2011

# A survey of the use of cellular automata and cellular automata-like models for simulating a population of biological cells

Jeremy D. Knutson
*Iowa State University*, knutsonj@iastate.edu

**A survey of the use of cellular automata and cellular automata-like models for simulating a population of biological cells**

by

Jeremy Knutson

A thesis submitted to the graduate faculty

in partial fulfillment of the requirements for the degree of

MASTER OF SCIENCE

Major: Mathematics (Applied Mathematics)

Program of Study Committee:

Michael Smiley, Major Professor

Anastasios Matzavinos

James Evans

Iowa State University

Ames, Iowa

2011

# TABLE OF CONTENTS

# LIST OF TABLES

# LIST OF FIGURES

# ABSTRACT

The purpose of this thesis is to survey a few of the cellular automata and cellular automata-like models which have been used in mathematical biology applications in recent years. Some of the benefits of using such models will be discussed. Also some issues pertaining to the analysis of such models will be addressed.

# CHAPTER 1. A survey of mathematical modeling in biology

*"A model is an object or concept that is used to represent something else. It is reality scaled down and converted to a form we can comprehend."* [1]

## 1.1 Introduction

Models come in a wide variety of forms. A familiar type of model is the hobbyist's wood or plastic scale replica. Take, for instance, the scale model airplane. When the hobbyist builds this replica he or she gains an understanding, in as far as the model is an accurate representation of the real airplane, without ever having had "hands on" experience with the assembly of a real plane.

The mathematician is also conversant with models and modeling. A mathematical model is a model built from mathematical concepts: constants, variables, functions, equations, etc. The design of such mathematical models is a chief area of study within applied mathematics, and is of no small importance. In fact, according to Shier and Wallenius, "The greatest success story of applied mathematics (and perhaps all of science) over the last three centuries has been its keen ability to model the *laws of nature* and to use these constraints to assist in traditional engineering problem solving."[2] (emphasis in the original) Although this golden age of mathematical modeling is relatively recent, mathematical models date back to ancient times. For many hundreds of years, people have modeled fluid flow (to better understand crop irrigation) and ballistics; for thousands of years they have modeled the movement of heavenly bodies. A notable astronomical model, *Almagest* of Claudius Ptolemy, dates back to 140 c.A.D.

While mathematical modeling originated in the physical sciences, it is certainly not restricted there. In recent years applied math has become increasingly important to the biologi-

cal sciences, especially as molecular biology has gained more and more preeminence. It is not difficult to understand why, when one considers the size of cells. Since laboratory experimentation may not always be possible or practical, a mathematical model may be used to help the biologist study cell behavior.

The amount of biological data that is now being collected is overwhelming; so much so that the "ability to collect new data outstrips our ability to heuristically reason mechanisms of cause and effect in complex systems." It is applied mathematics that gives the researchers the hope to make sense of all this data because "[mathematical modeling] allows us to formalize the cause and effect process and tie it to the biological observation."[3]

## 1.2 Mathematical Biological Problems

Biological problems have been studied using mathematics for hundreds of years, but for much of history these studied problems have been restricted to ecological modeling (i.e. population dynamics). Mathematical modeling for population dynamics has existed at least since the work of Thomas Malthus, during the early 19th century, after which the Malthusian growth model is named. (Fibonacci also gave a much idealized population model many centuries earlier, which will be mentioned in section 1.2.1).

Models of morphogenesis, that is the study of the biological processes that govern the development of an organism's shape, are perhaps the next oldest models within mathematical biology. D'arcy Thompson was a pioneer in this field, and is best remembered for his work *On Growth and Form* (1917). In his work Thompson attempted to explain changes in organism shape using mathematical concepts. For example, single celled organisms' shapes were studied as a problem of surface curvature minimization; the change in shape of several species of fish was exlpained through a transfromation by an affline mapping. Alan Turing also did early research in this field, using a reaction-diffusion system to to study the self-organization of biological pattern formation.[6]

Of course ecology and morphogenesis make up only a small part of the wide spectrum of fields in the life sciences. However, it was not until relatively recently that mathematical modeling would be applied to wide variety of other areas of biology. There are now numerous

ways that mathematics is applied to biological problems, ranging from population models, to epidemiology and medicine. Predator/prey models predict the populations within populations of animals having this predator/prey relationship. Tumor growth and virus behavior can be described. Neuron 'firing' can be studied with models like the Hodgkin-Huxley equations. These advancements all grant scientists the ability to predict biological phenomena.

### 1.2.1 Modeling methods employed in mathematical biology

This section highlights a few of the various methods of modeling employed within mathematical biology, including difference equations, ordinary differential equations, partial differential equations, and cellular automata.

A difference equation relates the value of some quantity $P$ at time $k+1$ to the value of quantity $P$ at previous times, for example $P_{t+1} = F(P_t)$, where $F(x)$ is a real valued function.[8] Such a model is especially well suited to modeling populations in which generations do not overlap (i.e. one generation dies out to be completely replaced by its progeny). Take, for example, annual plants, which die every year, leaving behind seeds which become the next generation the following growing season. If we assume that all seeds that survive the winter will germinate at the beginning of the next growing season (an oversimplification that is not actually that hard to correct), then a difference equation that models such a population is is solved to give the discrete growth equation $P_{k+1} = fsP_k$, where $f$ is the number of seeds produced per plant every year and $s$ is the proportion of seeds which will survive a winter. This can also be expressed as $P_k = (fs)^k P_0$.[7]

A particularly famous difference equation is the one which is used to determine the Fibonacci sequence, $P_k = P_{k-1} + P_{k-2}$. In fact, it was because of a biological problem that that sequence received its name. In 1202 A.D. Fibonacci studied the sequence in relation to the growth of an idealized population of rabbits..

The models described above are discrete Malthusian growth models. A more familiar form of Malthusian growth is the continuous case, which gives us the ordinary differential equation (ODE) $\frac{dP}{dt} = kP$, where $k$ is the growth rate. This model predicts that a population grows exponentially, which can only be the case in an environment with unlimited resources. More

realistically a popultion will follow a logistic curve $\frac{dP}{dt} = kP(1-P/K)$, where $K$ is the carrying capacity. In this model, growth of a population slows as the population grows large (due to higher competition for scarce resources).

Another famous ODE is the Lotka-Volterra equations (predator-prey equations), which also models populations, specifically populations of predators and prey which are dependent on each other. According to the Lotka-Volterra equations the number of predator animals $y$ and the number of prey animals $x$ should follow the equations $\frac{dx}{dt} = x(\alpha - \beta y)$ and $\frac{dy}{dt} = -y(\gamma - \delta x)$ where $\alpha$, $\beta$, $\gamma$, and $\delta$ are positive parameters representing the interaction of the two species; $\beta$ is the negative effect that a population of predators has on a population of prey, $\gamma$ is the negative effect that a predator population has on itself, as predators compete for prey.

Systems of ordinary differential equations are also used in epidemiology. A simple example of such a system is the SIR (susceptible, infectious, recovered) model. The system of equations is

$$\begin{aligned}\frac{dS}{dt} &= -\beta IS \\ \frac{dI}{dt} &= \beta IS - \nu I \\ \frac{dR}{dt} &= \nu I\end{aligned}$$

which may be used to study the dynamics of a single epidemic outbreak. In this model $S$ represents the number of susceptible individuals, $I$ represents the number of infected individuals, and $R$ represents the number of individuals in the recovered state and $\beta I$ represents the force of the infection.

A similar partial differential equation (PDE) describes an epidemic in greater detail. Assuming that a biological system or phenomenon depends only on time may neglect other important dependencies. Although the dynamics of an epidemic are time dependent, the susceptibility of a person to contracting a disease are usually age dependent. The individuals' ages may be too important of a variable to leave out of the SIR model; the SIR PDE (which is not simply a differential equation, but an integro-differential equation) adds this detail to the previously discussed ODE. The equations describing the number of indiviuals at time $t$ of age $a$ in each

of the three groups of interest are:

$$\frac{\partial}{\partial t}s(t,a) + \frac{\partial}{\partial a}s(t,a) = -\mu(a)s(a,t) - s(a;t)F(a,t,i(t,a_1))$$
$$\frac{\partial}{\partial t}i(t,a) + \frac{\partial}{\partial a}i(t,a) = -\nu(a)i(a,t) + s(a,t)F(a,t,i(t,a_1))$$
$$\frac{\partial}{\partial t}r(t,a) + \frac{\partial}{\partial a}r(t,a) = -\mu(a)s(a,t) + \nu i(a,t)$$

where $F(a,t,i(t,a_1))$ is the force of the infection and $a_1$ is some minimum age being considered. The total population in the groups susceptible, infected, and recovering at any time $t$ is given by

$$S(t) = \int_0^{a_M} s(t,a)da$$
$$I(t) = \int_0^{a_M} i(t,a)da$$
$$R(t) = \int_0^{a_M} r(t,a)da$$

where $a_M$ is a max age. [9]

Difference equations and differential equations have a long standing position in mathematics, having been studied for hundreds of years. The focus of the remainder of this paper is a comparatively recent model, known as a cellular automaton. Cellular automata are discrete models. The focus, and namesake, of a cellular automaton is the state of a grid cell in the lattice of grid cells which makes up the model. Each cell in the lattice has a finite number of possible states, such as on or off. Also, for each cell, there is a set of cells which make up this cell's neighborhood (for example, in a two-dimensional rectangular lattice, a possible neighborhood of a cell is the set consisting of the four cells which are above, below, to the right, and to the left of the cell). These models have a discrete time variable and a mathematical rule for how the state of a given cell should change based on the states of the cells in the cell's neighborhood, as time progresses.

Since some time has been spent discussing SIR models, it is interesting to note that cellular automata have also been used in modeling epidemics. Each grid cell could be used to represent individuals in the population, and these cells then have the elementary states susceptible, immune or recovered. In fact, the Greenberg-Hastings Automata, used to model epidemics, is related to the previously discussed SIR model.

When used as a model of infectious disease each cell in the Greenberg-Hastings Automata represents a single individual or a single region. The three possible states a cell may have are representative of whether that individual is susceptible, infectious or immune. Simple rules govern the change of the states of the cells. If a susceptible cell has at least one infected cell adjacent to it orthogonally (or possibly also diagonally, depending on the definition of the neighborhood), that cell becomes infected. An infected cell remains infected for a certain number of time steps, then becomes immune for a certain number of time steps, before becoming susceptible again.

The lengths of time for which a cell remains infected or immune are related to the parameters of a continuous case SIR model. One key difference, however, between the continuous ODE and PDE models and the Greenberg-Hastings model is that the SIR ODE and PDE assume that every susceptible individual has an equal chance of becoming infected. In the Greenberg-Hastings model, only a susceptible indiviual with an infected neighbor may become infected. This is due to an aspect of CA which may be called *locality*, i.e. the state of a cell (at any given time) is dependent on only the state of that cell's (local) neighbors (at the previous time). This is an important aspect driving the interesting pattern formation displayed by cellular automata.

Today, just as there are a wide variety of model types, there is a wide variety of biological problems being studied through mathematical modeling. Tumor growth is an important problem being modeled, as well as blood vessel formation (angiogenisis), as it relates to tumor growth.[10] Morphogensis continues to be an interesting problem and cellular automata are being used as a tool to study cell sorting.[11] Cellular automata can be used to model individuls in population ecology, this can be applied to immunology if the cellular automaton is made to represent a microbial population.[12] Just as there are a wide variety of patterns a CA may display, in nature, orgnanism markings show a wide variety, and cellular automata are being used to study these pigment pattern formations.[13] These are only a small sample of the applications of cellular automata to biological problems from recent years. Some of these will be discussed in greater detail in the chapters to follow.

## 1.3 Model considerations

There are a number of considerations that go into the design of a good model. Should the variables be continuous or discrete? Should the model be deterministic or should it have a stochastic element? How much of the real life biological complexity can/should be represented in the model; which variables ought to be omitted so that the model retains simplicity in its use? The answers to these questions are not always clear, but the way in which some researchers have answered them will be considered in the next few chapters.

The majority of this thesis will pertain to models with some similar charactaristics. The models that will be discussed are discrete in space and time, have a probabalistic element, and treat the biological cell as the smallest biological unit.

Discrete models in mathematical biology are of interest because many biological structures are discrete by nature (e.g. cells in a tissue), and thus are most naturally modeled by a discrete model.[3] In many cases it may be quite difficult to model these discrete structures using a continuum approach. For example, of the problem of modeling interactions between cells, one researcher says, "Certain cell processes, however, such as cell-cell adhesion would be difficult, if not *impossible*, to model at the continuum level." (emphasis added.)[14]

Biological phenomena are difficult to predict with absolute certainty; a stochastic model reflects this uncertainty. Many classical biological models would lead us to believe that biological processes are strictly deterministic: "current scientific wisdom views biological systems as essentially deterministic in character, with dynamics entirely predictable given sufficient knowledge of the state of the system (together with complete knowledge of the physics and chemistry of the interacting biomolecules)."[5] Given this statement, why shouldn't a deterministic approach be used in modeling any biological system; why is stochastic modeling necessary? The reason stochastic modeling is necessary is found in the words 'sufficient knowledge' and 'complete knowledge'.

A model with 'sufficient' and 'complete' detail is overly complicated and must be simplified. This simplified model is no longer deterministic, and thus it may be appropriate to introduce a probabilistic element to the model reflect this fact.

In modeling an organism, organ, or tissue there is certainly a question as to what level of detail to use. In this thesis evidence will be given that the cellular level may be the appropriate level of detail. According to Merks and Glazier "nature's solution to data hiding is the cell"[4]. What this means is that the cell can be treated as a "black box" module; that is, the inner workings of the cell can be ignored to focus on how the cell interacts with other cells and its environment.In the same paper, Merks and Glazier pose the question "If nature itself uses individual cells as an abstraction, why should we make our biological models more complicated by describing tissue-level structures in terms of subcellular behaviors, which do not affect it directly?"

# CHAPTER 2. Cellular Automata

Many models fail to properly describe biological structures at the cellular (and thus also, subcellular) level. Cellular automata, on the other hand, can provide a very good description of biological structure at both the cellular and subcellular levels.[14]

## 2.1 What is a Cellular Automata?

Cellular automata are discrete models in which the states of the variables, i.e. values associated with grid cell locations, are driven by simple rules dependent on the states of the neighbors of each variable. The definition of an cellular automaton includes:

1. a definition for a grid, which includes boundary conditions. Often, to avoid complications due to a boundary, periodic boundary conditions are used, so that a two-dimensional grid is the surface of a torus.

2. a finite (usually small) set of states that grid cells can have.

3. a neighborhood, which is a definition of which nearby cells may affect the state of a given grid cell

4. a local rule, by which a grid cell's state may change. This rule may be either deterministic, or in the case of a stochastic cellular automata, have a probabilistic element.

A strategy for updating the grid must also be defined. Will the grid be updated synchronously or asynchronously? Usually a synchronous update, where all updates to the grid are applied at the same time, is employed. In an asynchronous update individual cells update indiviually, thus the new state of a cell immediately affects the calculation of the state of a neighbor. In

many of the cellular automata with biological applications a certain subset of the all the cells may be randomly chosen to synchronously update on a given time step.

For the sake of clarification consider the simple (and famous) cellular automata, known as Conway's Game of Life, or simply the Game of Life. Originally 'played' on a Go game board, the Game of Life may be defined on any (usually a large) grid. Each grid cell exists in one of two possible states, alive or dead. The neighborhood of a given grid cell is made up of the eight next-nearest neighbors which are orthogonally adjacent and diagonally adjacent to that cell (this is sometimes called a Moore neighborhood). Each grid cell will be updated synchonously every time step according to the following rules:

1. If a living cell has less than 2 living neighbors it dies (as if by loneliness).
2. If a living cell has more than 3 living neighbors it dies (as if by overcrowding).
3. If a dead cell has exactly 3 living neighbors it becomes alive.

To begin, each cell in the grid is initialized with either the state living or dead. Some of the initial configurations which give rise to some of the more interesting patterns have been given names, such as 'glider', 'small exploder', and 'lightweight spaceship'.[15] Figure 2.1 gives images of five time steps of the Game of Life beginning with the initial configuration known as 'glider'. This example shows how some of these interesting patterns repeat themselves on successive time steps.

## 2.2 History

Cellular automata trace their beginning to the Los Alamos Laboratory in the 1940s, where mathematician John von Neumann was studying the concept of self-replicating robots. His idea was that moons or asteroids would be most efficiently mined by such automata, due to the exponential growth of their population. The cost of experimenting with such robots, however, was prohibitive. Stanislaw Ulam, also at Los Alamos, was working on the problem of crystal growth using a lattice model. At his suggestion von Neumann applied a lattice grid model

also to the problem of self-reproducing automata. This model became what is now known as cellular automata.

What von Neumann invented with pen and paper was later popularized by the the automaton known as the Game of Life, which is described above. Invented in the 1970s by John Conway, this model displays a wide variety of complex patterns despite its very simple rules for whether a given cell should 'live' or 'die'. This simple CA caught the attention of researchers in a wide variety of fields including computer science, physics, biology, economics, and mathematics.

In the 1980s Stephen Wolfram published a number of papers detailing his study of the universality of cellular automata and the complexity of their patterns. In 2002, after having sudied cellular automata for decades, Stephen Wolfram published a 1280 page text on the subject of their simple rules and complex patterns. This fact, that cellular automata can display complex patterns despite having simple rules, is a reason that they are considered to be such a useful model type.

At the University of Siegen, Düchting and Vogelsaenger did some of the the earliest research in the area of using cellular automata to model tumor growth. Their 1984 paper describes a three-dimensional simulation of tumor growth and describes what its application to tumor treatment might be.

## 2.3 Biological Motivation

Due to the locality of the rules which drive them, CA are a useful tool in studying pattern formation. This characteristic of CA makes them desirable models for biological systems and phenomena.[3] In section 1.2.1 a couple of ways to model the spread of an infection were discussed, inculding a CA. In the actual spread of an infection an uninfected individual (or region) will become infected only if the infection comes sufficiently close to itself (within its 'neighborhood'); neither the ODE or the PDE model can express this aspect of an epidemic. On a cellular scale, CA have been used to study morphogenesis by modeling cell differentiation[16] also they have been used to study the invasion of cancerous cells.[17]

A number of researchers have shown that in modeling cellular level behavior by CA and CA-

like models, realistic tissue level pattern formation can be observed in simulations (as will be discussed throughout the remainder of this thesis.) Modeling biological phenomena at mutliple scales is an important current problem and will be discussed in greater detail in chapter 4.

## 2.4 A case for cellular automata as a cellular scale model

It is an argument of this thesis that CA have characteristics that make them good models of a population of biological cells, but do they give results consistent with established biology theory? One effort to answer this question is presented in a 1993 paper by An-Shen Qi and co-workers. In this apper they compare results from a CA simulation to results from a more standard modeling approach.

The intention of the design of the model presented in this paper was for it to be a microscopic explanation of the purely phenomenological Gompertz model. Used since the 1960s, the Gompertz model is a well-established model of tumor growth, which says that tumor growth should follow the equation $X'(t) = \alpha \log(\frac{K}{X(t)})X(t)$, with the solution $X(t) = K \exp(\log(\frac{X(0)}{K}) \exp(-\alpha t))$, where $K$ is the carrying capacity and $\alpha$ is a constant related to the proliferation of tumor cells. Research has given evidence that this model is consistent with tumor growth.[18] .

The model presented by Qi et al. models the invasion of cancerous cells in a population of normal cells. Each grid cell represents a single biological cell; the state of each of these cells is one of the following: normal, cancerous, cancerous and bound by a white blood cell (also called a complex), or dead cancerous. The neighborhood of each cell is defined as the four nearest neighbors orthogonally adjacent to the cell.

If at least one of the cells in the neighborhood of a cancerous cell is normal (although in vitro this requirement is not a necessity), then that cancerous cell may proliferate with some probability $\tilde{k}_1 = k_1(1 - N_c/\phi)$, where $k_1$ is the proliferation rate in vitro, $N_c$ is the current number of cancerous cells and $\phi$, a constant, is the maximum possible number of cancer cells. If a cancerous cell proliferates, then the state of one of the normal cells in its neighborhood is changed to cancerous. This represents the invasion of cancerous cells into normal tissue.

It is also possible for a cancerous cell to become a complex; that is, a white blood cell may bind to a cancerous cell. A cell in the cancerous state changes to the complex state with

probability $k_2$. If a grid cell is in the complex state it changes to the dead state with probability $k_3$. Finally, if a grid cell is in the dead state it may change to the normal state with probability $k_4$, which represents the invasion of normal cells into a cancerous mass.

As part of the evidence of the legitimacy of their CA model they compare a plot of the number of tumorous cells in their model to a plot of the Gompertz curve. The plot of the data taken from their model follows the Gompertz curve quite closely. As the authors put it, "The agreement is clearly acceptable." It can be concluded that this is a promising model, since it matches previously established theory. Besides comparing the results of the CA to established biological theory, the authors also made use of a least-squares curve fitted to experimental data to establish that the model also agrees with biological observations. As the data from the CA agrees in both cases the intention of the design of this model was met. The authors also argue that this model is more realistic than some other established models, such as the Eden growth model, which assumes that a tumor only grows from its surface.

Although the model was successful in its purpose some researchers have since argued that the model neglects some important biological considerations, such as the influence of growth stimulants and inhibitors on tumor growth.[19] This will be discussed in greater detail in Chapter 4. Another potential weakness of a CA is that it neglects to take into consideration the unique sizes and shapes of the biological cells it models. This point will be further discussed in Chapter 3.

## 2.5 Mathematical theory and cellular automata

Although CA follow simple rules, mathmatical analysis of a CA may not be straightforward. First of all, there is the problem of classifying cellular automta. Stephen Wolfram's set of four CA classes is probably the most popular method of CA classififacation. It is, however, strictly qualitative and sufferes from a degree of subjectivity. According to this classification, all cellular automata may be categorized as being in one of the four Wolfram classes found in Table 2.1.

All cellular automata fit into on of these four categories, however it is not always clear which category a CA ought to be placed in. A more rigorous definition of these classes is given by a paper by Culik and Yu.[21] In the paper, the authors prove that the process of classifying

Table 2.1 Wolfram's CA Classifications[20]

| | |
|---|---|
| Class I: | these CA have the simplest behavior; almost all initial conditions result in the the same uniform final state |
| Class II: | different initial conditions yield different final patterns, but these different patterns consist of an arrangement of a certain set of structures, which stays the same forever or repeats itself within a few steps |
| Class III: | behavior is more complicated and appears random, but some repeated patterns are usually present (often in the form of triangles) |
| Class IV: | in some respects these are the most complicated class; these behave in a manner somewhere inbetween Class II and III, exhibitng sections both of predictable patterns and of randomness in their pattern formation |

a cellular automaton according to Wolfram's classes is an undecidable problem, i.e. it is not possible to design an algorithm to decide on a class membership for a given CA.

### 2.5.1 Markov Processes

The behavior of a CA may be thought of as a Markov process. A Markov process, named for the mathematician Andrey Markov, is a stochastic process that exhibits memorylessness, also known as the Markov property. This property of the process means that future probabilities of an event may be determined from the probabilities of events at the current time. A process has this property if the following equations holds:

$$P(X(t_n) = x \,|\, X(t_1) = x_1, X(t_2) = x_2, \ldots, X(t_{n-1}) = x_{n-1}) = P(X(t_n) = x_n \,|\, X(t_{n-1}) = x_{n-1})$$

The term Markov chain is sometimes used to refer to a discrete time Markov process.

One tool that is commonly used to study discrete time Markov chains is the transition matrix $P$:

$$P = \begin{pmatrix} p_{11} & p_{12} & p_{13} & \cdots \\ p_{21} & p_{22} & p_{23} & \cdots \\ p_{31} & p_{32} & p_{33} & \cdots \\ \vdots & \vdots & \vdots & \end{pmatrix}$$

where $p_{ij}$ represents the probability that the system change from state $i$ to state $j$ in one time step.

Lets consider a simple one-dimensional two-celled cellular automata (oriented horizontally). Define 2 possible states for each cell: living or dead. The system updates according to simple rules. When a living cell is selected for update it will do one of four things: 1.) give 'birth' (the adjacent cell comes to life) and continue living with probability $p = .25$, 2.) give 'birth' and die with probability $p = .25$, 3.) remain quiescent and continue living with probability $p = .25$, or 4.) remain quiescent and die with probability $p = .25$.

This system has $2^2 = 4$ possible states (base 2, since each cell has 2 possible states, to the second power since there are 2 cells): state 1: all cells are dead, state 2: the cell on the left is living and the cell on the right is dead, state 3: the cell on the right is living and the cell on the left is dead, or state 4: both cells are living. The probability transition matrix is as follows:

$$P = \begin{pmatrix} 1 & 0.25 & 0.25 & 0 \\ 0 & 0.25 & 0.25 & .25 \\ 0 & 0.25 & 0.25 & .25 \\ 0 & 0.25 & 0.25 & .5 \end{pmatrix}$$

(In this example the state where both cells are dead is called an "absorbing state", since it is impossible for the system to leave that state.)

This may seem be a good strategy to study a CA on a small grid. However, many of the CA used to study biological systems may be set in a 100 x 100 grid (or larger), which means that the system has $2^{100 \cdot 100}$ possible states and the probability transition matrix is a $2^{100 \cdot 100}$ x $2^{100 \cdot 100}$ matrix. Even though the matrix will likely be sparse, it is probably not a practical way to study the CA.

So what strategy ought to be used to analyze a CA? Some other methods of analysis which are related to Markov chain theory are the Chapman-Kolmogorov equation, which will be addressed in section 2.5.2 and mean-field theory, which will be addressed in section 2.5.3.

### 2.5.2 Chapman-Kolmogorov

Informally the Chapman-Kologorov equation, as it relates to a discrete time Markov chain, states that the probability of system in state $a$ going to state $s$ is equal to the sum of the probabilities of going from state $a$ to state $s$ through an intermediate state $\tilde{s}$ over all states $\tilde{s}$. This is formally stated as:

$$P_{k+1}(s) = \sum_{\tilde{s} \in \mathcal{S}} P_k(\tilde{s}) P(\xi_{k+1} = s | \xi_k = \tilde{s})$$

where $\mathcal{S}$ is the state space.

Now, this equality holds for probability of any given state of a CA, however, just as in the case of the transition matrix, it is usually not feasible to analyze the CA by this method. To analyze a CA some approximation is necessary.[6]

### 2.5.3 Analysis of a probabilistic Game of Life

Mean field theory reduces the problem of a cellular automaton down to a problem of giving the average number of cells on the lattice in a given state as a function of time. In 1978, statistical physicists Schulman and Seiden developed mean field equations for Conway's Game of Life. Part of their analysis will be considered in this section.

An aspect of the paper, not elaborated on in this thesis, is that the authors develop a modified stochastic Game of Life to apply mean field theory to. Although the Game of Life is a strictly deterministic cellular automaton, there is a probabilistic aspect if the initial configuration is unknown. For this reason it may also be of interest to apply mean field thoery to the deterministic Game of Life. It is this second analysis that will be considered here.

Let $\sigma_{ij}(t)$ represent the state of the cell at position $(i, j)$ at time $t$, where 1 represnts living and 0 represents dead. The observation is made that the state of a cell is given by the equation:

$$\sigma_{ij}(t+1) = \delta(3, \sum\nolimits^{'} \sigma(t)) + \sigma_{ij}(t)\delta(2, \sum\nolimits^{'} \sigma(t)) \tag{2.1}$$

[22] where $\delta$ is the Kroenecker delta and $\sum^{'} \sigma(t)$ is the sum of all $\sigma_{i'j'}$ in the eight-cell neighborhood of cell $(i, j)$. Let $\rho(t)$ represent the density of living cells on the grid at time $t$, thus

$\rho(t) = \frac{1}{n}\sum \sigma_(ij)(t)$, if there are $n$ cells in the grid. For the sake of simplicity let $\rho$ represent $\rho(t)$.

It is assumed that the initial configuration is completely random and that the states of the cells are independant. (Of course the state of a cell is dependent on the states of some other cells at the previous time, however this assumption makes calculation easier and will actually give similar results to considering the system with correlations). For any cell $(i, j)$ the probability that $n$ of its 8 neighbors are living follows a binomial distribution, thus the, using equation 2.1, the density of living cells at time $t + 1$ is given by

$$\rho(t+1) = \binom{8}{3}\rho^3(1-\rho)^5 + \rho\binom{8}{2}\rho^2(1-\rho)^6 = 28\rho^3(1-\rho)^5(3-\rho) \tag{2.2}$$

Thus $28\rho^3(1-\rho)^5(3-\rho) - \rho = 0$ may be solved to find the steady states $\rho \approx .19$ and $\rho \approx .37$. Table 2.2 shows the plot of equation 2.2 in the $\rho$, $\rho(t+1)$ plane. It can be seen that $\rho \approx .19$ is a stable fixed point and $\rho \approx .37$ is instable.

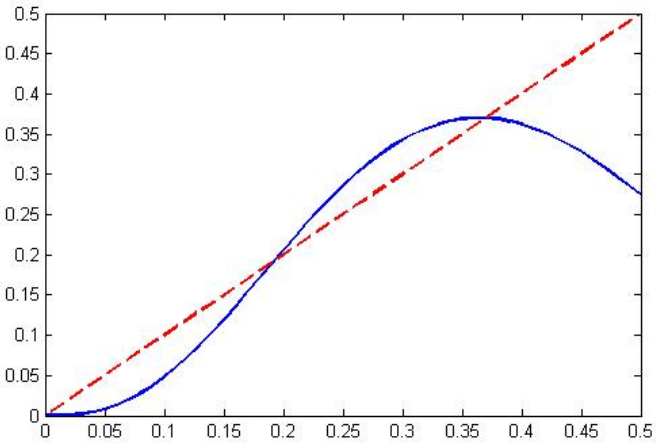


Table 2.2 Plot of density function: $\rho(t+1) = 28\rho^3(1-\rho)(3-\rho)$

The mean field approximation is a good approximation only on a large enough grid. In fact, the approximation gives the exact density only if the grid is infinitely large. The method will usually give qualitatively correct results, despite the fact that it is a very rough approximation.[6]

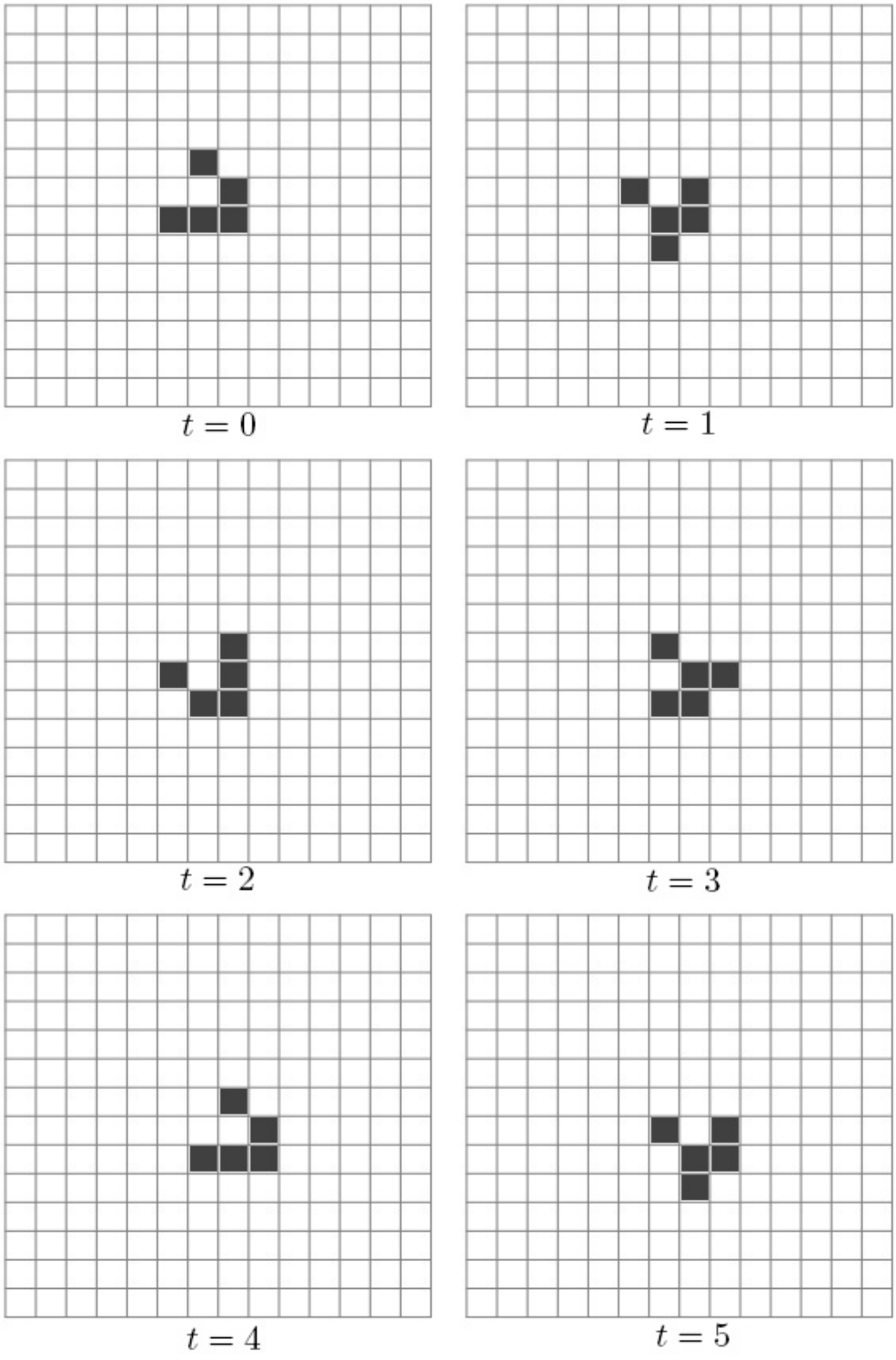


Figure 2.1 Five time-steps of the 'Game of Life'

# CHAPTER 3. Cellular Potts model

The cellular Potts model (CPM) is another model which uses lattice dynamics to study interactions of biological cells. It falls into the category of generalized CA. As with a stochastic CA, probabilistic rules are used to update lattice sites one at a time as time progresses.

## 3.1 History

The cellular Potts model is a generalization of a large-Q Potts model, which is itself a generalization of the Ising model. The Ising model, which was used in statistical physics to study ferromagnetism, was invented in 1920 by Wilhelm Lenz. The model consists of a lattice populated by a discrete number of variables called 'spins', which can take one of two values: 'up' or 'down', $-1$ or 1. Each node of the lattice is populated by a spin, and this spin may interact with its four nearest neighbors residing in the adjacent nodes. An energy function is defined, $E = -\sum J_{ij} S_i S_j (1 - \delta_{i,j})$, where $J_{ij}$ is the exchange constant and $S_i$ is the spin of cell $i$. As time is progressed spins will 'flip' in order to minimize this energy function. The cellular Potts model was introduced in a 1992 paper by Glazier and Graner, in which they used an extended Potts model to simulate the sorting of biological cells in two-dimensions.[23] Instead of two spins, there are $N$ spins used in the CPM to represent $N$ different biological cells. Also, each spin is assigned one of a possible certain number of *types*, this will be discussed in more detail in this chapter.

## 3.2 Biological Considerations

Certain cell types display the ability, when randomly mixed, to rearrange themselves again into homogeneous cell clusters, or possibly even coherent tissue. The differential adhesion

hypothesis, proposed by biologist Malcolm Steinberg in 1964 "explains cell sorting and related cell rearrangements as progressions of motile and mutually adhesive cell populations toward configurations of minimal interfacial (adhesive) free energy."[24]. According to this hypothesis, cell sorting is a matter of minimizing free energy and thus a model of cell sorting ought to have an energy function which is minimized as cells properly sort themselves. The cellular Potts model was originally designed to model this particular phenomenon.

Many lattice models use a single grid cell to represent a single biological cell. One interpretation of this characteristic of the model is that the model assumes all biological cells have identical size and shape, in another sense it may be interpreted that the model assumes cell size and shape are trivial and in this sense each biological cell is represented by a point. These assumptions are dangerous to make, because the shape and size of a biological cell are nontrivial characteristics of a cell, and the fact that the cellular Potts model does not make these same assumptions is a strength of the model.

### 3.3 The Glazier-Graner Extended Potts Model

As with a CA, the CPM was set in a two-dimensional lattice. Biological cells, as previously mentioned, are represented by one or more grid cells adjacent to each other with the same spin. (Spin is a term borrowed from the statistical physics, in this context spin is a unique identifier for a biological cell). Like a stochastic CA there is a simple rule for how the state of grid cells should change, depending on the state of nearby grid cells. In a model of a population of $N$ biological cells each grid cell is assigned one of the $N$ spins, $\sigma(i,j) = 1, 2, 3, \ldots, N$, where a grid cell is identified by $(i,j)$. All grid cells of one particular spin represent a single biological cell. As this is a model of cell sorting, there must be a concept of cell type (and this is where the model departs from previous Potts models). In the paper that introduced the CPM, Graner and Glazier considered three cell types, $\tau(\sigma) = l$, $d$, or $M$, where $l$ stands for light and represents cells with low-adhesivity, where $d$ stands for dark and represents cells with high-adhesively, and $M$ stand for medium and represents the extra-cellular matrix, i.e. the medium the cells exist in.

At each time step an attempt to change the spin of one of the lattice sites is made. The

probability that this change will occur depends on whether the new configuration of cells has higher, lower, or equal surface energy as compared to the previous configuration of cells. The energy of the system is defined by the following Hamiltonian: $\mathcal{H}_{sort} =$

$$\sum_{(i,j),(i',j') \text{ neighbors}} J(\tau(\sigma(i,j)),\tau(\sigma(i',j'))(1-\delta_{(i,j),(i',j')})+\lambda \sum_{\text{spin types } \sigma} [a(\sigma)-A\tau(\sigma)]^2(1-\delta_{\tau(\sigma),M}) \tag{3.1}$$

where $J(\tau;\tau')$ is the surface energy between spins of type $\tau$ and $\tau'$, $\lambda$ is a constant that specifies the strength of the constraint on cell area, $a(\sigma)$ is the area of the cell with spin $\sigma$ and $A_\tau$ is the target area of a cell of type $\tau$. The probability that a grid cell changes its spin from $\sigma$ to $\sigma'$ (let's call these two system states $A$ and $B$, respectively) is given by

$$P(\sigma(i,j)\to\sigma(i',j')) = \begin{cases} \{\exp(-\Delta\mathcal{H}/kT) : \Delta\mathcal{H}\geq 0; \quad 1 : \Delta\mathcal{H} < 0\} & \text{for } T > 0 \\ \{0 : \Delta\mathcal{H} > 0; \quad 0.5 : \Delta\mathcal{H} = 0; \quad 1 : \Delta\mathcal{H} < 0\} & \text{for } T = 0 \end{cases} \tag{3.2}$$

where $\Delta\mathcal{H} = \mathcal{H}(A) - \mathcal{H}(B)$ and $T$ represents temperature in a sense similar to simulated annealing and $k$ is the Boltzman constant.

Now, to determine system energy, a surface energy function, which reflects the adhesivity of the cells, must be defined. A cell with high-adhesivity should have a low surface energy between itself and another cell of that same type, likewise, a cell with low-adhesivity should have a high surface energy between itself and another cell of the same type. According to the authors of the paper the following relationship between the various surface energies ought to exist: $0 < J(d,d) < [J(d,d) + J(l,l)]/2 < J(d,l) = J(l,d) < J(l,l) < J(l,M) = J(d,M)$. It had been established by the authors, and others, that this relationship ought to exist for spontaneous cell sorting to occur.[23]

Updates to the system are made according to the Metropolis algortithm, described in more detail in section 3.4. The system updates as follows: At each step a lattice cell is chosen at random, then it is randomly determined whether the cell ought to flip its spin to that of one of its neighbor cells, according to the probabilities, dependenant on $\mathcal{H}$, given above.

It should be expected that the system will move towards a state where high-adhesivity cells are near the maximum number of cells of their same type, likewise low-adhesivity cells will attempt to be near as few cells of their same type as possible. This configuration is accomplished

in this manner: cells with high-adhesivity congregate in the middle of the cell mass while cells with low-adhesivity should form a thinner boundary between the high-adhesivity cells and the extra-cellular matrix. These same sorting behaviors are observed in vitro.[25] Although the constants chosen by the authors do not have a strong correlation to experimentally determined phenomena, they were in fact able to simulate cell sorting similar to what is observed in vitro. A similar, simplified model was written to accompany this thesis. That model will be discussed in greater detail in Chapter 5. Refer to Figure 3.1 for an example of cell sorting generated by this much simplified model, which is similar to cell sorting produced by the CPM.

## 3.4 Metropolis Algorithm/Simulated Annealing

The Boltzman distribution is a probability measure of the states in a system, which is found in physics and chemistry and other sciences. The Metropolis algorithm is a strategy for obtaining a random sample from a Boltzman distribution, which has also been generalized to obtain a random sample from any distribution which is proportional to the Boltzman distribution. This is relavent because the Hamiltonian of both the CPM and Ising model are both proportional to the Boltzman distribution, thus the Metropolis algorithm is useful for finding the minimum energy states for these two models.

The Metroplis algorithm as applied to the Ising model is as follows:

1. Begin in some system state $A$ with known energy $\mathcal{H}(A)$. Randomly choose a spin to flip to obtain a new state $B$

2. Evaluate $\mathcal{H}(B)$, the energy of state $B$.

3. If $\mathcal{H}(B) < \mathcal{H}(A)$, accept state $B$ with probability 1.

4. Otherwise keep the higher energy state, $B$, with probability $\exp\left(-(\mathcal{H}(B) - \mathcal{H}(A))/T\right)$ This means that for high temperature it will be easier for the system to go to higher energy states.[26]

The parameter $T$, in the previously stated probabilities, represents temperature, which is a parameter used in an optimization technique known as simulated annealing. In metallurgy,

annealing is a process of heating and cooling which is used to optimize the crystalline structure of metals to increase their strength and hardness. Simulated annealing will aid in the search for a state near the global minimum.

The algorithmic process begins with $T$ relatively high. If $T$ is too low, the system may get stuck in a local minimum; all system states near the local minimum would have higher system energy, and if $T$ is near 0 it is unlikely that the system will be allowed to leave that state. It is necessary, however, that the temperature not be too high, otherwise even if the system finds a minimum, it is unlikely that it will remain near that minimum. Some appropriately high temperature between these extremes must be chosen.

After the system progresses through a certain number of time steps at this initial temperature it is time that it begin a cooling schedule, where temperature is decreased with successive time steps. However, no cooling schedule was enacted in the CPM simulation by Graner and Glazier. Rather, they determined a sufficiently high starting point (but not too high, otherwise the cells would begin to disassociate), and then ran the simulation without decreasing the temperature for a predetermined number of time steps. To complete the process the simulation was ran for an additional two time steps with $T = 0$. This last step was done with the intention of getting the average cell size to reduce to a predetermined ideal cell size (in the case of this simulation this desired size is six grid cells per biological cell).

CPM have a nice relationship with the differential adhesion hypothesis. Also, they do not make the oversimplification of assuming that cells are shapeless points. There is continuing research with this model by Glazier and other researchers. Much of the new research combines this model with other successful modeling techniques to form a 'hyrbid' model.

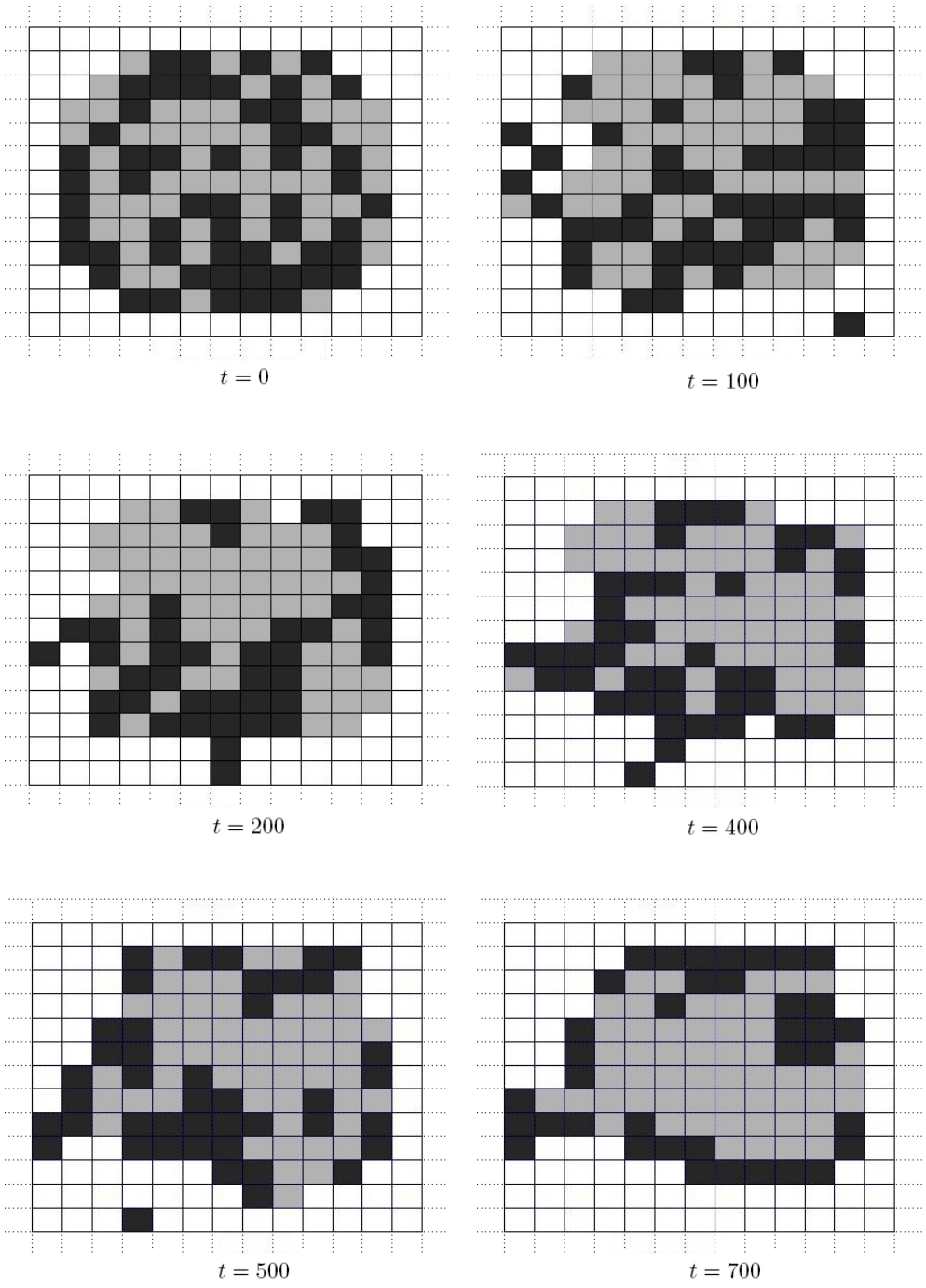


Figure 3.1 Time progression of a CPM-like CA simulation

# CHAPTER 4. Agent Based Models

The generalized model type that both cellular automata and cellular Potts models fall into is that of agent based models (ABM). Agent based models could also be called individual based, because that's what the variables are in some sense – individuals. These individuals, or agents, and a set of prescribed rules which governs the behavior of these individuals are the components that make up an ABM. In the case of CA, these agents are the individual lattice cells. In fact, von Neumann's CA are considered by some to be the first ABM. ABM have been used to model a wide variety of problems in fields including social sciences, economics, transportation management, and now cellular biology. [27]

This section will focus on some recent applications of agent based modelling in biology. In particular it will focus on some ABM that may be called hybrid CA. Hybrid CA are cellular automata which have been seemlessly combined with some other model type, usually to account for some sort of continuous variable that a strictly discrete CA could not realistically account for.

## 4.1 Introduction

Multiscale modeling is an important area of research within mathematical biology. The pertinant question is this: how do we develop mathematical models which are accurate at multiple scales? What if, for example, we need a model that not only accurately describes both cellular level interactions and tissue level behaviors but also models some phenomena on a molecular level? The biological processes which are being modeled are complex and "neither a true continuum nor a mere discrete model can describe all processes sufficiently". [27] Agent based models, and specifically hybrid CA, show promise in being able to better describe these

complex biological systems.

As has been argued previously in this thesis, CA are good models of cellular interaction and tissue pattern formation. They are not, however, good models of continuous biological properties, like chemoattractant and nutrient concentrations or any other fluid aspects of cell growth.[27] Rather than unrealistically considering these continuous elements as discrete variables, a continuous element can be introduced.

## 4.2 A hybrid model of tumor growth

In Chapter 2 evidence was given showing that using a CA to model early tumor development is a effective strategy. However there are a number of attributes of a tumor's environment that are not well described in discrete terms, such as nutrient levels, toxin levels and $H^+$ ion levels of the extracellular environment.

It is clear that tumor cells prefer to move towards regions with high nutrient levels, away from regions with higher toxin levels, and that cells will die in the absense of nutrients or in the presense of high toxicity. Researchers at the Complex Biosystems Modeling Labratory, Harvard-MIT have presented an ABM to explore these problems [27][19] The pH of the extracellular environment also plays an important role in tumor invasion. Low pH of the extracellular environment aids in tumor invasion in three ways: 1.) it is easier for the mutated tumor cell, in comparison to the healthy normal cell, to live in this environment, thus it aids them to outcompete the normal cells, 2.) low pH triggers the productions of enzymes that help in the breakdown of the extracellular matrix and 3.) low pH is also detrimental to the intercellular adhesion and communincation of normal cells, aiding the tumor cells in their invasion between normal cells.[28]

The presence of tumorous cells actually exacerbates this problem, as they produce lactic acid, since cancer cells seem to rely on anaerobic rather than aerobic respiration.

In 2001 Patel, Gatenby et al. published a paper describing an agent based model with the purpose of studying this problem. Defined on a large two-dimensional grid, the state of each cell of that grid is a vector of four components: 1.) one of four discrete grid cell states: tumorous cell, normal cell, microvessel, or vacant 2.) the continuous $H^+$ ion concentration, of

that grid cell (represented by $H^+$, for the duration of this chapter), 3.) the continuous glucose concentration, of that grid cell, and 4.) for microvessels only: a set of eight ghost values, four (for the four walls of the vessel) for each of the two chemical concentrations, to enforce gradient boundary conditions. The rules that govern the evolution of the states of the cells are as follows:

1. If the grid cell represents a tumor cell and $-\log(H^+) < pH_T^d = 6.0$ then the cell dies.
2. If the grid cell represents a normal cell and $-\log(H^+) < pH_N^d = 6.8$ then the cell dies.
3. If the grid cell represents a tumor cell and $pH_T^d < -\log(H^+) < pH_T^Q = 6.4$ then the cell is quiescent (the cell does nothing this time step).
4. If the grid cell represents a normal cell and $pH_N^d < -\log(H^+) < pH_N^Q = 7.1$ then the cell is quiescent.
5. If the grid cell represents a tumor cell and $-\log(H^+) > pH_T^Q$ and the cell has at least one vacant neighbor, mitosis occurs and a tumorous daughter cell is placed in the neighboring vacant cell with the highest glucose concentration.
6. If the grid cell represents a normal cell and $-\log(H^+) > pH_N^Q$ and the cell has at least one vacant neighbor, mitosis occurs and a normal daughter cell is placed in the neighboring vacant cell with the highest glucose concentration.

These states are updated in a series of sub-generations. The authors found that having all cells within a characteristic diffusion length all be updated at once caused spurious correlations, so they randomly selected a tenth of the total grid to update per sub-generation. After these first elements of the cell state vectors have been updated for a sub-generation, the other elements of the state vectors must be updated also to reflect these changes.

Considered as a time dependent parabolic diffusion equation the equation for the glucose conentration at position $\mathbf{r}$ is as follows: $D_G \nabla^2 G(\mathbf{r},t) - k(\mathbf{r})G(\mathbf{r},t) = \frac{\partial G(\mathbf{r},t)}{\partial t}$, where $D_G$ is the glucose diffusion constant. The authors argue, however, that because of the large difference in time scale between cell proliferation ($\mathcal{O}(10^2 hrs)$) and chemical diffusion ($\mathcal{O}(1s)$) it is possible to

coarsen the time scale and avoid having to solve this more computationally expensive equation. It suffices to solve the elliptic boundary-value problem: $D_G \nabla^2 G_t(\mathbf{r}) - k(\mathbf{r}) G_t(\mathbf{r}) = 0$, where $G_t(\mathbf{r})$ is the glucose concentration at cell $\mathbf{r}$ after sub-generation $t$ and $k(\mathbf{r})$ represents the glucose consumption rate, also at that cell. These rates are dependent on the state of the grid cell:

$$k(\mathbf{r}) = \begin{cases} k_N & \text{where the cell at location } \mathbf{r} \text{ is a normal cell} \\ k_T & \text{where the cell at location } \mathbf{r} \text{ is a tumor cell} \\ 0 & \text{where the cell at location } \mathbf{r} \text{ is a microvessel} \\ 0 & \text{where the cell at location } \mathbf{r} \text{ is vacant} \end{cases}$$

Glucose is able to enter the system by diffusion through the microvessel walls. This is modeled through the boundary conditions $-D_G \hat{n} \cdot \nabla^2 G_t|_{wall} = q_G(G_S - G_t|_{wall})$, where $q_G$ is the vessel permeability level, $\hat{n}$ is the unit normal vector pointing orthogonal to the cell wall and $G_S$ is the serum glucose level.

Similar equations govern the acid concentration also. The boundary-value equation governing the $H^+$ ion concentration is: $D_H \nabla^2 H_t(\mathbf{r}) - h(\mathbf{r}) = 0$, where $D_H$ is the diffusion constant for lactic acid and $h(\mathbf{r})$ is the acid production according to the following:

$$h(\mathbf{r}) = \begin{cases} \dot{H}_T^A & \text{where the cell at location } \mathbf{r} \text{ is an active tumor cell} \\ \dot{H}_T^Q & \text{where the cell at location } \mathbf{r} \text{ is a quiescent tumor cell} \\ 0 & \text{where the cell at location } \mathbf{r} \text{ is not a tumor cell} \end{cases}$$

Microvessels are capable of removing acid from the system and this is represented by the following boundary conditions: $-D_H \hat{n} \cdot \nabla^2 H_t|_{wall} = q_H(H_S - H_t|_{wall})$, where $q_H$ is the vessel permeability to lactic acid and $H_S$ is the serum lacic acid concentration.

The boundary-value problem for the glucose and lactic acid concentrations were then solved using the method of successive over-relaxation with Chebyshev acceleration, which is a standard iterative linear system solution method. It is interesting to note that the simulation took over 500 h of computing time for a grid of size 100 x 100. The simulation was found to agree with previously established biological theory.

## 4.3 Other recent hybrid models

New research continues to be done in this area of hybrid models. There are more than a few researchers who are continuing to develop these models to study biological phenomena. In a 2004 paper, mathematician Alexander Anderson describes another hybrid model of tumor growth, in this instance focusing on how tumor cells interact with and invade host tissue, and the importance of matrix-degrading enzymes and oxygen.[14] A 2009 paper by physicist Nikodem Poplawski describes a collaberation between himself and others, including Glazier and Anderson.[31] The hybrid model borrows characterstics from both the CPM and Anderson's hybrid model to simulate the morphology of avascular tumors.

Mathematician Amy Bauer et al. presented another hybrid CPM of tumor growth in a 2007 paper.[29] Expanded in 2009[30], this model focuses on tumor-induced angiogenesis.

The formation of new blood vessels from existing vasculature is called angiogensis. Tumors have the ability to induce the formation of these new blood vessels, and thus become vascularized. Since a vascularized tumor, having its growth sustained by a blood supply, poses a much greater danger to a patient and is much more difficult to treat than a non-vascularized tumor, this is certainly an item of interest to be studied. This model simulates the migration of the endothelial cells (the type of cells which form the inner lining of blood vessels) into the extracellular matrix, as a blood vessel grows towards a tumor mass.

As in the Glazier and Graner model, the spin of a grid cell, $\sigma$, is used to designate which biological cell that grid cell is part of. Each cell also has a type, $\tau$, asscociated with it. In this case the type may be one of the the four values $e$, $m$, $t$, or $f$, which stand for for endothelial cell , extracellular matrix fiber, tissue cell, and interstitial fluid (also called 'tissue fluid', which is the fluid that surrounds the cells of multicellular animals), respectively. A single spin, 1, and a single spin, 0, are assigned to all grid cells which make of the extracellular matrix and the interstitial fluid, respectively.

The grid updates according to the Metropolis algorithm and updates are accepted with the probability

$$P(\sigma(i,j) \to \sigma(i',j')) = \begin{cases} 1 & \Delta\mathcal{H} < 0 \\ \exp\left(\Delta\mathcal{H}/kT\right) & \Delta\mathcal{H} \geq 0 \end{cases}$$

and the energy function is defined as

$$\mathcal{H} = \sum_{(i,j),(i',j') \text{ neighbors}} J(\tau(\sigma(i,j)), \tau(\sigma(i',j'))(1 - \delta_{(i,j),(i',j')}) + \lambda \sum_{\text{spin types } \sigma} \gamma(\tau)[a(\sigma) - A\tau(\sigma)]^2 + \sum_{(i,j),(i',j') \text{ neighbors}} \chi(\sigma)\Delta V + \alpha \sum_{\text{spin types } \sigma} (1 - \delta_{a(\sigma),a(\sigma')}) \tag{4.1}$$

The first two terms of equation 4.1 are similar to the first two terms of the energy function for the CPM, given in equation 3.1. The third term accounts for chemotaxis; it represents the effect the chemical attractants, which are being being produced by the tumor, have on the newly forming endothelial cells; $V$ is the concentration of vascular endothelial growth factor (VEGF). The fourth term is a continuity constraint which prevents individual endothelial cells from splitting up (an unrealistic possibility in this model without this constraint).

The VEGF concentration, $V = V(x, y, t)$ follows the differential equation

$$\frac{\partial V}{\partial t} = D\nabla^2 V - \lambda V - B(x, y, V) \tag{4.2}$$

where $B$ represents the amount of VEGF which an endothelial cell can bind with and internalize,

$$B(x, y, V) = \begin{cases} \beta & \text{if } \beta \leq V \text{ and } \sigma(x,y) = e \\ V & \text{if } 0 \leq V < \beta \text{ and } \sigma(x,y) = e \\ 0 & \text{if } \sigma(x,y) \neq e \end{cases}$$

Every time step, before the grid is updated, equation 4.2 is solved for $V$. Just as the model outlined in section 4.2 considers chemical concentrations as continuous variables, the VEGF concentration is considered as a continuous component in this model.

In both the 2007 and 2009 papers the authors state that this model is a realistic simulation of new blood vessel sprouting, having compared it to experimental data. However, there is large number of interrelated underlying biological components which are responsible for the process of angiogenesis. The authors mention that there is still much research to be done to study

some of these other parameters. With their flexibility to accurately model both continuous and discrete biological systems, these hybrid model approaches are a promising strategy for simulating biological phenomena.

# CHAPTER 5. Sample Simulations

In this section, a sample computer simulation, which was written as part of the study for this thesis, will be discussed. See Appendix A for the relevant Matlab code.

## 5.1 A cell sorting CA

A simple cell sorting CA, similar to Graner and Glaziers CPM, was written in Matlab to accompany this thesis. The model is defined on a large grid. Each grid cell has 3 possible states: cell with high-adhesivity, cell with low adhesivity, extracellular matrix. At the beginning of the simulation a large circle in the center of the grid is randomly populated with both low and high adhesive cells, the rest of the grid is extracellular matrix.

Each time step consists of the random selection of a certain number of cells. These randomly selected cells then will attempt to exchange their state with one of the cells in their 8-cell neighborhood. Before and after the exchange is made the engergy of the system is calculated according to 3.1 and the exchange will be made according to the probability defined by 3.2. As a downfall of its simplicity, the model does not represent the uniqueness of the size and shape of cells.

Despite its simplicity, the simulation displays cell sorting similar to the CPM, as displayed by Figure 3.1. The simulation also accurately reflects the role that temperature ought to play in the model, according the CPM. If temperature is set to a moderately high level cells will sort such that low-adhesivity cells form a thin border between the large mass of high-adhesivity cells in the center and the extracellualr on the outside. If the temperature is set too low, then the low-adhesivity cells will not all have moved to the ousdide of the cell mass. Some of the low-adhesivity cells will form thin border bands on the inside of the cell mass,

separating high-adhesivity cells from other high-adhesivity cells. In this scenario the system has found a local minimum energy configuration. If, however, the temperature is set too high the system will allow too many energetically disfavorable moves. In the CPM cells will begin to disassociate from themselves. In the CPM and this model, the cell mass begins to disassociate and extracellular matrix creeps into the cell. Neither a local minium nor a state near the global minimum energy configuration will be found.

The simulation can be used to explore other aspects of the annealing process of such a model. Glazier and Graner's CPM does not employ a traditional cooling schedule, in which the temperature is decreased towards at each successive time step. Rather they run the model for a few $T = 0$ annealing steps at the end. In the simple cell sorting CA it seems that employing a cooling schedule is a better strategy (the system seems to be more likely to move towards the expected final sorted state). However, these observations are purely qualatative and are subjective. Also, there may be some element of the CPM not represented in the simple model that makes the employment of a cooling schedule unnecessary.

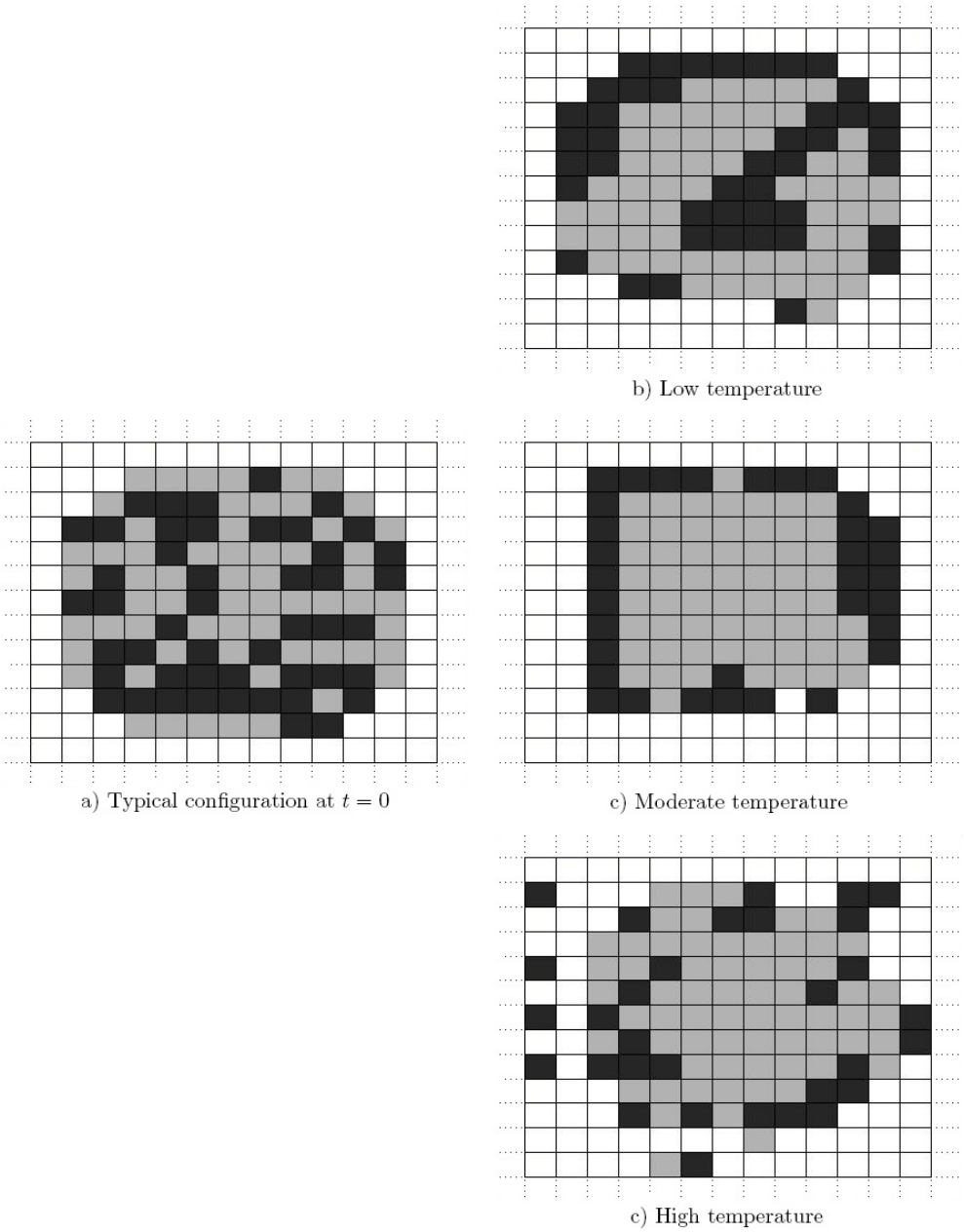


Figure 5.1 Simulation configurations resulting from low, medium and high temperatures.

# CHAPTER 6. End Remarks

There are a variety of reasons why cellular automata and cellular automata-like models attract researchers who want to model biological phenomena. They are simple models with complex behaviors that lend themselves well to modeling the behaviors of some sort of population.

It should come as no surprise that a model that was invented to simulate a population of artificial self-reproducing individuals can also model populations of living self-reproducing individuals, such as cells. Cells are the bulding blocks of life, maybe then they also be the building blocks of life models. There are many reasons that a CA may be chosen to perform this task.

There is much ongoing research in this field. These models continue to be improved, especially by means of combining successful models together in a hybrid model. There is much research to be done in developing useful analsyis for these models.

# APPENDIX A. Additional Material

This chapter includes Matlab code written to accompany this thesis.

## A simple cell-sorting cellular automaton

```
%Cell Sorting Model
function SimpleCellSortCA
  m = 15; n = 15; %grid dimensions
  t_b = 12; %number of time steps
  T_0 = 14; %Temperature (like simulated annealing)
  T = T_0;
  [Grid, Spin] = initialize_grid(m, n); %The idea of 'Spin' is not
                                        %currently used, perhaps it will be
                                        %implemented in a later version
  paint_cell(Grid)
  pause(1);

  for k1 = 1:t_b
    for k2 = 1:16*50 %many cells will attempt to trade places with
                     %each other during a single time step.
      %choose a random location in the grid
      i = max(ceil((m-1)*rand),2); j = max(ceil((n-1)*rand),2);
      [Grid, Spin] = cells_move(Grid, Spin, T, i, j);
    end
```

```
    paint_cell(Grid)
    pause(1);
    T = T - T_0/t_b;%unlike CPM we will undergo a cooling schedule with T
                    %being reduced towards zero.
  end

  %we'll sort for a while now with T = 0
  T = 0;
  for k2 = 1:16*50
    %choose a random location in the grid
    i = max(ceil((m-1)*rand),2); j = max(ceil((n-1)*rand),2);
    [Grid, Spin] = cells_move(Grid, Spin, T, i, j);
  end

  paint_cell(Grid)
  pause(1);
end

% initializes an m x n grid and corresponding figure
% for the cells to 'live' in
function [Grid, Spin] = initialize_grid(m, n)
  Grid = zeros(m,n); Spin = zeros(m,n); %We are not currently using 'spin'
  center_row = int8(m/2); center_col = int8(n/2);

  % 0 - extracellular matrix, 1 - low density cell, 2 - high density cell
  for i = 1:m
      for j = 1:n
          if (abs(center_col - i))^2 + (abs(center_row - j))^2 <
```

```
                                                    ((center_row+center_col)/2 - 2)^2
                p = rand;
                if p > 0.45
                    Grid(i,j) = 1;
                else
                    Grid(i,j) = 2;
                end
            end
        end
    end

    close all
    x = 0:m; y = 0:n;
    figure
    hold on
    for i = 1:m+1
      plot(x(i)*ones(1,n+1),y,':')
    end
    for j = 1:n+1
      plot(x,y(j)*ones(1,m+1),':')
    end
end

%the Hamiltonian of the system
% H_sort = sum over all neighbors (surface energy between cell types) Note:
% Glazier's model is more complicated, a cell might likely take up more
% than one grid cell location
function H_sort = hamiltonian(Grid, Spin)
```

```
  [m, n] = size(Grid);
  H_sort = 0;
  %J is surface energy between 2 cell types
  for i = 2: m-1
      for j = 2: n-1
          %check east, north and northeast neighbor of a cell location
          %every grid cell is considered a different biological cell
          H_sort = J(Grid(i,j), Grid(i, j+1)) + J(Grid(i,j), Grid(i+1, j)) +
                   J(Grid(i,j), Grid(i+1, j+1)) + H_sort;
      end
  end
end

%A cell will attempt to trade places with a random neighbors
function [Grid, Spin] = cells_move(Grid, Spin, T, i, j)
  H_sort = hamiltonian(Grid, Spin);
  Gtemp = Grid; Stemp = Spin;

  %the cell randomly chooses a neighbor to attampt to trade places with
  p = rand;
  if p < .125
    Gtemp(i, j) = Grid(i + 1, j); Gtemp(i + 1, j) = Grid(i, j);
  elseif p < .25
    Gtemp(i, j) = Grid(i, j + 1); Gtemp(i, j + 1) = Grid(i, j);
  elseif p < .375
    Gtemp(i, j) = Grid(i - 1, j); Gtemp(i - 1, j) = Grid(i, j);
  elseif p < .5
    Gtemp(i, j) = Grid(i, j - 1); Gtemp(i, j - 1) = Grid(i, j);
```

```
elseif p < .625
  Gtemp(i, j) = Grid(i + 1, j + 1); Gtemp(i + 1, j + 1) = Grid(i, j);
elseif p < .75
  Gtemp(i, j) = Grid(i - 1, j + 1); Gtemp(i - 1, j + 1) = Grid(i, j);
elseif p < .825
  Gtemp(i, j) = Grid(i - 1, j - 1); Gtemp(i - 1, j - 1) = Grid(i, j);
else
  Gtemp(i, j) = Grid(i - 1, j - 1); Gtemp(i - 1, j - 1) = Grid(i, j);
end
H_new = hamiltonian(Gtemp, Spin);

if T == 0
  if H_new < H_sort
      Grid = Gtemp;
  elseif H_new == H_sort
      p = rand;
      if p > .5
          Grid = Gtemp;
      end
  else
      %no move!
  end
elseif T > 0 %T > 0 allows for moves that will increase energy
  if H_new > H_sort
      p = rand;
      if p < exp((H_sort-H_new)/(T))
          Grid = Gtemp;
      end
```

```
    else
        Grid = Gtemp;
    end
  end
end

%J is surface energy between 2 cell types
%0 < J(2,2) < [J(2,2) + J(1,1)]/2 < J(2,1) < J(1,1) < J(1,0) = J(2,0)
function J = J(t1, t2)
  J = 0;
  if (t1 == 0) && (t1 ~= t2)
      J = 16;
  elseif t1 == 1
      if t2 == 1
          J = 14;
      elseif t2 == 2
          J = 11;
      elseif t2 == 0
          J = 16;
      end
  elseif t1 == 2
      if t2 == 1
          J = 11;
      elseif t2 == 2
          J = 2;
      elseif t2 == 0
          J = 16;
      end
```

```
  end
end

%update current grid location in the figure
function paint_cell(Grid)
  [m, n] = size(Grid);
  x = 0:m; y = 0:n;
  for i = 2:m-1
    for j = 2:n-1
      color = 'w'; %extracelluar matrix is white
      if Grid(i,j) == 1
        color = [.15,.15,.15]; %low density cells are blackish
      elseif Grid(i,j) == 2
        color = [.7,.7,.7]; %high density cells are gray?
      end
      a = x(i); b = x(i+1); % indexing offset by +1 in x,y
      c = y(j); d = y(j+1);
      u = [a,b,b,a,a];
      v = [c,c,d,d,c];
      fill(u,v,color)
    end
  end
end
```

# BIBLIOGRAPHY

[1] Meyer, Walter. (1984). *Concepts of Mathematical Modeling*, McGraw-Hill: New York.

[2] Shier D.R., K.T. Wallenius. (2000). *Applied Mathematical Modeling*, Chapman & Hall/CRC: Boca Raton.

[3] de Vries, Gerda, Thomas Hillen, Mark Lewis, Johannes Müller, Birgett Schönfisch. (2006). *A Course in Mathematical Biology*, SIAM: Philedelphia.

[4] Merks, Roeland, James Glazier. (2005). A cell-centered approach to developmental biology, *Physica A*, *352*, 113–130.

[5] Wilkinson, Darren. (2006). *Stochastic Modeling for System Biology*, Chapman & Hall/CRC: Boca Raton.

[6] Deutsch, Andreas, Sabin Dormann. (2005). *Cellular Automaton Modeling of Biologcial Pattern Formation*, Birkhäuser: Boston.

[7] Edelstein-Keshet, Leah. (2005). *Mathematical Models in Biology*, SIAM: Philedelphia.

[8] Allman, Elizabeth, John Rhodes. (2003). *Mathematical Models in Biology*, Cambridge University Press: Cambridge.

[9] Agheksanterian, Alen, Matthias Gobbert. (2007). Modeling the spread of epidemic cholera: an age-structured model, Technical Report number TR2007-9, University of Maryland, Baltimore County.

[10] Swat, Maciej, J. Scott Gens, Benjamin L. Zaitlen, Nikodem J. Poplawski, Maciej Swat, James A. Glazier. (2009). 3D multi-cell simulation of tumor growth and angiogenesis, *PLoS One 4*(10), Published online: e7190.

[11] Maedal, Takuya, I. Ajioka, K. Nakajima. (2007). Computational cell model based on autonomous cell movement regulated by cell-cell signalling successfully recapitulates the "inside and outside" pattern of cell sorting, *BMC Syst Biol 1*(43), Published onilne: 10.1186/1752-0509-1-43.

[12] Bru, Antonio, Pere-Joan Cardona. (2010). Mathematical modeling of tuberculosis bacillary counts and cellular populations in the organs of infected mice, *PLoS One 5*(9), Published online: e12985.

[13] Moreira, J., Andreas Deutsch. (2005). Pigment pattern formation in zebrafish during late larval stages: a model based on local interactions, *Developmental Dynamics 232*, 33-42.

[14] Anderson, Alexander. (2005). A hybrid mathematical model of solid tumor invasion: the importance of cell adhesion, *Mathematical Medicine and Biology 22*(2), 163-186.

[15] Gardner, Martin. (1970). Mathematical Games: The fantastic combinations of John Conway's new solitaire game "life", *Scientific American 223*, 120-123.

[16] Silva, H.S., M.L. Martins. (2003). A cellular automata model for cell differentiation, *Physica A 322*, 555-566.

[17] Qi, An-shen, X. Zheng, C.Y. Du, B.S. An. (1993). A cellular automaton model of cancerous growth *Journal of Theoretical Biology*, *161*, 1–12.

[18] Lo, C.F. (2007). Stochastic Gompertz model of tumour cell growth, *Journal of Theoretical Biology 248*(2), 317-321.

[19] Mansury, Yuri, Mark Kimura, Jose Lobo, Thomas Deisboeck. (2002). Emerging patterns in tumor systems, *Journal of Theoretical Biology 219*(3), 343-370.

[20] Wolfram, Stephen. (2002). *A New Kind of Science* Wolfram Media, Inc: Champaign, IL.

[21] Culik, Karel, Shen Yu. (1998). Undecidability of CA classification schemes, *Complex Systems 2*, 177-190.

[22] Schulman, L.S., P.E. Seiden. (1978). Statistical mechanics of a dynamical system based on Conway's game of life, *Journal of Statistical Physics 19*(3), 293–314.

[23] Graner, François, James Glazier. (1992). Simulation of biological cell sorting using a two-dimensional extended Potts model, *Physical Review Letters 69*, 2013-2016.

[24] Steinberg, M.S. (1975). Adhesion-guided multicellular assembly: a commentary upon the postulates, real and imagined, of the differential adhesion hypothesis, with special attention to computer simulations of cell sorting, *Journal of Theoretical Biology 55*(2). 431-443.

[25] Foty, R.A., C. Pfleger, G. Forgacs, M.S. Steinberg. (1996). Surface tensions of embryonic tissues predict their mutual envelopment behavior, *Development 122*, 1611-1620.

[26] Saeta, Peter. (2005). The Metropolis Algorithm: Statistical Systems and Simulated Annealing, Harvey Mudd College http://saeta.physics.hmc.edu/courses/p170/Metropolis.pdf.

[27] Zhang, Le, Z. Wang, J.A. Sagotsky, T.S. Deisboeck. (2009). Multiscale agent-based cancer modeling, *Mathematical Biology 58*, 545-559.

[28] Patel A., E. Gawlinki, S. Lemieux, R. Gatenby. (2001). Cellular automaton model of early tumor growth and invasion: the effects of native tissue vascularity and increased anaerobic tumor metabolism, *Journal of Theoretical Biology 213*(3), 315-331.

[29] Bauer, Amy, Trachette Jackson, Yi Jiang. (2007). A cell-based model exhibiting branching and anastomosis during tumor-induced angiogenesis, *Biophys J. 92*(9), 3105-3121.

[30] Bauer, Amy, Trachette Jackson, Yi Jiang. (2009). Topography of extracellular matrix mediates vascular morphogenesis and migration speeds in angiogenesis, *PLoS Computational Biology 5*(7), Published online p.e1000445.

[31] Poplawski, Nikodem, U. Agero, J. Gens, M. Swat, J. Glazier, A. Anderson. (2009). Front instabilities and invasiveness of simulated avascular tumors, *Society for Mathematical Biology 71*, 1189-1227