tissue growth in *Drosphila* is not related to epithelial folding which partly supports the results from the simulations.

The parameter sweep suggests that a combination of small deviations of cell properties for epithelial folding result in different cell configurations such as a flat or broken tissue layer. This comfirms that high regulation is required for epithelial folding, and small deviations lead to deformations and diseases. Some of the simulations yield interesting results that can be used to model other biological fenomena such as neural tube formation.

The force fields show that the forces during a simulation are higher than at the end of a simulation. This is to be expected because the tissue is under pressure before folding, which leads to higher forces. This pressure is smaller at the end of a simulation because the tissue has a wavelike shape and cells can move up or downwards to relieve the pressure. The forces are on average higher upwards from the equilibrium of the tissue than downwards for differential growth as anticipated. For higher growth increments the cells exert higher pressure onto their neighbors creating higher forces. These forces are approximately equal for equal growth between the cell layers, confirming that bigger forces are caused by higher growth increments. The forces between cluster cells are on average bigger than between cells of one cluster. However, the forces are smaller between cells of a tissue layer if they have low bond energies. This is probably caused by an equal number of succesful and unsuccesful copy attempts to a neighboring lattice site, which results in a low total force.

4.1 Future research

Multiple directions are available for future research. Most of the initial conditions are kept constant for all simulations. Several bond energies with for example the medium or between cells in a cluster are constant, and other values will change the behaviour of cells and clusters.

The initial cell volume and number of cells can be varied as well, which most likely do not have much influence on the simulation results. A different initial volume leads to the same results because the cells grow during the simulations. Varying the number of cells probably result in better oppertunities for frequency analysis, such as more available frequencies in the tissue wave.

Most epithelial cells are polarized, with cell-matrix anchoring molecules at the ECM side, and a surface of microvilli towards the intestinal cavity. The epithelial cells have different adhesion molecules at their homotypic contacts depending on the distance to the intestinal cavity (Alberts et al., 2008). Differential adhesion for both ends of the epithelial cells realizes a more realistic tissue layer, which can be implemented with clusters of epithelial cells.

The boundary conditions of the tissue can be varied. Periodic boundary conditions probably results in a periodic wave of the tissue which will lead to better results for the Fourier transform because it makes the COM are a periodic dataset.

The resistance to compression is for the model of epithelial folding the same for all cell types. However, different resistance to compression for the cell layers is more realistic, and provides a way to prevent a cell layer from breaking for high bond energies.

Mitosis could be implemented for higher realism of the model for epithelial folding. This adds another dimension to the growth of the tissue layer. Villi formation is linked to increased proliferation at the base of the villi, and implementing mitosis, with for example increased proliferation for locations with high pressure or curvature in the tissue, relates biological villi formation to the computational model.

The current model for epithelial folding is two dimensional. A two dimensional tissue layer, where the cell layers are stacked in the third dimension increases the realism of the model. The layers more accurately represent a biological tissue if implemented, and a two dimensional tissue provides villi to grow in different shapes such as finger-like or herringbone shapes as in Hannezo et al. (2012). The used implementation does not account for the pressure from surrounding cells in the tissue plane, and does not accurately represent a part of the intestine. However, a lot of relations between cell properties and tissue behaviour must be known before a realistic model for a tissue sheet with epithelial folding can be implemented, which makes the current model relevant for future research.

The significance of the results in this thesis are quite low because most parameter sets were ran five or ten times. More simulations of the same parameter set but with different random seeds increases the significance of the results. However, different results between simulations of the same parameter set make analysis of the results difficult. An example is a different wavelength of the tissue for multiple simulations, which increases the variance of the results. It is possible that multiple stable cell lattice configurations exist at the end of a simulation for the same parameter set which could cause the different simulation results.

The data for the force fields has higher potential than the used implementation. The current implementation stores the forces for the entire cell lattice, which makes it nearly impossible to analyze. If the forces are stored for each cell individually, the forces can be used to map the resulting and average force on cells, and gain insight in the pressure on and motion of individual cells through time. The forces do not have a correct order of magnitude for biological relevance, because some scalar factor is missing which is the reason that the forces in this project do not have a unit. The biological forces of cells exerted at adhesion sites are of order picoNewton (Mammoto et al., 2010) opposed to forces of Newton order for the model of epithelial folding. Future research can help to improve the realism of the forces and link them to biological and physical tissue properties.

Mechanical properties of cells and the tissue, such as the Young modulus, elastic modulus and Poisson's ratio can be derived using the forces on the cells. These properties can be used to relate experimental data to the theoretical model of epithelial folding, which can help research to find the cause of diseases. It also provides oppertunities to compare and improve the model of epithelial folding using mathematical relations found in other research.

Previous studies explain tissue folding using regulated gene expression (Leptin and Grunewald, 1990), mechanical properties (Forgacs et al., 1998) or a basement membrane force (Dunn et al., 2012). Other studies show that signalling pathways with morphogens, mechanical forces such as traction forces on adhesion molecules and cytoskeletal cell shape changes drive morphogenetic changes such as tissue folding. The main focus of developmental biology during the last century has been genetic regulation and chemical forces, but mechanical forces exerted on cells by different mechanisms play an important role aswell. It is therefore necessary to study the influence of mechanical forces on developing tissues with experiments and computational models (Mammoto et al., 2010). The implementation of the force field for the model of epithelial folding is an important development to link experimental data to the theoretical model, and can be of great importance in future studies that implement mechanical forces.

References

[1] D. Drasdo and G. Forgacs. Modeling the interplay of generic and genetic mechanisms in cleavage, blastulation, and gastrulation. *Developmental Dynamics*, 219:182–191, 2000.

[2] B. Alberts et al. *Molecular Biology of The Cell*. Garland Science, 2008.

[3] E. Hannezo et al. Instabilities of monolayered epithelia: shape and structure of villi and crypts. *Physical Review Letters*, 107, 2012.

[4] G. Forgacs et al. Viscoelastic properties of living embryonic tissues: a quantitative study. *Biophysicall Journal*, 74:2227–2234, 1998.

[5] J.R. Spence et al. Vertebrate intestinal endoderm development. *Developmental Dynamics*, 240:501–520, 2011.

[6] M. Swat et al. *CompuCell3D Reference Manual, Version 3.6.2*.

[7] M.A. Wyczalkowski et al. Computational models for mechanics of morphogenesis. *Birth Defects Research Part C: Embryo Today: Reviews*, 96:132–152, 2012.

[8] S-J. Dunn et al. Modelling the role of the basement membrane beneath a growing epithelial monolayer. *Journal of Theoretical Biology*, 298:82–91, 2012.

[9] J.A. Glazier and F. Graner. Simulation of the differential adhesion driven rearrangement of biological cells. *Physical Review E*, 47:2128–2154, 1993.

[10] L. Greengard and J-Y. Lee. Accelerating the nonuniform fast fourier transform. *Society for Industrial and Applied Mathematics*, 46:443–454, 2004.

[11] M. Leptin and B. Grunewald. Cell shape changes during gastrulation in drosophila. *Development*, 110:73–84, 1990.

[12] T. Mammoto and D.E. Ingber. Mechanical control of tissue and organ development. *Development*, 137:1407–1420, 2010.

[13] B. L. Welch. The generalization of 'student's' problem when several different population variances are involved. *Biometrika*, 34:28–35, 1947.

[14] B.L. Zaitlen. Introduction to compucell3d. *The Biocomplexity Institute*.

5 Appendix

5.1 Constant variables during simulations

The bond energies between cluster cells and between cells and medium are kept constant during the simulations. The bond energies of the tissue are symmetrical for the bottom and top cell layer, as shown in Table 5. The bond energies between top and bottom cells are high compared to the bond energies with the center cells to prevent mixing of cells. The medium is free to move, and does not add to the Hamiltonian.

Table 5: *All bond energies for medium and cells of the tissue during simulations.*

Bond energy between clusters	Value
Medium - Medium	0
Top - Medium	30
Center - Medium	100
Bottom - Medium	30
Top - Top	$J_{t,t}$
Center - Center	$J_{c,c}$
Bottom - Bottom	$J_{b,b}$
Top - Center	100
Top - Bottom	200
Center - Bottom	100

Bond energy inside clusters	Value
Top - Center	10
Top - Bottom	25
Center - Bottom	10

5.2 Template simulation

The XML file defines the basic properties of the CPM, cell types, bond energies between the cell types and clusters, the file containing the initial cell latice and where to store the cell latice during a simulation. It also contains the settings for the force field implementation as shown in Listing 1.

Listing 1: *XML code for force field implementation*

```
    <ForceFieldFrequency>10000</ForceFieldFrequency>
    <ForceFieldName>[SIMID]</ForceFieldName>
    <ForceFieldCount>5</ForceFieldCount>
```

The Python scripts contain the code that is executed every MCS to modify parameters or calculate and store data. The Python scripts also contain steppables that describe the additional conditions for the CPM such as the change of the target volume for the cells and the definition of the resistance to compression of the cells. The Python code essential for the model of epithelial folding is shown in Listing 2. The Python file that generates the simulation files replaces "[*parameter name*]" with the correct parameter value. Other steppables check if the tissue layer is intact or calculate and store the growth and COM of the cells.

Listing 2: *Steppable for growth increments*

```
def step(self,mcs):
  for compartmentList in self.clusterList:
    for cell in CompartmentList(compartmentList):
      # growth increments
      if cell.type==1:
        cell.targetVolume+=[TVOLUMETOP]
```

```
      elif cell.type==2:
        cell.targetVolume+=[TVOLUMECENTER]
      elif cell.type==3:
        cell.targetVolume+=[TVOLUMEBOTTOM]

      # resistance to compression
      cell.lambdaVolume= [LVOLUME]/cell.surface
```

5.3 Force field implementation

The change of the Hamiltonian for each copy attempt is needed to calculate the force on each lattice site. The `potts3D.cpp` file contains the code for copy attempts and monte carlo steps, and the function `createCellField` has been modified which is responsible for initialization of the used variables. The code that has been added is shown in Listing 3.

Listing 3: *Modifications of the createCellField function*

```
  forceFieldFrequency = 0;
  forceFieldName = "";
  forceFieldCount = 1;
  for (int i = 0; i < dim.x; i++) {
    std::vector<Point3Df> row;
    for (int j = 0; j < dim.y; j++) {
      row.push_back(Point3Df(0,0,0));
    }
    forceField.push_back(row);
  }
```

The `metropolisFast` function is altered to sum all the forces and save them in a data file ready for analysis, as shown in Listing 4. The change of the Hamiltonian is availalble in the `metropolisFast` function, and the distance of the copy attempt as well.

Listing 4: *Alterations to the metropolisFast function*

```
if (forceFieldFrequency != 0 && (sim->getStep()+forceFieldCount)
                    % forceFieldFrequency < forceFieldCount) {
  forceField[n.pt.x][n.pt.y] += (Point3Df(n.pt) - Point3Df(pt)) *
                    ( ( -(float)change) / ((float)n.distance * (float)n.distance) );
}

if (forceFieldFrequency != 0 && (sim->getStep()+1) % forceFieldFrequency == 0) {
  ostringstream outForceFieldStreamR;
  outForceFieldStreamR << "forceFieldR_" << forceFieldName
                    << sim->getStep() << "." << "txt";
  ofstream* outForceFieldR = new ofstream(outForceFieldStreamR.str().c_str());

  ostringstream xs; ostringstream ys;
  for (int i = 0; i < forceField.size(); i++) {
    for (int j = 0; j < forceField[0].size(); j++) {
      if ((int)forceField[j][i].x != 0 || (int)forceField[j][i].y != 0) {
        xs << j << ",";
        ys << i << ",";
      }
    }
  }
```

```
  ostringstream dxs; ostringstream dys;
  for (int i = 0; i < forceField.size(); i++) {
    for (int j = 0; j < forceField[0].size(); j++) {
      if ((int)forceField[j][i].x != 0 || (int)forceField[j][i].y != 0) {
        dxs << (((int)(forceField[j][i].x*100.0f))/(100.0f*forceFieldCount)) << ",";
        dys << (((int)(forceField[j][i].y*100.0f))/(100.0f*forceFieldCount)) << ",";
        forceField[j][i] = Point3Df();
      }
    }
  }

  *outForceFieldR << "x=c(" << xs.str();        *outForceFieldR << "0)" << endl;
  *outForceFieldR << "y=c(" << ys.str();        *outForceFieldR << "0)" << endl;
  *outForceFieldR << "dx=c(" << dxs.str();      *outForceFieldR << "0)" << endl;
  *outForceFieldR << "dy=c(" << dys.str();      *outForceFieldR << "0)" << endl;
  *outForceFieldR << "df <- data.frame(x,y,dx,dy)" << endl;
  *outForceFieldR << " " << endl;
}
```

Code has been added to the `update` function to give the user the oppertunity to set force field parameters, as shown in Listing 5. The available parameters are the frequency at which the forces are to be stored, the name of the files and the number of MCS that are used to calculate the forces.

Listing 5: *Variables can be set with the XML file of the simulation*

```
if(_xmlData->getFirstElement("ForceFieldFrequency")) {
  if(forceFieldFrequency != _xmlData->getFirstElement("ForceFieldFrequency")->getUInt()){
    forceFieldFrequency = _xmlData->getFirstElement("ForceFieldFrequency")->getUInt();
    cerr << "ForceFieldFrequency = " << forceFieldFrequency << endl;
  }
}
if(_xmlData->getFirstElement("ForceFieldName")) {
    forceFieldName = _xmlData->getFirstElement("ForceFieldName")->getText();
    cerr << "ForceFieldName = " << forceFieldName << endl;
}
if(_xmlData->getFirstElement("ForceFieldCount")) {
  if(forceFieldCount != _xmlData->getFirstElement("ForceFieldCount")->getUInt()){
    forceFieldCount = _xmlData->getFirstElement("ForceFieldCount")->getUInt();
    cerr << "ForceFieldCount = " << forceFieldCount << endl;
  }
}
```

5.4 Simulation analysis script

The statistical package R is used to analyze the simulation data. The data is read from the stored files, and if needed transformed with the NDFT as shown in Listing 6. The code for all simulations with the same parameter set is shown. The data is stored in a matrix ready for statistical analysis.

Listing 6: *R script for loading and transforming simulation data*

```
for (j in 1:length(total_sim)) {
    com <- source(paste(rbind(filePath,"/",sim_name[j],
                    "/COM_100000.txt"),collapse=""))$value
```

```
    # move equilibrium to 0
    com[,2] <- com[,2] - mean(com[,2])

    # non-uniform discrete fourier transform
    e = 2.71828182
    b = c(1:length(com[,1]))
    for (k in 1:length(com[,1])) {
      b[k] = sum( com[,2]*e^(-complex(1,0,1)*2*pi*(k-1)*com[,1]
                    /(max(com[,1])-min(com[,1]))))/33
    }
    com <- Mod(b)

  # collect all values from simulation
  simndft[j,] <- com[,2]
}

assign(paste(rbind("mat",param_set),collapse=""), val)
```

The matrix with simulation data is analyzed with the Welch t-test. The code for the hypothesis test for two parameter sets i and j is shown in Listings 7. The p-values for all growth ratios or frequencies are printed.

Listing 7: *R script for hypothesis test*

```
# get variables
mat_names = grep("mat", ls(), value=TRUE)
param1 <- eval(parse(text=mat_names[i]))
param2 <- eval(parse(text=mat_names[j]))

# set standard deviation, sample size, degrees of freedom
s1 <- apply(param1,2,sd)
s2 <- apply(param2,2,sd)
n <- sampleSize
dfm <- (s1^2/n + s2^2/n)^2/((s1^2/n)^2/(n-1) + (s2^2/n)^2/(n-1))
s <- sqrt(s1^2/n+s2^2/n)

# perform t-test and calculate p-values
t <- (apply(param1,2,mean)-apply(param2,2,mean))/s
print(round(2*pt(-abs(t),df=dfm), 10))
```