University of Szeged

Doctoral Thesis

Mathematical models of cell cultures

Author:
Péter **Boldog**

Supervisor:
Gergely **Röst**

Doctoral School of Mathematics and Computer Science,
Bolyai Institute, University of Szeged

2022

Contents

Köszönetnyilvánítás

Köszönöm Témavezetőmnek, dr. Röst Gergelynek, a támogatását és azt a számtalan dolgot, amit tanított nekem - ha akarta, ha nem. Nélküle ma egészen biztosan más ember lennék. Minden beszélgetés inspiráló volt, akkor is, ha sosem volt elég. Biztos vagyok abban, hogy ha valaha lesz szerencsém szakmai sikerekben, akkor annak nagy részét a mellette töltött éveknek köszönhetem és annak, amit tőle tanultam.

Köszönettel tartozom korábbi témavezetőimnek, sorrendben dr. Hideg Évának, dr. Udvardy Andornak (akivel sajnos csak rövid ideig volt alkalmam együtt dolgozni), dr. Pál Margitnak hogy bevontak a kutatásaikba. Külön kiemelem dr. Nagy Lászlót, a hajdani Biofizika Intézetből, nem emlékszem, kifejeztem-e, hogy mekkora szerencsém volt, hogy vele dolgozhattam életem első cikkén.

Dr. Szabó G. Gábornak hasonlóan nagy köszönettel tartozom. Ő az, akit soha nem kértem, hogy legyen a "témavezetőm", ezért inkább az egyik legfontosabb ember lett az életemben.

Ha valaha fele olyan kiváló mentor, vagy tanár leszek, mint a felsorolt emberek (átlagosan:), akkor nagyon sikeres életem lesz. Köszönöm a Bolyai Intézet oktatóinak és munkatársainak az inspiráló közösséget és a barátságukat. Szerencsés vagyok, hogy fizikusként jó matek kurzusokat választottam. Imádtam dr. Polner Mónika, dr. Krisztin Tibor és dr. Waldhauser Tamás óráit. Köszönöm dr. Karsai Jánosnak és dr. Szabó Tamásnak a bíztató beszélgetéseket. Dr. Vizi Zsoltot és Bogya Norbertet (szigorúan ebben a sorrendben, hogy Zsolt végre első szerzet lehessen egy névsorban), szeretném külön megemlíteni a trauma feldolgozó csoportunkból. A turbulens időszak ellenére nagy élmény volt együtt dolgozni.

Köszönöm dr. Árpádffy-Lovas Tamásnak a fordításban nyújtott segítségét, értékes észrevételeit és azt, hogy lelkiismeretesen gondozta a szöveget. Nélküle ez a dolgozat sokkal szerényebb lenne a szó szoros értelmében.

Legvégül pedig szeretném megköszönni a legfontosabb emberek támogatását, a párom, Horváth Mari, a szüleim, a nagymamám és a barátaim végtelen szeretetét. Nélkülük sokkal nehezebb lenne minden. Velük sokkal könnyebb minden.

Legeslegvégül pedig szeretném megköszönni két kiváló pszichológusnak, dr. Papp Edinának és dr. Menis Yousrinak, hogy segítettek megérteni a felsorolt nagyszerű emberek tanításait.

1

Foreword

In this thesis, we address the modeling issues of cell movement and division with a special focus on the phenomenon of volume exclusion in lattice-based, exact stochastic simulation frameworks.

After a brief introduction, in Chapter 2, we present the central modeling framework of the thesis, Gillespie's stochastic simulation algorithm (SSA) for well-stirred chemical systems. We then show how Baker and Simpson [35] applied the SSA to model populations of cells that move, divide and die on a lattice. Finally, with a short example we also point out the limitations of this algorithm in the case of a crowded environment.

In Chapter 3, Preliminaries, we introduce the basic assumptions in our lattice based framework and the new reaction types we shall incorporate to our novel modeling framework: contact-inhibited, contact-promoted, and spontaneous reactions.

Reactions are random events that may occur with a given probability during the life of the simulated cells according to the considerations of the model. The rate of contact-inhibited reactions of a cell decreases as the number of its free neighbors increases. Therefore, migration and proliferation naturally fall into this reaction type, and further physiological processes that are inhibited in a crowded environment may also be accounted for in this category. In a similar manner, the rate of contact-promoted reactions of a cell increases as the number of occupied neighbors increase. On the other hand, the rate of spontaneous reactions is not affected by the number of occupied neighboring lattice sites. We shall explain these concepts in details in 3.2.

To the best of our knowledge, these reaction types have not been taken into account in lattice-based stochastic simulations of cell cultures. These new types of reactions may be easily applied to complicated systems, enabling the generation of biologically feasible stochastic cell culture simulations.

In Chapter 4 we describe in detail the extended version of Baker and Simpson's method which we call the Prompt Decision Method (PDM). Then, we propose a novel exact method, called Reduced Rate Method (RRM), that is based on different premises and substantially quicker than the PDM for large number of cells.

Furthermore, we show that the PDM algorithm and our RRM algorithm are mathematically equivalent in the sense that the next reaction to be realized and the corresponding waiting time both belong to the same reaction and time distributions in the two approaches – even with the newly introduced reaction types. However, the number of calculations required by the two algorithms to produce a single stochastic realization of the system can be drastically different. At the end of the chapter we present numerical comparisons of the running times and two toy models to illustrate the effect of contact dependent and spontaneous reactions.

In Chapter 5 we take a step back and incorporate realistic cell cycle length to our models. We present the surprising behavior of the dynamics due to the non-Markovian approach: the synchronization of cells. We also fit our model to *in vitro* experimental data measured by Vitadello et al. [46].

In Chapter 6, we bring together the tools we have developed so far to propose the Go or Grow algorithms. These methods are created to model cells with realistic cell cycle distribution in a lattice based framework. Finally, in Chapter 7, a delay differential equation mean field model is presented and compared with the stochastic model.

Exact, agent-based, stochastic methods of cell culture simulations seem to be undervalued and are mostly used as benchmarking tools to validate deterministic approximations of the corresponding stochastic models. Our proposed methods are exact, they are easy to implement, have a high predictive value, and can be conveniently extended with new features. We believe these approaches promise a great potential.

2

Introduction

With the emergence of personalized therapeutic approaches, computational drug testing, and even artificial tissue engineering modeling of cell cultures have become a prominent area of mathematical and *in silico* biology. Understanding the complex collective behavior emerging from simple phenomena, such as migration, proliferation, or intercellular communication of individual cells is essential to be productive in these fields. Researchers use a wide variety of mathematical and numerical tools to describe the behavior of populations. Parunak et al. [45] distinguishes between equation-based models and agent-based models.

2.1 Equation based models

Equation-based models are typically composed of deterministic or stochastic approaches and provide well-established methods for both analytical and numerical treatments of the problem. For example, to determine the evolution of a population over time or to predict certain events. They are also used to prove or rule out specific behaviors of the system: such as extinction of a population, chaotic phenomena, or the absence of periodic behavior, for instance.

Deterministic models consist of systems of ordinary, partial, or delay differential equations and are used with great success to predict and analyze the behavior of real-world systems [2] [32]. These models usually treat the state variables, as well as the actual individuals in the population model, as if they were continuous quantities (densities or concentrations), although, populations or even chemical solutions are composed of discrete individuals (particles). Moreover, for these models to be predictive, one has to assume that the size of the population is large or apply corrections into the model. We refer to the infamous atto-fox problem and Fowler's article [20] for a detailed discussion and feasible solutions. The time evolution of the model can be obtained with various numerical methods.

When it is important to work with discrete individuals or when the number of indi-

viduals is small, the theory of stochastic processes may provide a feasible alternative, especially when stochastic effects play an important role in the evolution of the system. These are mostly Markovian stochastic processes with a discrete state space in discrete or continuous time [3] [32] [40]. Analytical treatment of complex stochastic models is usually not trivial, the gold standard among the numerical approaches is Gillespie's stochastic simulation algorithm (SSA) [27] [28] and its variants [12] [29]. We shall adapt this algorithm to the simulation of cell cultures in this article.

2.2 Agent based models

Agent-based models (ABMs) or individual-based models [21] [45] [47] are feasible alternatives to the equation based models. ABMs, unlike continuum models, regard every particle as an individual that follows a prescribed set of rules. Information about the system can be obtained by analyzing the collective behavior of the agents with statistical methods. This technique offers a broader description of individual behavior and accounts for stochastic effects caused by the finite number of agents and their interactions. As Bonabeau [8] points out: "Individual behavior exhibits memory, path-dependence, and hysteresis, non-Markovian behavior, or temporal correlations, including learning and adaptation". However, the rigorous mathematical analysis is usually difficult, and simulations can be resource-intensive. ABMs are subject for existential-type proofs, to prove the existence of a certain behavior of the population, even in cases when the formulation of a corresponding equation-based model is not evident [47]. Equation-based models have a well-developed mathematical theory, yet agent-based models are mostly expressed as computer routines.

In the case of equation-based population models, we usually decrease the complexity of the investigated phenomena by applying a classification on the individuals according to the trait we intend to examine – assuming that the differences in this attribute can be averaged out within a class or can be accounted for with a corresponding distribution. We usually aim to keep the number of classes, and, hence, the number of parameters required as low as possible. Meanwhile, the model remains of a manageable size, yet it still captures the essence of the modeled phenomena. In contrast, ABMs are mainly used when the number of these classes is close to the number of members in the population, i. e., the classifying traits are parameter-specific to the individuals and highly influence their behavior. Therefore, in these complex cases, we may use ABMs and focus on simulations.

2.3 Gillespie's stochastic simulation algorithms

Gillespie [27, 28] formulated the SSA in order to produce computer-based numerical experiments of chemical reactions. It is assumed that chemical species or reactants $(S_1, S_2, \ldots, S_m)$ interact via reaction channels $(R_1, R_2, \ldots, R_M)$ in a well-stirred reaction chamber.

Time evolution of the amount of reactants $(\underline{S}(t) = (S_1(t), S_2(t), \ldots, S_m(t)))$ is described by a Markov process that is continuous in time $(t \in [0, \infty))$ and discrete in state space $(\underline{S}(t) \in \mathbb{N}_0^m)$. The SSA obtains stochastic realizations of the system by answering two questions [27]:

(i) In what time (τ) will the next reaction occur?

(ii) What kind of reaction $(R_1, R_2, \ldots, R_M)$ will the next reaction be?

To answer these questions, Gillespie developed two separate but mathematically equivalent variants of the SSA: the 'direct method' (DM) and the 'first-reaction method' (FRM). The main difference is in the way of deciding which reaction channel to fire. The *first-reaction method* generates M random time values $(\tau_1, \ldots, \tau_M)$ for the M possible reactions from the corresponding exponential distributions and selects the reaction channel with the least time to fire. Alternatively, the *direct method* only requires two random numbers (one from a uniform distribution and one from an exponential distribution). Considering the fact that if there are more than 2 possible reactions then the FRM requires more computation and memory than the DM, moreover drawing M random numbers from exponential distributions requires calculating M logarithms, the computational cost of the *first-reaction method* is greater [27]. For a comparison of the variants of the SSA, we refer to Gillespie's original article [27] and to the work by Cao et al. [12].

The direct method algorithm

Starting from the vector of initial values $(S_1(0), \ldots, S_m(0)) \in \mathbb{N}_0^m$, the time series of the state variables $\underline{S}(t)$ is generated by constantly updating the state variables in properly generated subsequent times (τ) according to the answers to question (i) and (ii). To this end, the *reaction probability density function*, $P(\tau, \mu)d\tau$, is defined [27]. That is, at time t, the next reaction in the reaction chamber will occur in the differential time interval $(t + \tau, t + \tau + d\tau)$ with the probability $P(\tau, \mu)d\tau$, and will be an R_μ reaction $(\mu \in 1, \ldots, M)$.

With the procedure of conditioning [27], the two-variable density function $P(\tau, \mu)$ can be written as the product of two one-variable probability density functions:

$$P(\tau, \mu) = P_1(\tau) \cdot P_2(\mu|\tau). \tag{2.1}$$

In particular, from [27], it turns out that

$$P(\tau, \mu) = a_\mu \cdot e^{-\tau \cdot a},$$

where the symbol a_μ ($\mu \in \{1, \ldots, M\}$) stands for the so-called propensity function that characterizes reaction R_μ, and may depend on the quality and the actual number of the reactants, their number of combinations or environmental factors, etc. In [27] it is assumed that $a_\mu = c_\mu h_\mu$, where c_μ corresponds to the rate constant of the reaction R_μ and h_μ is the number of distinct combinations of reactants participating in R_μ, being present in the reaction chamber at time t. Although a_μ and h_μ clearly depend on time via the number of reactants at time t, the time dependence is usually not indicated to simplify the notations. Variable a stands for the sum of the propensity functions: $a = \sum_{\mu=1}^{M} a_\mu$. In Table 2.1 we summarized the formula for h_μ in case of some important reactant combinations. For detailed explanation consult [27].

Reaction type	Reactant combinations
$S_i \rightarrow$ products	$h_\mu = S_i(t)$
$S_i + S_k \rightarrow$ products $(i \neq k)$	$h_\mu = S_i(t)S_k(t)$
$2S_i \rightarrow$ products	$h_\mu = S_i(t)(S_i(t) - 1)/2$
$S_i + S_k + S_l \rightarrow$ products $(i \neq k \neq l \neq i)$	$h_\mu = S_i(t)S_k(t)S_l(t)$
$S_i + 2S_k \rightarrow$ products $(i \neq k)$	$h_\mu = S_i(t)S_k(t)(S_k(t) - 1)/2$
$3S_i \rightarrow$ products	$h_\mu = S_i(t)(S_i(t) - 1)(S_i(t) - 2)/6$

Table 2.1: **Common reactant combinations in chemical reactions** A subset of chemical species $\{S_1, \ldots, S_n\}$ may react in various combinations h_μ in the reaction chamber. The first row corresponds to the decay of species S_i. The second and third row correspond to the two possible bimolecular reactions: the reactants are different or the same chemical species, respectively. Rows 4, 5, and 6 correspond to the possible trimolecular reaction types in which all three molecules are of different species, two are of the same species, and all of the same species, respectively.

Let $P_1(\tau)d\tau$ be the probability that the next reaction will occur between times $t+\tau$ and $t + \tau + d\tau$, independent of which reaction it might be. Similarly, $P_2(\mu|\tau)$ is the probability that the next reaction will be the R_μ reaction, given that the next reaction occurs at $t+\tau$. The probability $P_1(\tau)d\tau$ is the marginal distribution for τ, and obtained by summing the reaction probability density function over all possible μ values:

$$P_1(\tau) = \sum_{\mu=1}^{m} P(\tau, \mu) = a \cdot e^{-\tau \cdot a}. \tag{2.2}$$

Substituting $P_1(\tau)$ into (2.1) and solving for $P_2(\mu|\tau)$ we obtain

$$P_2(\mu|\tau) = \frac{a_\mu}{a}. \tag{2.3}$$

From which it is clear that $\int_0^\infty P_1(\tau)d\tau = \int_0^\infty a \cdot e^{-\tau \cdot a} = 1$ and $\sum_{\mu=1}^{M} P_2(\mu|\tau) =$

$\sum_{\mu=1}^{M} a_\mu/a = 1$. The so-called sojourn time (or waiting time) τ between consecutive reactions is an exponentially distributed continuous random variable with parameter a, the sum of the reaction propensity functions: $\tau \sim Exp(a)$. The algorithm randomly selects reaction index μ, which is a discrete random variable, from the discrete (categorical) distribution (Eq. 2.3).

One may generate such τ from Eq. (2.2) by selecting $r_1 \in (0,1)$ with a continuous uniform distribution and calculate $\tau = 1/a \ln(1/r_1)$. To generate a reaction index μ from Eq. (2.3), one has to select $r_2 \in (0,1)$ with a continuous uniform distribution and choose μ to be the index for which $\sum_{\nu=1}^{\mu-1} a_\nu < r_2 a \leq \sum_{\nu=1}^{\mu} a_\nu$ holds. This is discussed in details in [27], and most programming languages have highly optimised built-in functions to execute this task.

Upon choosing reaction R_μ and τ, the algorithm has to obtain the change vector $(\Delta S_1, \ldots, \Delta S_m) \in \mathbb{Z}^m$ that describes the change in the amount of reactants according to R_μ. Finally, both time and the state variables have to be updated. Gillespie's SSA requires a simple data structure to record the time evolution of the system: a list containing the subsequently appended values $(S_1(t), \ldots, S_m(t))$ after each iteration. With these notations and probability distributions, the direct method SSA can be summarized as follows.

Direct Method Stochastic Simulation Algorithm

1. **Initialization:**
   (i) Initialize data structure that stores time series.
   (ii) set $t \leftarrow 0$,
   (iii) set initial values $(S_1(0), \ldots, S_m(0)) \in \mathbb{N}_0^m$,
   (iv) prescribe halting conditions H.
2. **Calculate propensity functions** a_μ for all $\mu \in \{1, 2, \ldots, M\}$, calculate a.
3. **Decide when the next reaction will occur:** choose τ according to Eq. (2.2),
4. **Reaction selection:** draw μ according to Eq. (2.3),
5. **Update time and state variables:**
   (i) set $t \leftarrow t + \tau$
   (ii) obtain the change $(\Delta S_1, \ldots, \Delta S_m) \in \mathbb{Z}^m$ in number of reactants according to R_μ,
   (iii) change the number of molecules according to

$$(S_1(t+\tau), \ldots, S_m(t+\tau)) = (S_1(t), \ldots, S_m(t)) + (\Delta S_1, \ldots, \Delta S_m).$$

   (iv) Update data structure.
6. **Halt if** $H = True$ **else continue the process with Step (2).**

The first reaction method algorithm

The *first-reaction method* algorithm generates M so-called *putative times* $(\tau_1, \ldots, \tau_M)$ for the M possible reactions. Each τ_ν is drawn from exponential distribution with parameter a_ν: $\tau_\nu \sim Exp(a_\nu), \nu \in \{1, \ldots, M\}$. Then, the algorithm selects reaction with index μ for which the corresponding putative time is the minimal element of the set of the generated putative times: $\tau_\mu = \min\{\tau_1, \ldots, \tau_M\}$. Finally, let the waiting time until the next reaction be $\tilde{\tau} := \tau_\mu$.

The DM and FRM algorithms are mathematically equivalent since in a given state $\underline{S}(t)$ of the system the probability of selecting reaction R_ν for all $\nu \in \{1, \ldots, M\}$ is the same in both approaches and the waiting times between two consecutive reactions τ and $\tilde{\tau}$ are both of exponential distribution with parameter a. In fact, this follows from the following well known theorem.

Theorem 2.1. *Suppose $\tau_1, \ldots, \tau_M$ are independent exponentially distributed random variables with rate parameters a_ν and cumulative distribution functions $F_{\tau_\nu}(x) = 1 - e^{-a_\nu x}$ for $\nu \in \{1, \ldots, M\}$. Then*
(i) $\tau_\mu = \min\{\tau_1, \ldots, \tau_M\}$ is exponentially distributed, with parameter $a = \sum_{\nu=1}^{M} a_\nu$ and
(ii) index μ is chosen with probability $P(\tau_\mu = \min\{\tau_1, \ldots, \tau_M\}) = a_\mu / a$.

Proof. (i) Considering the cumulative distribution functions, direct calculation yields:

$$\begin{aligned} P(\min\{\tau_1, \ldots, \tau_M\} < \tilde{\tau}) &= 1 - P(\tau_1 > \tilde{\tau}, \ldots, \tau_M > \tilde{\tau}) \\ &= 1 - P(\tau_1 > \tilde{\tau}) \ldots P(\tau_M > \tilde{\tau}) \\ &= 1 - (1 - F_{\tau_1}(\tilde{\tau})) \ldots (1 - F_{\tau_M}(\tilde{\tau})) \\ &= 1 - e^{-a_1 \tilde{\tau}} \ldots e^{-a_M \tilde{\tau}} = 1 - e^{-a\tilde{\tau}} \end{aligned}$$

Thus, $\tilde{\tau} \sim Exp(a)$, which proves the first part of the theorem.
(ii) Let $I = \{1, \ldots, \mu - 1, \mu + 1, \ldots M\}$.

$$\begin{aligned} P(\tau_\mu = \min\{\tau_1, \ldots, \tau_M\}) &= \int_0^\infty P(\tilde{\tau} = \tau_\mu) \cdot P(\tilde{\tau} < \tau_\nu, \ \forall \nu \in I) d\tilde{\tau} \\ &= \int_0^\infty a_\mu e^{-a_\mu \tilde{\tau}} \prod_{\nu \in I} e^{-a_\nu \tilde{\tau}} d\tilde{\tau} = a_\mu \int_0^\infty e^{-a\tilde{\tau}} d\tilde{\tau} = \frac{a_\mu}{a}. \end{aligned}$$

■

First Reaction Method Stochastic Simulation Algorithm

1. **Initialization:**
   (i) Initialize data structure that stores time series.
   (ii) set $t \leftarrow 0$,
   (iii) set initial values $(S_1(0), \ldots, S_m(0)) \in \mathbb{N}_0^m$,
   (iv) prescribe halting conditions H.
2. **Calculate propensity functions** a_ν for all $\nu \in \{1, 2, \ldots, M\}$.
3. **Calculate putative times** $\tau_\nu \sim Exp(a_\nu)$ for all $\nu \in \{1, 2, \ldots, M\}$,
4. **Reaction selection:** choose μ for which $\tau_\mu = \min\{\tau_1, \ldots, \tau_M\}$,
5. **Update time and state variables:**
   (i) set $t \leftarrow t + \tau_\mu$
   (ii) obtain the change $(\Delta S_1, \ldots, \Delta S_m) \in \mathbb{Z}^m$ in number of reactants according to R_μ,
   (iii) change the number of molecules according to

$$(S_1(t+\tau), \ldots, S_m(t+\tau)) = (S_1(t), \ldots, S_m(t)) + (\Delta S_1, \ldots, \Delta S_m).$$

   (iv) Update data structure.
6. **Halt if** $H = True$ **else continue the process with Step (2).**

Strengths and limitations of the SSA

It became apparent that Gillespie's approach is particularly beneficial in case of certain biological systems when only a small quantity of reactants are present in the reaction chamber, thus stochastic fluctuations play an important role in the modeled physiological process. A well-known example of a process of this type is intracellular gene regulation [36], where only a few copies of the corresponding regulatory molecules are present in the cell [19]. In addition to chemical kinetics [17] [30] [44], the SSA has been proven useful for studying epidemic spread [1] [7] and describing ecological phenomena [15], among a variety of further applications.

Since the SSA produces the time evolution of the system by recording every single reaction, it may be time consuming for systems that are characterized by large state space or contain reactions that have a high rate compared to other reaction rates in the system. There have been several attempts to solve these problems, the most influential of these may be the one by Gibson and Bruck [25]. Their Next Reaction Method (NRM) is an extension of Gillespie's first reaction method. They introduced the so-called reaction dependency graph, that specifies which propensity functions need to be changed after the selected reaction R_μ is executed. Cao et al. [12] made a detailed comparison of the DM, the FRM, and the NRM, and found that "even with the best data structure, the NRM is less efficient than the DM, except for a very specific class of problems." The NRM handles a complicated data structure and according to

Schwehm [41]: "the simulator engine spends most of its execution time maintaining the priority queue of the tentative reaction times".

Cao et al. (2004) [12] optimized the algorithm by using the clever observation that in case of robust systems, the reactions are usually multiscale, which means that some of the reactions fire more frequently than others. Therefore, by sorting the indices of the reactions in decreasing order based on how often they fire, they were able to speed up the reaction selection step. Moreover they adopted the core idea of the NRM [25] and only updated the propensities of the reaction channels that were affected by the last reaction. The idea of separating fast and slow reactions was further improved in the *slow-scale stochastic simulation algorithm* by Cao et al. (2005) [10]. Finally, we would also like to mention Gillespie's tau-leaping method [29] and its improved versions [9] [11]. It is an approximate method analogous to the Euler method in deterministic systems, and it executes several reactions in a time interval before updating the propensities.

2.4 Modeling cellular behavior

Cells are mesoscopic entities in multicellular organisms. Depending on the scale at which their population is examined and on the complexity of the investigated phenomena, one may find an appropriate approach among the modeling paradigms described above. However, a typical approach is to approximate the behavior of the population with a continuum mean field model [16], such as a variant of the logistic equation or the Fisher–Kolmogorov equation [4] [5], incorporate corrections if spatial effects play an important role [6] [35], and analyze the obtained deterministic model with rigorous analytical means. Thus, the underlying ABM play the role of a benchmarking tool. We would like to acknowledge the early, but still significant review by Fredrickson et al. [22] and the excellent review by Charlebois and Balázsi [13].

The development of cell cultures is influenced by a number of important phenomena. The two most studied ones may well be cell movement and cell division. These two phenomena play notable roles in several important processes, including wound healing, tissue development, or even cancer propagation (Maini and Baker [34]; or for a more experimental viewpoint, see Méhes and Vicsek [37]). In most cases, approximating cells as point-like particles may not be sufficient, since spatial effects usually affect the development of the cell population [6]. Volume exclusion is such a significant spatial effect, which expresses that two cells cannot occupy the same volume in space, thus cells are physical obstacles to each other. Technically, the 'excluded volume' of a cell is the space that is unavailable to other cells in the population as a result of the presence of the first cell.

2.4.1 The lattice-based variant of the DM-SSA

Probably the most convenient way to incorporate volume exclusion into cell culture models is discretizing the space by an underlying lattice. Even though there are several well known on-lattice methods: cellular automaton models, cellular Potts models, or cellular lattice gas models, we improved a variant of Gillespie's SSA by Baker and Simpson that is extended to work in spatially inhomogeneous environment and incorporates volume exclusion [6]. The algorithm incorporates volume exclusion and it was used as a tool to produce numerical experiments to benchmark the mean field approximations of the authors.

Spherical cells (with uniform size of diameter d) are placed on a finite, regular, square lattice of two dimension with grid constant d and von Neumann neighborhood. A cell may engage in three events: movement (R_1), division (R_2) or death (R_3) with the corresponding rates r_1, r_2, r_3, in order. In every iteration, the corresponding propensity functions $a_\mu = r_\mu N$, $\mu \in \{1, 2, 3\}$, $a = a_1 + a_2 + a_3$ are calculated, where $N = N(t)$ is the number of cells in the population at t.

A random waiting time is generated from the distribution $\tau \sim Exp(a)$ and the time of the system is updated: $t \leftarrow t + \tau$. A target cell is randomly selected from the population with discrete uniform distribution. Then, the next possible reaction is selected from the distribution $P(\mu|\tau) = a_\mu/a$ and executed on the target cell with the following constraints:

- in the case of movement, the cell moves to a randomly selected neighboring lattice site if it is empty; if the lattice site is occupied, the reaction is rejected,
- in the case of division, the cell places a daughter cell on a randomly selected neighboring lattice site if it is empty and the population size is increased by 1; if the lattice site is occupied, the reaction is rejected,
- in the case of death, the cell dies without any particular restriction, in this case the lattice site of the cell is turned empty and the size of the population is decreased by 1.

Thus, whenever movement or proliferation cannot be executed because of the volume exclusion, we reject the selected reaction. However, we always keep the generated sojourn time τ. We will not present the algorithm in any more detail here, as we will soon introduce an extended version of it.

A major limitation of the approach is that the target cell, that is about to take action in a given iteration of the simulation, is selected randomly from all the cells in the population with a discrete uniform distribution. Thus, there is a chance that the algorithm selects a cell that is unable to move or divide. In fact, with nonzero proliferation rate and zero death rate the population grows monotonically. On a finite lattice the probability of choosing a cell that cannot engage in any of the possible events

grows as the cells would have less empty adjacent sites over time.

For an extreme example, suppose that the grid is finite of size $K = n_1 \cdot n_2$ with a periodic boundary condition, thus the maximum number of cells on the lattice is K, and suppose that the current number of cells is $K - 1$ while the death rate is zero. In such a setting, it is easy to show that the selected event is expected to be rejected $K - 1$ times before acceptance. Indeed, in these cases, i. e., in a crowded environment, it is computationally intensive to run the simulation, not only because *"nothing seems to happen"* most of the time, but also because generating the random waiting time from an exponential distribution requires calculation of a logarithm, which is a rather demanding task. Clearly, in this framework we cannot just choose the cell 'wisely' or choose an empty adjacent site 'on purpose' to speed up the simulation, as we are also sampling the distribution of the waiting time. Therefore, the number of times we discard a reaction is just as important as the reaction itself.

In our reduced rate method we are using another approach to define the reactions and the corresponding propensities. In addition, the algorithm has a different basis, so that all the reactions that are drawn are always executed.

2.4.2 The lattice-based variant of the FRM-SSA

We may also formulate the reaction selection and the waiting time selection of the previous lattice based algorithm in terms of the First Reaction Method. Naturally, according to Thm. 2.1, the two approaches would be mathematically equivalent if the implementation of the target cell selection and the way we decide whether to accept or reject the reaction remains the same.

3

Preliminaries

3.1 Lattice discretization and state space

All of the presented methods make the following assumptions: consider spherical cells of equal size, with diameter d, place the cells on a 2-dimensional square lattice with grid constant d and size $(n_1 \times n_2)$ where $n_1, n_2 \in \{1, 2, \dots\}$ are finite or infinite. Thus, the lattice has unit length spacing, allowing us to compare various cell lines with their characteristic cell sizes. We use the terms grid and lattice interchangeably. Also the terms neighbor, grid neighbor and adjacent site refer to the same concept.

We assume that the cells use chemical or mechanical mechanisms to sense each other's presence – specifically, whether or not their immediate grid neighbors are occupied. In the case of a finite lattice, we use a reflective boundary condition: the boundaries of the grid are perceived by the cells as occupied neighbors. A cell has only free and occupied neighbors: an adjacent site that is not free is occupied and vice versa. In our example models, we assume von Neumann neighborhood, thus a given cell has four adjacent neighbors on the lattice, $\eta = 4$ (up, down, left, right). Consequently, a cell may have $j \in \{0, \dots, \eta\}$ free and $(\eta - j)$ occupied neighbors. From now on, index j will always denote the number of free adjacent sites of the concerned cell or cells. The method we shall discuss may be applied immediately to three dimensions, other lattice types, other neighborhood types (see Fig.3.1), and even to lattices with special shapes.

We assume that the principle of volume exclusion applies for all cells: there can be at most one cell per lattice site under the above conditions. Thus, by volume exclusion, the capacity of a finite grid with size $(n_1 \times n_2)$, where $n_1, n_2 < \infty$, is $K = n_1 \cdot n_2$. Which means that the lattice can support at most K cells.

State descriptors Let $t \in [0, \infty)$ represent time, and $N_j(t)$ denote the number of cells in the population with j free neighbors at time t. Then let $N = N(t) = \sum_{j=0}^{4} N_j(t)$ be the total number of cells in the grid at time t (the time dependence is usually not

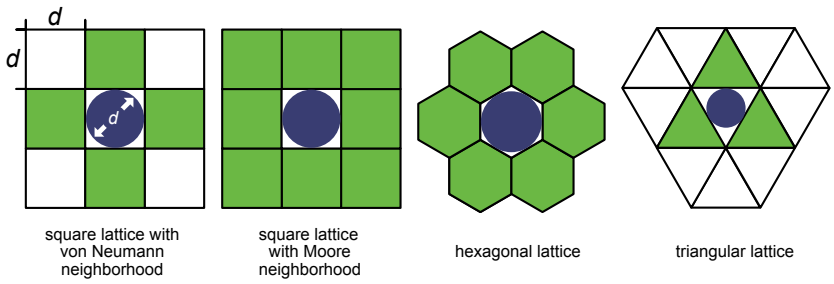

Figure 3.1: **Common lattice types in two dimensions.** Square and hexagonal lattices are widely used in lattice-based cell population models, triangular lattices are mostly used in lattice gas models. Each lattice contains one circular cell (purple) and the adjacent lattice sites are shaded green. The number of neighboring sites in these types are $\eta = 4, 8, 6, 3$ in order. Thus a cell may have $j \in \{0, \dots, \eta\}$ free and $(\eta - j)$ occupied adjacent sites. Throughout the paper, we use a square lattice with von Neumann neighborhood (leftmost figure). In this case we assume that the cell has diameter d matching with the grid constant of the lattice as indicated on the figure.

denoted for simplicity).

We introduce the state descriptor

$$\underline{\mathrm{X}}(t) = (C_1(t), \dots, C_N(t), S_1(t), \dots, S_m(t)) \tag{3.1}$$

where $C_i(t)$ is the coordinates of the cell with index $i \in \{1, \dots, N\}$ at time t on the lattice. By volume exclusion, it is clear that no two cells have the same coordinates at t. Scalar variables $S_s(t) \in \mathbb{N}_0$ $(s \in \{1, \dots, m\})$ store the amount of chemical species present in the cell space at time t. In order to ease the notation we refer to the state of the cells and the state of chemicals the following way:

$$\underline{\mathrm{C}} = (C_1(t), \dots, C_N(t)) \text{ and } \underline{\mathrm{S}} = (S_1(t), \dots, S_m(t)).$$

Fig.3.2 shows the time evolution of a cell population along with a snapshot of the lattice configuration at time t_*.

3.2 Reactions

We use the term 'reactions' in the sense it is used in chemical modeling. We also refer to movement, division or cell death as reactions. We will adopt the usual notation to denote the reaction index by $\mu \in \{1, \dots, M\}$. Our goal is to provide a generalized and therefore, a rather flexible modeling framework for cell culture simulation that can be utilized to capture exceptionally complex mechanisms occurring in cell cultures. In order to achieve this, any number and combination of the reactions described in this section can be included in the model the researcher is working on, which covers a multitude of possible activities in a modeled cell culture with very few technical restrictions.

It is important to keep in mind that selecting the reaction and selecting the target cell taking the action are executed in substantially different ways in the PDM and

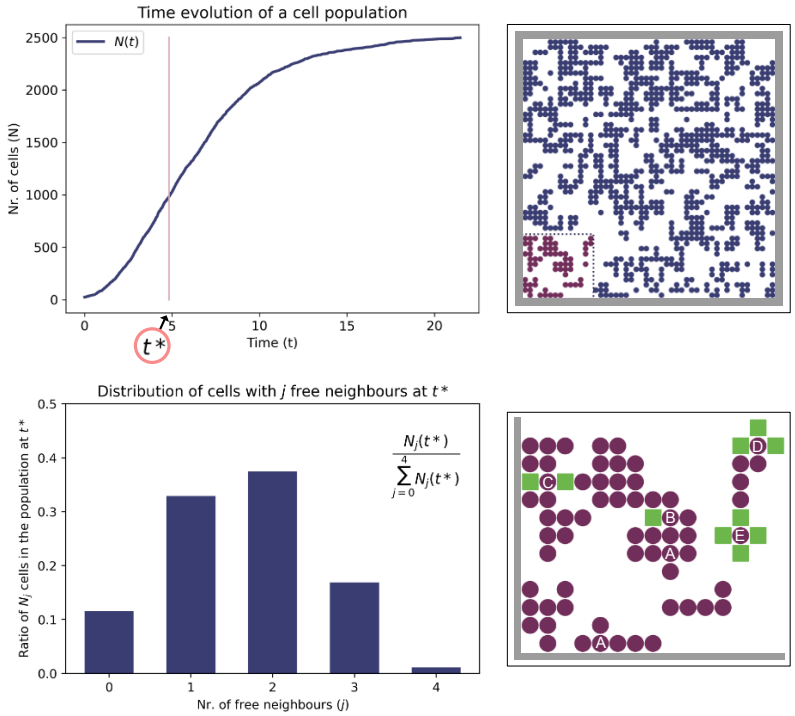

Figure 3.2: **Time evolution of a cell culture.** The *top left figure* shows the time evolution of a proliferating cell population on a lattice with carrying capacity $K = 2500$. In this experiment we chose the proliferation rate to be 1 and we assumed that no other reaction can happen to the cells. The *bottom left figure* shows a snapshot of the ratio of cells with j free neighbors in the population at $t* = 4.8$ (see the formula in the figure). In this state the total number of cells is $N(t*) = 1000$ and very few cells – only 1.1% of the population – have $j = 4$ free adjacent sites. In fact, most of the cells have $j = 2$ neighbors at $t*$. The *top right figure* shows a snapshot of the lattice configuration at $t*$ and the *bottom right figure* is an enlarged part of the highlighted area of the whole lattice. On this latter figure we selected some cells: A, B, C, D, and E having $0 \dots 4$ free adjacent sites, in order. Their corresponding free neighbors are colored green and the border of the lattice is marked in gray. We assume that the cells perceive the boundary as an occupied neighbor – thus both 'A' cells have zero neighbors at $t*$.

RRM algorithms. At this point, we only focus on what happens to the state variables if a reaction is '*successful*' in the sense that it is selected and *can be* executed. The differences will be highlighted in the following sections, when we introduce the methods. Next, we describe the reactions (Table 3.1.) that are introduced into the model during the prompt decision and reduced rate methods.

Reaction		In a finite environment...
Contact-inhibited	Movement ($\otimes$) Division ($\otimes$) Cell death Cellular biochemical reaction	the per capita rate decreases as the population grows,
Spontaneous	Cell death Cellular biochemical reaction Non-cellular biochemical reaction	the per capita rate is independent of pop. size,
Contact-promoted	Cell death Cellular biochemical reaction	the per capita rate increases as the population grows.

Table 3.1: **Common types of reactions that may occur in a cell culture**. Symbol $\otimes$ indicates volume exclusion. The reactions defined in this table are intended to provide a modeling framework for the exact simulation of cell cultures, that can be easily adapted to the problem the reader is considering. A prominent aspect of this work would be to show that, using these types of reactions, the PDM and RRM algorithms are mathematically equivalent. For detailed explanations and possible interpretations see Sec. 3.2.

3.2.1 Non-biochemical reactions

We define the contact-dependent and contact-independent movement, proliferation, and death of the cells as non-biochemical reactions.

Contact-inhibited reactions: movement, proliferation, and cell death Contact-inhibited reaction rates of a cell decreases as the number of its free neighbors decreases. Typical examples are the volume excluding reactions, movement and proliferation, but we shall define two other types of such reactions as well. First, consider the movement and division of the cells.

Cells may perform random walk on the grid with a motility rate r_{mot} per unit time, moving to another adjacent lattice site or divide with a proliferation rate r_{div} per unit time giving rise to another agent, such that the principle of volume exclusion applies. Consequently, a movement rate of r_{mot} and a division rate of r_{div} can only be measured for a sufficiently sparse population of cells. In a dense population, the cells are expected to interfere with each other, thus lower r_1 and r_2 rates are observed.

After selecting the movement or division reaction, a target cell is selected from the population. If the reaction is a *successful* movement, then the cell is moved to its selected free adjacent position on the lattice and its coordinates are updated in $\underline{\mathrm{X}}(t)$. If the reaction is a *successful* cell division, the target cell places a daughter cell (X_{N+1}) onto a selected free adjacent site on the lattice and the new cell along with its coordinates is appended to the state descriptor $\underline{\mathrm{X}}(t)$.

Cell-cell communication with local cellular signals are crucial factors to maintain a cell culture. In fact, abandoned cells that do not respond to the environment are destined to die in a healthy tissue. *Contact-inhibited cell death* may be a model for death due to the lack of these local effects. In case of a cell death reaction, the target cell is removed from the lattice and its coordinates are deleted from the state descriptor $\underline{\mathrm{X}}(t)$.

Spontaneous (contact-independent) reaction: cell death It is also assumed that cells can be involved in events that are not influenced by the number of free adjacent sites. Such a reaction is, for example, *spontaneous cell death*, which is a standard model of naturally occurring death events due to the finite lifespan of the individuals. Spontaneous cell death affects all cells equally, regardless of their adjacencies, and when it occurs, the cell that is randomly selected from the whole population would be removed from the lattice, and also from the state descriptor.

Contact-promoted reaction: cell death The rate of contact-promoted reactions increase as the environment gets more crowded. Thus, contact-promoted cell death may serve as a model for death due to competition for resources, such as oxygen or

glucose, between the cells. For example, in tumor pathology, it is a well known effect that the core of some tumors suffer from hypoxia and substrate deprivation due to the lack of sufficient vascularization. Contact-promoted death may be a good candidate to model death in such cancer models. In case of a *successful* contact-promoted death, the target cell is removed from the lattice and also from the state descriptor.

3.2.2 Biochemical reactions

These reactions include (bio)chemical reactions that do or do not involve the cells themselves, called cellular and non-cellular biochemical reactions respectively. Examples of these reactions would be cell-cell communication, product synthesis (proteins, nucleic acids, hormones, signaling factors, toxins etc.), or substrate consumption.

Cellular biochemical reactions We assume that a subset of reactants $\{S_1, \ldots, S_m\}$ in reactant combination h_μ and a cell $X_i(t)$ may interact and engage in *cellular biochemical reactions*. These reactions may be contact-dependent (inhibited or promoted) or spontaneous reactions. Cellular biochemical reactions may have four types of outcomes: i) cell movement, ii) division, iii) death or iv) a $(\Delta S_1, \ldots, \Delta S_m) \in \mathbb{Z}^m$ change in the number of the reactants. The state descriptor has to be updated according to the ways we discussed so far in this section or in Sec. 2.3.

Non-cellular biochemical reactions A subset of the reactants $\{S_1, \ldots, S_m\}$ may also participate in spontaneous *non-cellular chemical reactions* in reactant combination h_μ. Non-cellular biochemical reactions are independent from the cells and only depend on the reactants. These reactions may result in a $(\Delta S_1, \ldots, \Delta S_m) \in \mathbb{Z}^m$ change in the number of the reactants. The state descriptor has to be updated according to the ways we explained in Sec. 2.3.

Note that if we exclude all other reactions, including movement and cell division, and keep only the non-cellular chemical reactions, the cell culture simulation algorithm reduces to Gillespie's SSA.

3.2.3 Some restrictions concerning the reactants

We intend to show that the PDM and RRM algorithms are mathematically equivalent ways to obtain realizations of a given stochastic process concerning a cell culture. In order to achieve this, we must apply some restrictions on the reactants and the state space, limiting the algorithms for the sake of this proof. We assume that the state of the cells is only determined by their individual positions on the lattice (and its neighborhood), and they do not differ in any other way. Also, reactants $S_1, \ldots, S_m$ do not diffuse in the intercellular space.

These conditions restrict the possibilities enabled by the RRM algorithm, but these conditions are necessary for a direct comparison and proof of equivalence between the RRM and the more limited PDM approaches. A model defined using the RRM algorithm without these restrictions may be impossible to implement in the PDM approach.

Under these confined conditions, we may still implement a multitude of biological mechanisms: the reactants may be secretions from an internal or external secretory gland, which are not released into the intercellular space but into the bloodstream or into a lumen. Or we may also assume homogenous mixing of the reactants in the cell space so that the products of the cellular or non-cellular chemical reactions reach all cells equally. This can be easily achieved in an experimental setup where we assume a layer of fluid over the cells on a 2D surface, homogenized by continuous mixing.

4

Cell Culture Simulation Algorithms

4.1 The aim of this chapter

In this chapter we present the prompt decision method (PDM) that applies the volume exclusion principle directly and incorporates the newly defined reactions. Then, we introduce our other novel approach, the reduced rate method (RRM), which applies volume exclusion in an indirect way. We also obtain an equivalent, but simplified version of the RRM algorithm, the marginal RRM (mRMM), that is expected to be easier to implement. After that, we intend to prove that in a given state of the system the *time until the next reaction to be realized* and also *the next reaction to be realized* are drawn from the same time and event distribution in both the PDM and the RRM algorithms, therefore, they are mathematically equivalent and are interchangeable. Finally, we experiment with some toy models to illustrate the potential in the newly incorporated contact dependent reactions.

4.1.1 Some important notations

In order to ease the further notations, for every reaction $\mu \in \{1, , \ldots, M\}$ we introduce

$$r_\mu = c_\mu h_\mu. \tag{4.1}$$

Where c_μ is the reaction rate of reaction μ. We have to adjust the definition of h_μ, as there may be reactions in the Cell Culture Simulation Algorithms in which chemicals $S_1, \ldots, S_m$ are not involved.

$$h_\mu = \begin{cases} 1 & \text{if chemical species are not involved in } \mu \\ \text{Nr. of distinct reactant combinations} & \text{if chemical species are involved in } \mu \end{cases} \tag{4.2}$$

See Table 2.1 and [27] for some reactions and the corresponding h_μ. Note that this notation slightly differs from Gillespie's original notation, where $c_\mu h_\mu$ would be the propensity function corresponding to reaction μ. In the cell culture simulation

algorithms most reactions depend on the state $\underline{\mathrm{C}}$ of the cells, this is what we are intended to emphasize with this distinguished notation. Basically, $a_\mu = r_\mu \cdot f(\underline{\mathrm{C}})$, where the definition of the function $f : \underline{\mathrm{C}} \to [0, \infty)$ is essentially different in the PDM and the RRM formulation.

Given that the two simulation methods differ not only in the algorithm but also in the definition of the reactions, we will denote reactions, events, and variables specific to the PDM with a hat ($\hat{\bullet}$) symbol. In case of the RRM, we do not use any specific notation. Some variables, events, and their probabilities depend only on the state of the system: e.g. the number of cells N_j with j free neighboring lattice sites, the reaction rates c_μ, reactant combinations h_μ etc. In a given state their values are the same in both methods, so we use the same notations for these in the algorithms.

4.2 The Prompt Decision Method (PDM)

The PDM is an extended version of the stochastic simulation algorithm introduced by Baker et al. [6]. It is a classical acceptance-rejection sampling method that samples the two-variable density function $P(\tau, \mu)$ given the state $\underline{\mathrm{X}}(t)$ of the system, that contains information of the number $N(t)$ and positions $C_i(t)$ of the cells, as well as the amount of the interacting chemical species $S_s(t)$. Thus, the PDM directly incorporates the principle of volume exclusion by rejecting reactions that are impossible due to the configuration of the cells on the lattice.

4.2.1 Propensities and probabilities

For the reactions $\hat{R}_\mu$ of types we discussed in Sec. 3.2 (also see Table 3.1) at time t, state $\underline{\mathrm{X}}(t)$, we define propensity functions $\hat{a}_\mu$ and their sum $\hat{a}$ the following way:

$$\hat{a}_\mu = r_\mu N, \qquad \hat{a} = \sum_{\mu=1}^{M} \hat{a}_\mu = N \sum_{\mu=1}^{M} r_\mu. \tag{4.3}$$

Notice that at first glance non-cellular chemical reactions do not fit in this formulation, since by definition, their propensities only depend on $\underline{\mathrm{S}}(t)$ and do not depend on the state of the cells. To overcome this we may think of the number of cells as a scaling factor $1/N$ to the reaction rate constant. Then, for a non-cellular chemical reaction μ we have

$$\hat{a}_\mu = c_\mu h_\mu = \frac{c_\mu h_\mu}{N} N = r_\mu N.$$

With this notation, non-cellular chemical reactions reduce to a special case of the spontaneous type cellular biochemical reactions. Let $\hat{P}_1(\hat{\tau})$ denote the probability density function of the sojourn time until the next *possible* event and $\hat{P}_2(\mu|\hat{\tau})$ denote

the probability mass function of the next possible event:

$$\hat{P}_1(\hat{\tau}) = \hat{a}e^{-\hat{a}\hat{\tau}}, \quad \hat{P}_2(\mu|\hat{\tau}) = \hat{a}_\mu/\hat{a}, \;\; \mu \in \{1, \ldots, M\}. \tag{4.4}$$

4.2.2 Assumptions in the PDM algorithm

The algorithm is derived from the following assumptions:

- **Selecting the next possible reaction:** from $\hat{P}_2(\mu|\hat{\tau})$ (Eq. 4.4) we draw reaction index μ. We will use the notation $\hat{R}_\mu$ for a selected reaction in the PDM algorithm. It is easy to see that the probability of selecting $\hat{R}_\mu$ does not depend on the actual number of cells in this approach (cf. Eq. 4.3 and 4.4).
- **Target cell selection:** a target cell is selected from the population with a discrete uniform distribution, without taking the position of the cells into account. Let $\hat{E}_j$ denote the random event that we select a target cell that has j free neighbors.
- **The execution of the reaction depends on the neighborhood of the target cell**, the spontaneous reactions are always executed, contact-inhibited or contact-promoted events may be discarded depending on the neighborhood of the target cell. A more detailed explanation of this acceptance-rejection step is about to come later. During one iteration we make trials to decide if the selected possible reaction can be the next reaction to be realized.
- **Selecting the waiting time:** as we may discard the selected *possible reactions* several times during one iteration, we introduce $l \in \{1, 2 \ldots, \}$ to index the trials during one iteration. Thus, in a given iteration for the l-th trial we generate sojourn time $\hat{\tau}_l \sim Exp(\hat{a})$ (Eq. 4.4) corresponding to the next *possible* event.
  For the sake of clarity (and the convenience of the proof), we define the variable $\hat{T}$, into which we sum the sojourn times generated for the *possible events* during one iteration. Accordingly, the time elapsed between two consecutive *realized events* is
  $$\hat{T} = \sum_{l=1}^{k} \hat{\tau}_l, \tag{4.5}$$
  given that we were able to execute the reaction in the k-th attempt (i. e., $k-1$ possible events were discarded in this case). Naturally k is a discrete random variable that is geometrically distributed, as we shall see later. In practice, book the intervent time $\hat{\tau}_l$ right after generating it: $\hat{T} \leftarrow \hat{T} + \hat{\tau}_l$.
- **During the decision step** at the l-th attempt
  - In case of contact-inhibited reactions, randomly select a neighboring site of the target cell from its neighbors with discrete uniform distribution. If the site is *occupied*, or it is outside the border, reject the reaction. If the site is

empty execute the reaction and increase the system time: $t \leftarrow t + \hat{T}$ and set $\hat{T} = 0$.

- In case of spontaneous reactions execute the reaction on the selected cell and increase the system time: $t \leftarrow t + \hat{T}$ and set $\hat{T} = 0$.
- In case of contact-promoted reactions, randomly select a neighboring site of the target cell from its neighbors with discrete uniform distribution. If the site is *occupied*, or it is outside the border, execute the reaction, increase the system time: $t \leftarrow t + \hat{T}$ and set $\hat{T} = 0$. If the site is *empty*, reject the reaction.

The steps of the algorithm do not follow the order of the assumptions, we chose this order to streamline the explanation. We solely introduced the variable $\hat{T}$ to make the proof more convenient. In practice, it is not necessary to occupy memory space for $\hat{T}$ and waste machine time by updating it. One may simply increase the system time right after generating the waiting time.

4.2.3 Data structure

From a practical point of view, the implementation of $\underline{\mathrm{X}}(t)$ is a list or an array. However, we need a data structure that allows us to easily determine whether a lattice site (x_*, y_*) is empty or occupied ($x_* \in \{1, \ldots, n_1\}, y_* \in \{1, \ldots, n_2\}$). Searching in arrays or lists typically has $O(N)$ time complexity, thus it is more convenient to use the underlying lattice to book-keep the occupied lattice sites.

Hence, we define the matrix $L \in \{0, 1\}^{n_1 \times n_2}$ for $n_1, n_2 < \infty$, where 0 and 1 stand for an empty and an occupied site, respectively. Checking the value of an element $L[x_*, y_*]$ usually has a constant time complexity, $O(1)$. Using $\underline{\mathrm{X}}(t)$ and L simultaneously has two advantages. First, randomly selecting the target cell from the list $\underline{\mathrm{X}}(t)$ is of $O(1)$, but it can be very time consuming to randomly select a target cell from L. Second, with using L we may obtain the occupancy of a lattice site in constant time. These assumptions hold for most programming languages, e. g., for Python that we shall use to implement our toy models. With the introduced notations, the Prompt Decision Method is as follows.

Algorithm Prompt Decision Method

1. **Initialization:**

   (i) set $t \leftarrow 0$, prescribe halting conditions H,

   (ii) initialize the lattice L with size $n_1 \times n_2$,
   place $N(0) \leq n_1 \cdot n_2$ cells on the lattice according to essay,
   store location of cells and amount of chemical species in $\underline{\mathrm{X}}(t)$

2. **Calculate propensity functions** $\hat{a}_\mu$ for all $\mu \in \{1, \ldots, M\}$ according to Eq. (4.3) and **set** $\hat{T} = 0$.
3. **Generate intervent time:** choose $\hat{\tau} \sim Exp(\hat{a})$ (Eq. 4.4) and **set** $\hat{T} \leftarrow \hat{T} + \hat{\tau}$.
4. **Reaction selection:** choose μ according to $\hat{P}_2(\mu|\hat{\tau})$ (Eq. 4.4).
5. **Cell selection:** draw target cell $X_k(t)$ from the population $\underline{\mathrm{X}}(t)$ with discrete uniform distribution.
6. **Decision step:**
   - In case of a non-cellular chemical reaction execute the reaction on the concerned chemical species.
     Update data structure. Go to Step 7.
   - In case of a spontaneous reaction, execute the reaction on target cell and update data structures.
     Go to Step 7.
   - In case of a contact-inhibited reaction, randomly select a neighboring site of the target cell from its adjacent sites, with discrete uniform distribution.
     - If the target site is *occupied*, or it is outside the border, reject the reaction. Go to Step 3.
     - If the site is *empty* execute the reaction and update data structures. Go to Step 7.
   - In case of a contact-promoted reaction, randomly select a neighboring target site of the target cell from its adjacent sites, with discrete uniform distribution.
     - If the site is *empty*, reject the reaction. Go to Step 3.
     - If the site is *occupied*, or it is outside the border, execute the reaction, update data structures. Go to Step 7.
7. **Update time:** set $t \leftarrow t + \hat{T}$.
8. **Halt if** $H = True$ **else continue the process with Step 2.**

4.3 The Reduced Rate Method (RRM)

The essence of the method is the classification of the cells according to the number of their free neighbors on the lattice. From a theoretical viewpoint, this approach enables us to define a state space in which the system is 'well-stirred' in the sense that the probability of a reaction to happen in the next time interval δt does not depend explicitly on the positions of the reactants (cells), it only depends on the number of their free adjacent sites. Thus, with the proper scaling of the reaction rates and some appropriate assumptions, we reduce the problem to Gillespie's SSA formalism.

4.3.1 Assumptions in the RRM algorithm

According to the restrictions we made, the products of the cellular and non-cellular (bio)chemical reactions neither change the state of the cells nor diffuse: the cells only differ in their positions on the lattice. Based on this assumption, we may classify the cells at state $\underline{\mathrm{X}}(t)$, according to the number of their free adjacent sites j. We refer to cells that have j free adjacent sites as *class-j cells*. At any state $\underline{\mathrm{X}}(t)$, the value of $N_j = N_j(t)$ for $\forall j \in \{0, \ldots, \eta\}$ and the positions of the cells in class-j are known, thus, we can define the propensities and choose the target cells based on this information.

We assume that **contact dependent reactions occur between cells and either their j free or $(\eta - j)$ occupied neighbors**. Thus, in the RRM algorithm selecting a reaction defines two things: i) the reaction index μ, that defines the change in the state $\underline{\mathrm{X}}$ and ii) the class-j of the target cell. We define reaction $R_{\mu,j}$, that is reaction with index μ acting on a cell having j free neighbors (and possibly causes a $(\Delta S_1, \ldots, \Delta S_m)$ change in the number of the chemical species).

4.3.2 Propensities

For all $\mu \in \{1, \ldots, M\}$ and all $j \in \{0, \ldots, \eta\}$ we define propensity functions $a_{\mu,j}$ according to the following considerations.

Contact-inhibited reactions occur when a cell 'reacts' with one of its empty neighbors. Thus, in this case the reactants are the cells and their empty adjacent sites on the lattice. The number of possible combinations between the class-j cells and their free adjacent sites is $j \cdot N_j$. The reaction rate of an elementary reaction, with one empty neighbor, is r_μ/η. Thus, the propensity function of the contact-inhibited reaction μ acting on a cell with j free adjacent sites is:

$$a_{\mu,j} = \frac{r_\mu}{\eta} j N_j, \quad \forall j \in \{0, \ldots, \eta\}. \tag{4.6a}$$

The propensities of **spontaneous reactions** does not depend on the number of free neighbors of the cell. Nevertheless, we use the two-index notation to simplify the proof later.

$$a_{\mu,j} = r_\mu N_j, \quad \forall j \in \{0, \ldots, \eta\}. \tag{4.6b}$$

For non-cellular chemical reactions we use the same scaling we used in the PDM algorithm:

$$a_{\mu,j} = \frac{c_\mu h_\mu}{N} N_j = r_\mu N_j \quad \forall j \in \{0, \ldots, \eta\}. \tag{4.6c}$$

Contact-promoted reactions occur when a cell 'reacts' with one of its occupied neighbors. The number of possible combinations between the class-j cells and their occupied $\eta - j$ neighbors is: $(\eta - j) \cdot N_j$. The rate of such an elementary reaction is

r_μ/η. The propensity function of the contact-promoted reaction μ acting on a class-j cell is:

$$a_{\mu,j} = \frac{r_\mu}{\eta}(\eta - j)N_j(t), \quad \forall j \in \{0, \ldots, \eta\}. \tag{4.6d}$$

We assume that reaction indices $\mu \in \{1, \ldots, \alpha\}$ belong to the contact-inhibited, $\mu \in \{\alpha+1, \ldots, \alpha+\beta\}$ belong to the spontaneous and $\mu \in \{\alpha+\beta+1, \ldots, M\}$ belong to the contact dependent reactions. The sum of the propensities is:

$$a = \sum_{\mu=1}^{M}\sum_{j=0}^{\eta} a_{\mu,j} = \frac{1}{\eta}\sum_{\mu=1}^{\alpha} r_\mu \sum_{j=0}^{\eta} jN_j + \sum_{\mu=\alpha+1}^{\alpha+\beta} r_\mu N + \frac{1}{\eta}\sum_{\mu=\alpha+\beta+1}^{M} r_\mu \sum_{j=0}^{\eta}(\eta - j)N_j \tag{4.7}$$

Unlike in the PDM, in case of the RRM a may be zero in some models even with $N > 0$. We have to ensure that this does not cause problems while drawing the waiting time.

With these notations $a_{\mu,j}\delta t$ is the probability, to first order in δt, that a reaction with index μ will occur to a class-j cell on the lattice and possibly cause a $(\Delta S_1, \ldots, \Delta S_m)$ change in the number of the chemical species in the next time interval δt. The propensity functions explicitly specify the corresponding (μ, j) pairs, thus, we obtain the joint probability of executing reaction μ on a cell of class-j. We denote the corresponding event $R_{\mu,j}$.

4.3.3 Obtaining the waiting time distribution and the probabilities of the reactions from first principles

Now, we obtain the formula for the joint probability $P(\tau, R_{\mu,j})$ with the direct application of Gillespie's reasoning [27]. The probability that $R_{\mu,j}$ occurs in the differential time interval $[t+\tau, t+\tau+d\tau]$ is

$$P(\tau, R_{\mu,j})d\tau = P_0(\tau) \cdot a_{\mu,j}d\tau, \tag{4.8}$$

where $P_0(\tau)$ is the probability at time t that none of the reactions occur in $(t, t+\tau)$.

We divide the interval $(t, t+\tau)$ to κ subintervals with equal length of $\epsilon = \tau/\kappa$. Assume the Markov property holds for the system, then the probability that none of the $R_{\mu,j}$ reactions occur in any of the subintervals is $P_0(\tau) = [1 - \sum_{\nu,l} a_{\nu,l} \cdot \epsilon + o(\epsilon)]^\kappa = [1 - \sum_{\nu,l} a_{\nu,l} \cdot \tau/\kappa + o(\kappa^{-1})]^\kappa$. Taking the limit $\kappa \to \infty$ we obtain

$$P_0(\tau) = \lim_{k\to\infty}\left[1 - \sum_{\nu=1}^{M}\sum_{l=0}^{\eta} a_{\nu,l} \cdot \tau/\kappa + o(\kappa^{-1})\right]^\kappa = \exp\left(-\sum_{\nu=1}^{M}\sum_{l=0}^{\eta} a_{\nu,l}\tau\right) = e^{-a\tau}. \tag{4.9}$$

Substituting Eq. (4.9) into Eq. (4.8) yields

$$P(\tau, R_{\mu,j})d\tau = e^{-a\tau} \cdot a_{\mu,j}d\tau. \tag{4.10}$$

As the reactions are mutually exclusive we may obtain the marginal distribution $P_1(\tau)$ for the waiting time by summing Eq. (4.10) over all (μ, j) pairs:

$$P_1(\tau)d\tau = e^{-a\tau} \sum_{\mu=1}^{M} \sum_{j=0}^{\eta} a_{\mu,j}d\tau = a \cdot e^{-a\tau}d\tau. \tag{4.11}$$

Finally, by conditioning $P(\tau, R_{\mu,j})$ on τ, we obtain

$$P(\tau, R_{\mu,j})d\tau = P_1(\tau) \cdot P_2(R_{\mu,j}|\tau)d\tau. \tag{4.12}$$

Substituting Eq. (4.10) and (4.11) into Eq. (4.12) and solving for P_2 we obtain the joint probability that the next reaction would be of index μ and would be executed on a class-j cell given waiting time τ:

$$P_2\left(R_{\mu,j}|\tau\right) = \frac{a_{\mu,j}}{a} =: P(R_{\mu,j}) \quad \mu \in \{1, \ldots, M\}, \quad j \in \{0, \ldots, \eta\}. \tag{4.13}$$

Hence, following Gillespie's argument, we have derived the probability density function of the waiting time and the probabilities of the reactions in the Reduced Rate Method Cell Culture Simulation Algorithm from first principles.

4.3.4 Data structure

Note that in case of a movement, proliferation, or a cell death event, the class of several cells can change at once. Thus, we need to reach the cells as quick as possible to adjust the corresponding changes. A convenient data structure for this (in Python) consists of five dictionaries, that we will store in one list, called $\underline{\mathrm{D}}$. Dictionary $\underline{\mathrm{D}}[j]$ contains the cells that have j free adjacent sites in the following way: the indices of cells are the keys and the corresponding coordinates are the values. Naturally, the length of $\underline{\mathrm{D}}[j]$ is N_j. We also store the amount of chemical substances in a dedicated list $\underline{\mathrm{B}}$.

We keep the underlying lattice L to check if a particular lattice site is occupied or empty. A convenient data structure is a matrix $L \in \mathbb{N}^{n_1 \times n_2}$, where entry $L[x_*, y_*]$ is the index of the cell in the population. Note that every newborn cell has to be given a unique index to avoid complications. Thus, in some models the value of some indices may be greater than the capacity K of the population. With these notations and assumptions, the Reduced Rate Method is as follows.

Algorithm Reduced Rate Method

1. **Initialization:**
   (i) set $t \leftarrow 0$, prescribe halting conditions H,
   (ii) initialize the lattice L with size $n_1 \times n_2$, initialize $\underline{\mathrm{D}}$,
   place $N(0) \leq n_1 \cdot n_2$ cells on the lattice according to essay,
   store location of class-j cells in dictionary $\underline{\mathrm{D}}[j]$,
   (iii) initialize $\underline{\mathrm{B}}$, store initial amount of chemical species in list $\underline{\mathrm{B}}$.
2. **Calculate propensity functions** $a_{\mu,j}$ and a for all $\mu \in \{1, \ldots, M\}$ and $j \in \{0, \ldots, 4\}$ according to Eqs. (4.6) and (4.7). If $a = 0$ halt the algorithm.
3. **Generate intervent time:** choose τ according to $P_1(\tau)$ from Eq. (4.11).
4. **Select event and cell type:** choose (μ, j) according to $P(R_{\mu,j})$ from Eq. (4.13).
   Select target cell from class-j with discrete uniform distribution.
   In case of a movement, or cell division, select one of the j free adjacent sites of the target cell with discrete uniform distribution.
5. **Execute and update:**
   (i) Update time: set $t \leftarrow t + \tau$.
   (ii) Execute the event. Update $\underline{\mathrm{D}}, \underline{\mathrm{B}}$ and L.
6. **Halt if** $H = True$ **else continue the process with Step (2).**

It is clear that this formulation of the RRM is of direct method type. We will refer to this implementation as RRM in the thesis. However, we would like to point out that Theorem 2.1 guarantees that we may easily apply Gillespie's first reaction method formalism to the RRM.

Algorithm Reduced Rate Method (First Reaction Method formalism)

1. **Initialization:**
   (i) set $t \leftarrow 0$, prescribe halting conditions H,
   (ii) initialize the lattice L with size $n_1 \times n_2$, initialize $\underline{\mathrm{D}}$,
   place $N(0) \leq n_1 \cdot n_2$ cells on the lattice according to essay,
   store location of class-j cells in dictionary $\underline{\mathrm{D}}[j]$,
   (iii) initialize $\underline{\mathrm{B}}$, store initial amount of chemical species in list $\underline{\mathrm{B}}$.
2. **Calculate propensity functions** $a_{\nu,l}$ for all $\nu \in \{1, 2, \ldots, M\}$ and $l \in \{0, \ldots, \eta\}$.
3. **Calculate putative times** $\tau_{\nu,l} \sim Exp(a_{\nu,l})$ for all $\nu \in \{1, 2, \ldots, M\}$ and $l \in \{0, \ldots, \eta\}$,
4. **Reaction selection:** choose $R_{\mu,j}$ for which $\tau_{\mu,j} = \min\{\tau_{\nu,l} : \nu \in \{1, 2, \ldots, M\}$ and $l \in \{0, \ldots, \eta\}\}$,
   - Select target cell from class-j with discrete uniform distribution.
   - In case of a movement, or cell division, select one of the j free adjacent sites of the target cell with discrete uniform distribution
5. **Update time and state variables:**
   (i) set $t \leftarrow t + \tau_{\mu,j}$

(ii) Execute the event. Update $\underline{\mathrm{D}}, \underline{\mathrm{B}}$ and L.

6. **Halt if $H = True$ else continue the process with Step (2).**

4.3.5 The well stirred case

We may reach the well stirred system of the cell population by extending the neighborhood of the cells to the whole volume of the container minus the volume of the cell itself. Thus, we would have only one adjacency class that contains all cells $N(t)$ of the population. The number of free sites would then be $K - N$ and the number η of adjacent sites would be $K - 1$, which we one may approximate with $K - 1 \approx K$ for large K. Thus, we may obtain the following propensities for this 'well stirred' system:

$$a_\mu = \begin{cases} \frac{r_\mu}{K-1}(K-N)N(t) & \mu \in \{1, \dots, \alpha\} \\ r_\mu N(t) & \mu \in \{\alpha+1, \dots, \alpha+\beta\} \\ \frac{r_\mu}{K-1}(N-1)N(t) & \mu \in \{\alpha+\beta+1, \dots, M\} \end{cases} \tag{4.14}$$

Cell division is contact inhibited by our earlier assumptions, cell death and also biochemical reactions may be modeled by any type of reaction, see Chapter 3.

A remarkable consequence of this is the fact that $a_{\text{division}} = \frac{r_\mu}{K-1}(K-N)N(t)$ is itself able to model a population that grows in a logistic fashion. Classical stochastic logistic models and variants are assumed to be birth and death processes [3]. Unlike those, the logistic model we obtained here does not involve cell death. Since cell division is supposed to be a reaction between cells and free adjacent sites, the effect the shrinking 'free' space takes on the cells is sufficient to produce logistic behavior. Thus, we gained a new interpretation of the logistic growth model, which, to the best of our knowledge has not been considered so far. In Chapter 5 we will improve this idea to obtain a method that efficiently simulates the dynamics of populations with realistic cell cycle length.

4.4 The marginal Reduced Rate Method (mRRM)

Next, we describe a variant of the RRM in which the reaction index μ and the cell type j are selected in two distinct steps. With this approach, we may reduce the number of propensity functions which may make the implementation more convenient and may reduce the simulation runtime. In the mRRM algorithm we use the same data structure we introduced in case of the RRM algorithm. Assume that reactions with index $\mu \in \{1, \dots, \alpha\}$ are contact-inhibited, $\mu \in \{\alpha+1, \dots, \alpha+\beta\}$ are spontaneous and $\mu \in \{\alpha+\beta+1, \dots, M\}$ are contact dependent reactions.

4.4.1 Propensities and probabilities

Consider the propensity functions $a_{\mu,j}$ of the RRM (Eq. 4.6) and the joint probabilities $P(R_{\mu,j}) = a_{\mu,j}/a$ obtained from them. First, for every reaction $\mu \in \{1,\dots,M\}$, we sum over j obtaining the marginal probabilities $P(R_\mu) = a_\mu/a$ for reaction R_μ. The corresponding propensities $a_\mu = \sum_{j=0}^{\eta} a_{\mu,j}$ are of form:

$$a_\mu = \begin{cases} \frac{r_\mu}{\eta}\sum_{j=0}^{\eta} jN_j(t) & \mu \in \{1,\dots,\alpha\} \\ r_\mu N(t) & \mu \in \{\alpha+1,\dots,\alpha+\beta\} \\ \frac{r_\mu}{\eta}\sum_{j=0}^{\eta}(\eta-j)N_j(t) & \mu \in \{\alpha+\beta+1,\dots,M\} \end{cases} \tag{4.15}$$

Naturally, the sum of the propensity functions are the same in the RRM and mRRM algorithms:

$$a = \sum_{\mu=1}^{M}\sum_{j=0}^{\eta} a_{\mu,j} = \sum_{\mu=1}^{M} a_\mu. \tag{4.16}$$

In case $a > 0$, the probability density function $P_1(\tau)$ of the waiting time τ until the next reaction, and the probability $P_2(\mu|\tau)$ of the next reaction R_μ are:

$$P_1(\tau) = ae^{-a\tau}, \quad P_2(\mu|\tau) = \frac{a_\mu}{a} =: P(R_\mu), \quad \mu \in \{1,\dots,M\}. \tag{4.17}$$

4.4.2 Target cell selection

Finally, we have to obtain a method for selecting the target cell in the mRRM algorithm. It would be tempting to simply choose the class-j of the target cell according to N_j/N, in case of all reaction types. But this would lead to a wrong result in general. Suppose, we chose cell movement to execute and suppose $N_0 > 0$. There would be a positive probability to choose class-0 from which to select the target cell, and thus select a cell with no free neighbors, even though such a cell is not able to move.

In the original RRM algorithm the probability of executing reaction μ on a type j cell is $P(R_{\mu,j})$. We saw that in the mRRM algorithm, selecting reaction R_μ and selecting the type j of the target cell is executed in two distinct steps, but these steps are not independent from each other. In the mRRM algorithm, let A_j denote the event that the target cell is of class j. For every reaction R_μ, we look for the probability $P(A_j|R_\mu)$, with which $P(A_j|R_\mu)\cdot P(R_\mu) = P(R_{\mu,j})$ in the original RRM algorithm. That is:

$$P(A_j|R_\mu) = \begin{cases} \frac{jN_j}{\sum_{j=0}^{\eta} jN_j} & \mu \in \{1,\dots,\alpha\} \\ N_j/N & \mu \in \{\alpha+1,\dots,\alpha+\beta\} \\ \frac{(\eta-j)N_j}{\sum_{j=0}^{\eta}(\eta-j)N_j} & \mu \in \{\alpha+\beta+1,\dots,M\} \end{cases} \tag{4.18}$$

Notice that with Eq. (4.18) it is not possible to select class-0 in case of a contact-

inhibited reaction. Likewise, it is not possible to select class-4 in case of a contact-promoted reaction. On Fig. 4.1 we illustrate Eq. (4.18) in a concrete example, the case of the snapshot of the simulation introduced on Fig.3.2 (Sec.3.1). After selecting the class j of the target cell according to Eq. (4.18), the algorithm chooses a target cell from the class j cells with discrete uniform distribution.

It easy to see that the RRM and mRRM algorithms are mathematically equivalent formulations to obtain realizations of a given stochastic process concerning a cell culture.

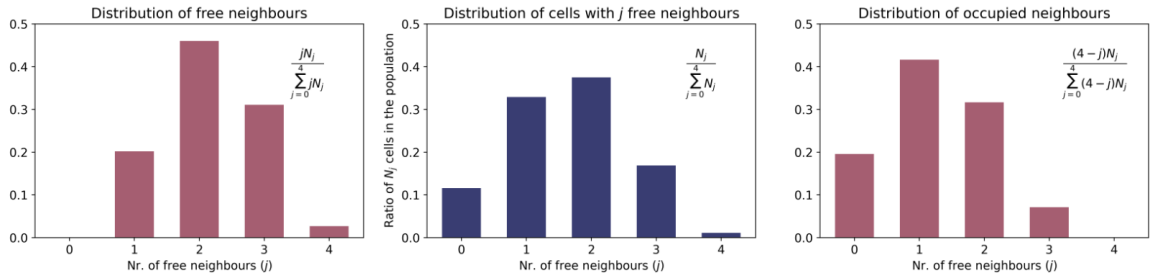

Figure 4.1: **Illustration of the distributions in Eq. (4.18)** The figures correspond to the snapshot of Fig.3.2. Suppose we had all three reaction types in the simulation at state $\underline{X}(t_*)$ of the system. In case the next reaction would be a *spontaneous reaction*, we should choose the class j of the target cell according to the distribution on the middle figure, and then choose a particular target cell from class j at random. This is equivalent of choosing a cell randomly from the total population. It is clear from the figure that we would choose the cell class $j = 2$ cell with the highest probability, as the ratio of cells with two free adjacent sites is the highest in the population. Since there are cells in all five classes, we may choose any class with nonzero probability in a spontaneous reaction. If the next reaction would be a *contact-inhibited reaction*, we should choose the class of the target cell according to the distribution on the left figure. Since $0 \cdot N_0 = 0$, we choose class $j = 0$ with zero probability. Thus cells with zero free neighbors do not participate in contact-inhibited reactions. At this state of the system we would choose a class $j = 2$ cell with the highest probability. If the next reaction would be a *contact-promoted reaction*, we should choose the target cell type from the distribution on the right. Since $(4-4) \cdot N_4 = 0$, we choose class $j = 4$ with zero probability – cells with 4 free adjacent sites do not participate in contact-promoted reactions. We would choose class $j = 1$ with the highest probability at this state.

Algorithm marginal Reduced Rate Method

1. **Initialization:**
   (i) set $t \leftarrow 0$, prescribe halting conditions H,
   (ii) initialize the lattice L with size $n_1 \times n_2$, initialize $\underline{D}$,
   place $N(0) \leq n_1 \cdot n_2$ cells on the lattice according to essay,
   store location of class-j cells in dictionary $\underline{D}[j]$,
   (iii) initialize $\underline{B}$, store initial amount of chemical species in list $\underline{B}$.
2. **Calculate propensity functions** a_μ and a for all $\mu \in \{1, \ldots, M\}$ according to Eq. (4.15) and (4.16).
   If $a = 0$ halt the algorithm.
3. **Generate intervent time:** choose τ according to $P_1(\tau)$ from Eq. (4.17).
4. **Select event:** choose μ according to $P_2(\mu|\tau)$ from Eq. (4.17).
5. **Select cell:**
   - Choose the class j of cells according to $P(A_j|R_\mu)$ from Eq. (4.18).
   - Choose a target cell randomly with uniform distribution from class j.

- If R_μ correspond to movement or division, then choose one of the free neighbors of the target cell with discrete uniform distribution.

6. **Execute and update:**
   (i) Update time: set $t \leftarrow t + \tau$.
   (ii) Execute the event.
   (iii) Update $\underline{D}, \underline{B}$ and L.
7. **Halt if $H = True$ else continue the process with Step (2).**

4.5 Equivalence of the algorithms

Before stating our main theorem, we would like to emphasize the differences between the PDM and the RRM algorithms. The first important difference is in the definition of the propensity functions: in case of the PDM, the propensity functions always depend on the total number of cells (Eq. 4.3). In contrast, in the RRM we classify the cells according to the number of their free neighbors. Only 'spontaneous' reactions and the corresponding propensities depend on the total number of cells (Eq. 4.6b) other reactions occur between cells and their empty neighbors (Eq. 4.6a) or between cells and their occupied neighbors (Eq. 4.6d).

The second important difference is that the drawn event does not always occur in the PDA approach, whereas the drawn event is always executed in the RRA approach. The third important difference is that drawing reaction $\hat{R}_\mu$ and randomly selecting a target cell (that has j free neighbors, event $\hat{E}_j$) are independent in the case of PDM. In the RRM, $R_{\mu,j}$ uniquely determines that the reaction with index μ has to be executed on a cell having j free neighbors and maybe alters the state of the chemical species as well.

During the proof we assume that there are M possible reactions in the PDM and RRM and the order of reaction indices are the same in both algorithms. Suppose that reactions with index $\mu \in \{1, \ldots, \alpha\}$ are contact-inhibited, $\mu \in \{\alpha + 1, \ldots, \alpha + \beta\}$ are spontaneous and $\mu \in \{\alpha + \beta + 1, \ldots, M\}$ are contact dependent reactions.

Theorem 4.1 (On the equivalence of the PDM and RRM algorithms)**.** *Using the notations introduced so far, consider the state $\underline{X}(t)$ of the system at time t. Suppose that, in this state both $\hat{a}, a > 0$ in PDM and RRM, respectively. Suppose we start both algorithms from $\underline{X}(t)$ (i. e., $\hat{T} = 0$ in case of the PDM).*

In the PDM algorithm let $\hat{R}_\mu$ denote the drawn reaction in the k-th attempt during the 'Reaction selection step' and let $\hat{E}_j$ denote the event that we choose a target cell with j free neighbors in the corresponding 'Cell selection step' and let $\hat{S}$ denote the random event that we would be able to execute $\hat{R}_\mu$ in the 'Decision step'.

In the RRM algorithm let $R_{\mu,j}$ denote the next reaction. Let $\hat{T}$ and τ denote the

soujurn times until the next executed reaction in the PDM and the RRM, respectively. Then

(a) $\hat{P}((\hat{R}_\mu \cap \hat{E}_j)|\hat{S}) = P(R_{\mu,j})$*, and*

(b) $\hat{T} \sim Exp(a)$ *and* $\tau \sim Exp(a)$ *with parameter* a *defined in (4.7).*

Proof. We will use a direct proof. Let us start with the probability of the events.

(a) In the RRM algorithm $P(R_{\mu,j}) = a_{\mu,j}/a$ (Eq. 4.13). We will show that in the PDM $\hat{P}((\hat{R}_\mu \cap \hat{E}_j)|S) = a_{\mu,j}/a$ with the definition of $a_{\mu,j}$ (Eq. 4.6) and a (Eq. 4.7).

In the PDM algorithm let $\hat{P}(\hat{S})$ denote the probability of the random event $\hat{S}$, that is, we would execute the selected $\hat{R}_\mu$ in the 'Decision step'. With Bayes theorem we obtain:

$$\hat{P}(\hat{R}_\mu \cap \hat{E}_j|\hat{S}) = \frac{\hat{P}(\hat{S}|\hat{R}_\mu \cap \hat{E}_j) \cdot \hat{P}(\hat{R}_\mu \cap \hat{E}_j)}{\hat{P}(\hat{S})}. \tag{4.19}$$

During the 'Cell selection step' in the PDA algorithm, $\hat{E}_j$ is the event when we choose a target cell that has j free neighbors. The probability of this is the proportion of cells with j neighbors in the population: $\hat{P}(\hat{E}_j) = N_j/N$. Now, because $\hat{R}_\mu$ and $\hat{E}_j$ are independent events:

$$\hat{P}(\hat{R}_\mu \cap \hat{E}_j) = \hat{P}(\hat{R}_\mu) \cdot \hat{P}(\hat{E}_j) = \hat{a}_\mu/\hat{a} \cdot N_j/N. \tag{4.20}$$

In the PDM, during the Decision step spontaneous reactions are executed with a probability of 1. Contact-inhibited reactions are executed if we choose a target cell that has at least one free neighbor and we choose one of its free neighbors. Thus, in the PDM algorithm, we introduce the random event $\hat{F}$, which occurs when a contact-inhibited reaction is executed. Contact-promoted reactions are executed if we choose a target cell that has occupied neighbor and we choose one of them. We shall denote the corresponding event $\hat{\overline{F}}$. Since, $\hat{R}_\mu$ and $\hat{F}$ are independent, we may obtain:

$$\hat{P}(\hat{S}) = \frac{\sum_{\mu=1}^{\alpha} \hat{a}_\mu}{\hat{a}} \hat{P}(\hat{F}) + \frac{\sum_{\mu=\alpha+1}^{\alpha+\beta} \hat{a}_\mu}{\hat{a}} + \frac{\sum_{\mu=\alpha+\beta+1}^{M} \hat{a}_\mu}{\hat{a}} \hat{P}(\hat{\overline{F}}). \tag{4.21}$$

Now, we need the probabilities $\hat{P}(\hat{F})$ and $\hat{P}(\hat{\overline{F}})$. The probability that a randomly chosen neighbor of a cell is free, given that the cell has $j \in \{0, \ldots, \eta\}$ free neighbors is $\hat{P}(\hat{F}|\hat{E}_j) = j/\eta$. Since $\hat{E}_0, \ldots, \hat{E}_\eta$ are pairwise disjoint events whose union is the entire sample space (i. e., we randomly choose a cell from the population that has 0 to η free neighbors), we may obtain $\hat{P}(\hat{F})$ by the law of total probability:

$$\hat{P}(\hat{F}) = \sum_{j=0}^{\eta} \left(\hat{P}(\hat{F}|\hat{E}_j) \cdot \hat{P}(\hat{E}_j) \right) = \frac{1}{\eta N} \sum_{j=0}^{\eta} j N_j. \tag{4.22}$$

Since $\hat{\overline{F}}$ is the complement event of $\hat{F}$: $\hat{P}(\hat{\overline{F}}) = 1 - \hat{P}(\hat{F}) = \frac{1}{\eta N} \sum_{j=0}^{\eta} (\eta - j) N_j$. Thus,

we have a formula for every term in Eq. (4.21).

Finally, we have to evaluate $\hat{P}(\hat{S}|\hat{R}_\mu \cap \hat{E}_j)$ that is the probability that the currently selected reaction can be executed given that it is a reaction $\hat{R}_\mu$ and the target cell has j free neighbors. Given the reaction and the target cell, we have to choose one neighbor of the target cell at random with discrete uniform distribution. In case of a CIR we can execute $\hat{R}_\mu$ if the site is empty, with probability j/η. Similarly, a CPR can be executed if the chosen site is occupied, it occurs with probability $(\eta - j)/\eta$, and we can always execute a spontaneous reaction. For all pairs (μ, j), where $j \in \{0, \ldots, \eta\}$:

$$\hat{P}(\hat{S}|\hat{R}_\mu \cap \hat{E}_j) = \begin{cases} j/\eta & \mu \in \{1, \ldots, \alpha\} \\ 1 & \mu \in \{\alpha+1, \ldots, \alpha+\beta\} \\ (\eta - j)/\eta & \mu \in \{\alpha+\beta+1, \ldots, M\} \end{cases} \tag{4.23}$$

Now, from Eqs. (4.3), (4.7), (4.21) and from the definition of $\hat{a}$ we have:

$$\hat{a}\hat{P}(\hat{S}) = \frac{1}{4}\sum_{\nu=1}^{\alpha} r_\nu \sum_{j=0}^{4} jN_j + \sum_{\nu=\alpha+1}^{\alpha+\beta} r_\nu N + \frac{1}{4}\sum_{\nu=\alpha+\beta+1}^{M} r_\nu \sum_{j=0}^{4}(4-j)N_j = a. \tag{4.24}$$

Thus, in the case of the contact-inhibited reaction $\mu \in \{1, \ldots, \alpha\}$, for the probability (4.19), with using (4.20), (4.21), (4.22), (4.23), and (4.24) we obtain:

$$\hat{P}(\hat{R}_\mu \cap \hat{E}_j|\hat{S}) = \frac{\frac{j}{4}\frac{N_j}{N} \cdot \frac{\hat{a}_\mu}{\hat{a}}}{\hat{P}(\hat{S})} = \frac{\frac{1}{4N}(jN_j) \cdot r_\mu N}{a} = \frac{\frac{r_\mu}{4} jN_j}{a} = \frac{a_{\mu,j}}{a} = P(R_{\mu,j}) \tag{4.25}$$

For the spontaneous reactions $\mu \in \{\alpha+1, \ldots, \alpha+\beta\}$:

$$\hat{P}(\hat{R}_\mu \cap \hat{E}_j|\hat{S}) = \frac{1 \cdot \frac{N_j}{N} \cdot \frac{\hat{a}_\mu}{\hat{a}}}{\hat{P}(\hat{S})} = \cdots = \frac{a_{\mu,j}}{a} = P(R_{\mu,j}) \tag{4.26}$$

For the contact-promoted reactions $\mu \in \{\alpha+\beta+1, \ldots, M\}$:

$$\hat{P}(\hat{R}_\mu \cap \hat{E}_j|\hat{S}) = \frac{\frac{4-j}{4} \cdot \frac{N_j}{N} \cdot \frac{\hat{a}_\mu}{\hat{a}}}{\hat{P}(\hat{S})} = \cdots = \frac{a_{\mu,j}}{a} = P(R_{\mu,j}) \tag{4.27}$$

This proves part *(a)* of the theorem.

(b) We now show that in PDM the sojourn time $\hat{T}$ until the next successful event has the same probability density function as the sojourn time τ in RRM. From the previous part of the proof we know that the reaction in PDM is only successful with probability $\hat{P}(\hat{S})$ (Eq.4.21), thus we discard the reaction with probability $1 - \hat{P}(\hat{S})$. Let us introduce the discrete random variable $\xi \in \{1, 2, \ldots\}$ that, in a given state $\underline{\mathrm{X}}(t)$ of the system, counts how many times the PDM algorithm has to repeat the Decision step in one iteration, until a reaction can be executed. In this case, the reaction is

discarded $\xi - 1$ times until the next realized events.

ξ is of a geometric distribution with parameter $\hat{P}(\hat{S})$, the probability of repeating the Decision step k times in one iteration of the algorithm is:

$$\hat{P}(\xi = k) = (1 - \hat{P}(\hat{S}))^{k-1}\hat{P}(\hat{S}) \tag{4.28}$$

In every trial $l \in \{1, \dots, k\}$ the algorithm generates a random sojourn time $\tau_l \sim Exp(\hat{a})$. Given that there were k trials, and since $\hat{\tau}_1, \dots \hat{\tau}_k$ are independent and exponentially distributed with parameter $\hat{a}$, the sum of these soujurn times $\hat{T} = \sum_{l=1}^{k} \hat{\tau}_l$ is a continuous random variable of Erlang-k distrubution. Its probability density function can be calculated with convolution resulting $f(t, k, \hat{a})$ and its cumulative distribution function is $F(t, k, \hat{a})$ with $t, \hat{a} > 0$, and $k \in \{1, 2, \dots\}$ [43]:

$$f(t, k, \hat{a}) = \frac{\hat{a}^k \cdot t^{k-1}}{(k-1)!} \cdot e^{-\hat{a}t} \qquad F(t, k, \hat{a}) = 1 - e^{-\hat{a}t} \sum_{l=0}^{k-1} \frac{1}{l!}(\hat{a}t)^l.$$

We may obtain a formula for the joint probability $\hat{P}(\hat{T} < t, \xi)$ by conditioning on ξ:

$$\hat{P}\left(\sum_{l=1}^{k} \hat{\tau}_l < t, \xi = k\right) = \hat{P}\left(\sum_{l=1}^{k} \hat{\tau}_l < t | \xi = k\right) \cdot \hat{P}(\xi = k) = \left(1 - e^{-\hat{a}t} \sum_{l=0}^{k-1} \frac{1}{l!}(\hat{a}t)^l\right) \cdot (1 - \hat{P}(\hat{S}))^{k-1}\hat{P}(\hat{S}).$$

The marginal distribution $\hat{P}(\hat{T} < t)$ can be derived with summation for k:

$$\begin{aligned}\hat{P}(\hat{T} < t) &= \sum_{k=1}^{\infty} \left(1 - e^{-\hat{a}t} \sum_{l=0}^{k-1} \frac{1}{l!}(\hat{a}t)^l\right) \cdot (1 - \hat{P}(\hat{S}))^{k-1}\hat{P}(\hat{S}) \\ &= \sum_{k=1}^{\infty} (1 - \hat{P}(\hat{S}))^{k-1}\hat{P}(\hat{S}) - \sum_{k=1}^{\infty} e^{-\hat{a}t} \sum_{l=0}^{k-1} \frac{1}{l!}(\hat{a}t)^l (1 - \hat{P}(\hat{S}))^{k-1}\hat{P}(\hat{S}).\end{aligned}$$

Where the first sum is trivially 1. To obtain the second sum, we rearrange the terms and change the summation index l to $m = l + 1$:

$$\begin{aligned}
\hat{P}(\hat{T} < t) &= 1 - e^{-\hat{a}t} \sum_{k=1}^{\infty} \sum_{m=1}^{k} \frac{(\hat{a}t)^{m-1}}{(m-1)!} (1 - \hat{P}(\hat{S}))^{k-1} \hat{P}(\hat{S}) \\
&= 1 - \hat{P}(\hat{S}) e^{-\hat{a}t} \sum_{m=1}^{\infty} \frac{(\hat{a}t)^{m-1}}{(m-1)!} \sum_{k=m}^{\infty} (1 - \hat{P}(\hat{S}))^{k-1} \\
&= 1 - \hat{P}(\hat{S}) e^{-\hat{a}t} \sum_{m=1}^{\infty} \frac{(\hat{a}t)^{m-1}}{(m-1)!} \sum_{n=0}^{\infty} (1 - \hat{P}(\hat{S}))^{n+m-1} \\
&= 1 - \hat{P}(\hat{S}) e^{-\hat{a}t} \sum_{m=1}^{\infty} \frac{(\hat{a}t(1 - \hat{P}(\hat{S})))^{m-1}}{(m-1)!} \sum_{n=0}^{\infty} (1 - \hat{P}(\hat{S}))^{n} \\
&= 1 - e^{-\hat{a}t} e^{\hat{a}t(1-\hat{P}(\hat{S}))} = 1 - e^{-\hat{a}\hat{P}(\hat{S})t},
\end{aligned}$$

since $\sum_{n=0}^{\infty}(1-\hat{P}(\hat{S}))^n = 1/\hat{P}(\hat{S})$. With Eq. (4.24), we have $\hat{P}(\hat{T} < t) = 1-e^{-at}$, which is the cumulative distribution function of the exponential distribution with parameter a. Hence, we proved that $\hat{T} \sim Exp(a)$ with parameter a defined in the RRM algorithm (Eq. 4.7). ■

Corollary 4.2 (On the interchangeability of the algorithms)**.** *The immediate consequence of the theorem is that in any state of the system, with $\hat{a}, a > 0$, after the last step of the currently running algorithm (which is step 8 for PDM and step 6 for RRM), the running algorithm can be interchanged with the other one.*

4.5.1 Some notes on the equivalence

A distinct advantage of the RRM and mRRM methods compared to the PDM is that they use much more information about the state of the system. Naturally, equivalence only holds when both algorithms can be implemented and this is guaranteed by the restrictions we made in Sec. 3.2. We may relax these restrictions, but it is not guaranteed that we would be able to implement the desired model in the PDM framework.

One possible and useful extension is considering not one, but several cell types in the model. This requires a non-trivial data structure to efficiently follow the different types of cells in the population. In our paper, *Propensity matrix method for age dependent stochastic infectious disease models*, we proposed such a framework, the so called 'propensity matrix' that is designed to account an inhomogeneous population. In the original context, we applied age stratification on a human population, but this approach may be easily transferred to the context of cell cultures. This way one may build and efficiently simulate exact stochastic tissue models with several different cell types.

4.6 Running time

We may expect that different machine times are required due to the different formulation of the algorithms to execute a full iteration in a given state of the system. The key point between the PDM and the RRM is that the PDM may repeat Steps 3-6 several times according to the model and the state of the system. We shall examine this problem in this section.

4.6.1 The number of trials in the PDM algorithm

First, we obtain an exact formula for the expected number of trials m_ξ (the number of repeating the Decision step between two consecutive, realized reactions) in the PDM algorithm.

Proposition 4.3. *In the PDM algorithm, given that the system is at state $\underline{X}(t)$, during one iteration the expected number of trials is:*

$$m_\xi = \frac{\sum_{\mu=1}^{M} r_\mu N}{\frac{1}{4}\sum_{\nu=1}^{\alpha} r_\nu \sum_{j=0}^{4} jN_j + \sum_{\nu=\alpha+1}^{\alpha+\beta} r_\nu N + \frac{1}{4}\sum_{\nu=\alpha+\beta+1}^{M} r_\nu \sum_{j=0}^{4}(4-j)N_j}. \quad (4.29)$$

Where N is the number of cells in the population and N_j is the number of cells with j free neighbors in state $\underline{X}(t)$.

Proof. As shown in the proof of the equivalence, the number of trials during one iteration ξ is of geometric distribution with parameter $\hat{P}(\hat{S})$. Thus, the expected number of trials is $m_\xi = (\hat{P}(\hat{S}))^{-1}$. After simplification, we arrive at the relationship of the proposition. ■

Proposition 4.4. *In the PDM algorithm, consider a model in which the capacity K of the lattice is finite, i. e., there can be at most K cells in the population.*
If $N \to K$, then

$$m_\xi \to \frac{\sum_{\mu=1}^{M} r_\mu}{\sum_{\nu=\alpha+1}^{M} r_\nu}.$$

Proof. Because of $N \to K$ for the number of empty sites $e = (K - N)$, we have $e \to 0$. Hence, for the number of cells with 0 free neighbors we have $N_0 \to K$. And because of $N - N_0 \to 0$ we have $N - N_0 = (N_1 + \cdots + N_4) \to 0$. And since $N_0, \ldots, N_4 \geq 0$ we have $\sum_{j=0}^{4} jN_j \to 0K = 0$ és $\sum_{j=0}^{4}(4-j)N_j \to 4K$. Taking $N \to K$ in Eq. (4.29) and substituting these limits we obtain the proposition. ■

Corollary 4.5. *In the PDM algorithm, consider a model in which the capacity K of the lattice is finite, and suppose that the model only has contact-inhibited reactions ($r_\nu = 0$ for $\nu \in \{\alpha+1, \ldots, M\}$). Now, if $N \to K$, then $m_\xi \to \infty$.*

In this latter case, the number of trials in the PDM algorithm, and also its running time, grows without bonds. Naturally, this would never happen in practice, as we would stop the simulation whenever $N = K$. Nevertheless, these propositions show that the number of trials and the simulation time may be extremely large. In contrast, in the RRM algorithm we never face this issue, since the selected reactions are always executed.

4.6.2 Numerical comparison

Fig. 4.2 shows a numerical comparison of the running times of the PDM and the mRRM algorithms. We choose a lattice of size $K = 100 \cdot 100$ and fixed proliferation rate $r_p = 1$. Then we experimented with 3 different mobility rates: $r_m = 1, 0.5, 0$. In case of each r_m we randomly occupied $q = N(0)/K$, $q \in \{5\%, 10\%, \ldots, 95\%\}$ of the lattice with initial cells, then run and averaged out 50 simulations. The markers show the corresponding simulation times $\Theta_{mRRM}(r_m, q)$ for each initial setting q normalized to the running time $\Theta_{PDM}(r_m, q)$ of the PDM algorithm, see the horizontal line on the figure, for easier comparison.

With our implementation and with the data structure we chose, for $r_m = 1$ and a small number of initial cells the mRRM is 30% *slower* than the PDM, however with zero mobility rate the mRRM may be 30% *faster* than the PDM. In the other hand, for greater than 50% initial occupancy the mRRM always perform better, in fact it may be up to 10 times faster than the PDM. Therefore, depending on the implementation and data structure, mRRM can indeed provide a notable speed increase.

In some situations, e. g., in a scratch essay the number of modeled cells may affect the outcome of the simulation as the cells may engage in other reactions, such as the cellular or non-cellular (bio)chemical reactions, that are able to influence the time evolution of the population. Therefore, enabling larger lattices can offer extensive benefits in terms of the reliability and the biological relevance of the model.

4.6.3 Methods to decrease running time

In Sec. 2.3 we described several resourceful methods to speed up the algorithm. Probably the most convenient approach at first is to only update the propensities if there were changes involving them. Rearranging the reaction indices and checking the rapid reactions first is also a feasible approach. Identifying and separating fast and slow reactions or tau leaping can be easily applied to product synthetic reactions, but must be treated with care in case of motility, division or any type of death event as conflict between reactions may arise (movement of a dead cell to a non empty site, as an extreme example).

Finally, according to corollary 4.2, the PDM and any version of the RRM can be

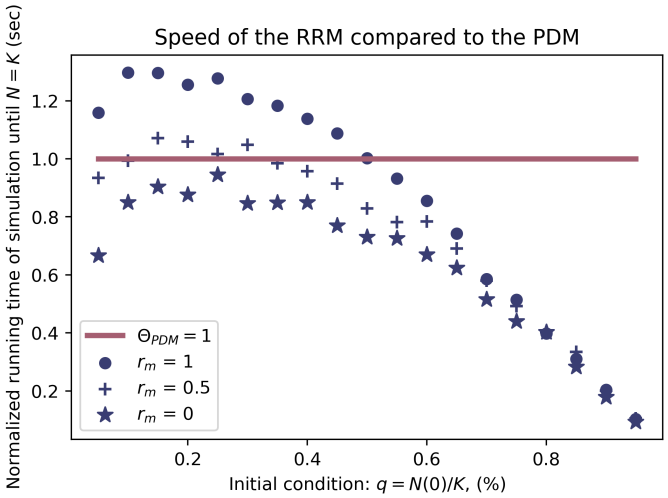

Figure 4.2: **Normalized running times of the mRRM algorithm compared to the PDM algorithm.** In these experiments the proliferation rate $r_p = 1$ is fixed and different mobility rates $r_m = 1, 0.5, 0$ were set (corresponding to the markers). For each r_m $q \in \{5, 10, \ldots, 95\}$ percent of the lattice was occupied by randomly placed cells, then 50 simulations were run and averaged out. We may observe a dramatic relative speed increase of the mRRM algorithm compared to the PDM after 50% of lattice utilization.

used interchangeably. Therefore, after some pre-runs and considering Proposition 4.3, one may identify the algorithm that performs the best at the regime of the state space under consideration.

4.7 Examples

In this section, we present two toy models to illustrate the newly introduced reactions. We used the mRRM algorithm in both simulations. We ran and averaged out 30 stochastic realizations with synthetic parameters. Any of the experiments can be readily repeated with the provided Python code. We present our findings without rigorous mathematical treatment.

4.7.1 Contact dependent and spontaneous product synthesis

Synthesis and secretion of biochemical substances are fundamental physiological processes in living organisms. Cells communicate, defend themselves, and influence each other with these processes. In this section, we demonstrate experiments with the three essentially different types of synthesis reactions we introduced in Sec. 3.2 (see Table 3.1). As it can be seen on Fig. 4.3, we started to grow a cell population on a lattice with capacity $K = 50 \cdot 50 = 2500$. The initial number of cells were 1% of the capacity, that is $N(0) = 25$ cells randomly placed on the lattice. We assumed that the cells engage in contact-inhibited, spontaneous and contact-promoted cellular biochemical synthesis reactions $\mu = 2, 3, 4$, producing one unit amount of chemical substances S_{ci}, S_s, S_{cp}, respectively. We chose $S_{ci}(0) = S_s(0) = S_{cp}(0) = 0$ for initial conditions. Let the index of the proliferation reaction be $\mu = 1$. We assumed there are no other reactions in

this system. We assume that these products are synthesized independently from each other, thus in r_2, r_3, r_4 the reactant combinations $h_2 = h_3 = h_4 = 1$, we chose $r_1 = 1$. The corresponding propensities are:

$$a_1 = \frac{r_1}{4}\sum_{j=0}^{4} jN_j, \qquad a_2 = \frac{r_2}{4}\sum_{j=0}^{4} jN_j, \qquad a_3 = r_3 N, \qquad a_4 = \frac{r_4}{4}\sum_{j=0}^{4}(4-j)N_j.$$

With little calculation we can simplify the sum a to:

$$a = (r_3 + r_4)N + \frac{r_1 + r_2 - r_4}{4}\sum_{j=0}^{4} jN_j.$$

The waiting time and the next reaction is selected according to Eq.(4.17) and the class of the proliferating cell is selected according to Eq.(4.18). We made the following assumptions: there is no cell death in the population, the probability of cell division is

$$p_1 = \frac{\frac{r_1}{4}\sum_{j=0}^{4} jN_j(t)}{(r_3 + r_4)N + \frac{r_1+r_2-r_4}{4}\sum_{j=0}^{4} jN_j} > 0$$

whenever $0 < N < K$, and only zero when $N = K$. Therefore, for the number of cells we expect $N \to K$ as $t \to \infty$, as it can be seen on the figure.

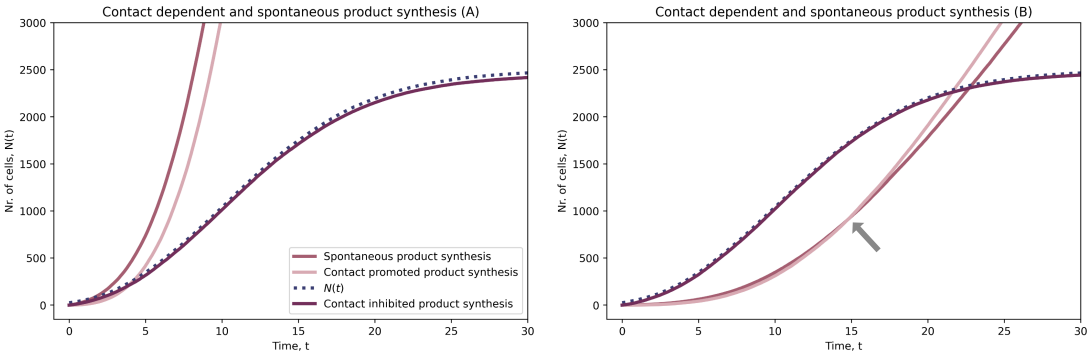

Figure 4.3: **The effect of contact dependent and spontaneous product synthesis**

We now estimate the time evolution of products $S_{ci}(t), S_s(t), S_{cp}(t)$. Following the reasoning of Proposition 4.4, it is easy to show that as $N \to K$, $p_1, p_2 \to 0$ and $p_3 \to r_3/(r_3 + r_4)$ and $p_4 \to r_4/(r_3 + r_4)$. Thus we expect $S_{ci}(t)$ to go into saturation after the size of the population reached the capacity of the lattice, leaving only the spontaneous and contact-promoted synthesis reactions to fire. From their limiting probabilities we may expect that the averaged curves $S_s(t), S_{cp}(t)$ approximate straight lines, as the same amount of product S_s and S_{cp} is generated on average per unit time, and the population remains constant after it reached the available capacity. Notice that $0 < 1/4 \cdot \sum_{j=0}^{4} jN_j \leq 1/4 \cdot \sum_{j=0}^{4} 4N_j = N$ for all $N \in \{1, \ldots, K\}$. Thus, in case $r_2 \leq r_3$, the probability of the contact-inhibited synthesis reaction is less than

or equal to the probability of the spontaneous synthesis reaction. We may expect a similar relation between the contact-promoted reaction and the spontaneous reaction as $0 \leq 1/4 \cdot \sum_{j=0}^{4}(4-j)N_j = N - 1/4 \cdot \sum_{j=0}^{4} jN_j \leq N$.

In the experiment of subfigure (A) of Fig.4.3, we chose $r_2 = r_3 = r_4 = 1$. We can see that $S_s(t) > S_{cp}(t)$ and the curves smooth to two parallell lines (since $r_2 = r_3$) as we may expect based on the above. On subfigure (B), we set $r_2 = 1, r_3 = 0.085, r_4 = 0.1$. In the beginning, the amount of product synthesized in the contact-inhibited reaction is much grater than the amount of products generated by the other two reactions, due to the difference in their reaction rates. Since $r_3 \leq r_4$ $S_s(t)$ and $S_{cp}(t)$ do not align to parallel lines for any N, in fact, they intersect as it is indicated in the figure.

4.7.2 Contact dependent and spontaneous cell death

In this section, using this toy model, we illustrate the effect of the three possible types of death that were introduced in Sec.3.2. Fig. 4.4 represents the averaged output of the corresponding 30 independent simulations of six different experimental settings. Specifically, in every single experiment we set the proliferation rate to $r_1 = 1$ and chose exactly one type of cell death also with rate 1. Initially $N(0)$ cells were randomly placed on the lattice. The waiting time and the next reaction is selected according to Eq.(4.17) and the class of the target cell is selected according to Eq.(4.18).

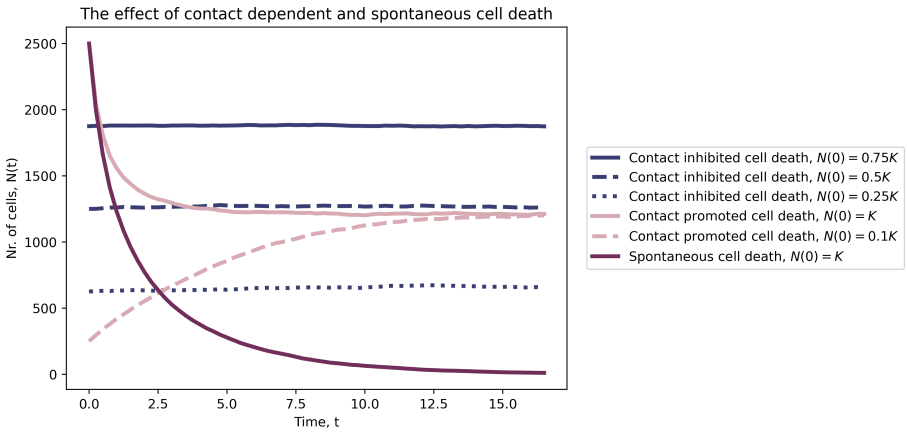

Figure 4.4: **The effect of different types of cell death on the population.** The figure shows the average outcome of six independent experiments after 30 runs. We assumed that in all cases the capacity of the lattice was $K = 50 \cdot 50 = 2500$ and the proliferation rate was $r_1 = 1$; each curve corresponds to one experiment. Blue pink, and purple curves illustrate the effect of contact-inhibited, contact-promoted, and spontaneous cell deaths, respectively.

During the first three experiments (blue curves) we assumed contact-inhibited cell death with three different initial conditions, $N(0) = 0.75K, 0.5K, 0.25K$. The corresponding propensities and probabilities are:

$$a_{\text{birth}} = \frac{1}{4}\sum_{j=0}^{4} jN_j(t), \qquad a_{\text{death}} = \frac{1}{4}\sum_{j=0}^{4} jN_j(t), \qquad p_{\text{birth}} = p_{\text{death}} = \frac{1}{2}.$$

As proliferation is also a contact-inhibited event, the probabilities of proliferation and death are the same at any state $\underline{X}(t)$. Thus, we may expect that the population can be sustained in a constant level if $N(0)$ is sufficiently large that $N(t)$ cannot drop to the absorbing state $N = 0$ due to stochastic fluctuations.

In our second experiment (pink curves) we assumed contact-promoted cell death and chose $N(0) = K, 0.1K$. The corresponding propensities are:

$$a_{\text{birth}} = \frac{1}{4}\sum_{j=0}^{4} jN_j(t), \quad a_{\text{death}} = \frac{1}{4}\sum_{j=0}^{4}(4-j)N_j(t).$$

With both initial conditions the size of the population tends to $K/2 > N \approx 1205$ where the probabilities of proliferation and death equals. Notice that these curves are under the blue dashed curve, which is around $K/2$. This is due to the reflecting boundary conditions we use in the experiments (see Sec. 3.1): any cell on the boundary can have at most 3 free adjacent sites.

In the third experiment (purple curve) we assumed spontaneous cell death. The propensities are

$$a_{\text{birth}} = 1/4 \cdot \sum_{j=0}^{4} jN_j, \qquad a_{\text{death}} = N.$$

And since $r_1 = r_2 = 1$ and $1/4 \cdot \sum_{j=0}^{4} jN_j \leq N$ for $N \in \{1, \ldots, K\}$ the probability of death is always greater than the probability of cell division, thus we may expect the population to die out in the long term, as it can be seen on the figure.

4.8 Discussion

In this chapter, we provided an extended version of the cell simulation algorithm introduced by Baker and Simpson [6] that includes new reaction types. We call this algorithm the Prompt Decision Method (PDM).

The PDM algorithm is a classical accept–reject montecarlo simulation algorithm that may be very computationally intensive due to the possible large number of rejections in a crowded environment.

We introduced the Reduced Rate Method (RRM) algorithm that is based on the classification of cells according to their free adjacent sites. In this algorithm, every selected reaction is executed, and therefore, the running time and the resource requirements see dramatic reductions. Furthermore, we also showed that the PDM and the RRM methods generate the same stochastic process as the distribution of the waiting time until the next successful reaction; also, the probabilities of the next successful reaction are the same in both approaches in a given state of the system. The RRM algorithm is an exact stochastic simulation algorithm that is derived from Gillespie's

method from its first principle. Since the RRM algorithm is equivalent to to PDM method, the PDM method is also an exact stochastic simulation method. Indirectly, we showed that the algorithm introduced by Baker and Simpson, which is a special case of PDM, is also exact.

We also introduced an equivalent, but simplified version of the RRM, called the marginal Reduced Rate Method (mRRM). Even though this algorithm is easier to implement, it still retains all the benefits of the RRM algorithm. The RRM and mRRM algorithms have several advantages compared to PDM:

- RRM is considerably faster in a crowded population. In contrast with PDM, in the RRM and mRRM algorithms the selected reaction is always executed due to the different formulation of the propensities. This is a particularly important feature in cell culture models, as the immotile cells or the ones that are unable to proliferate do not slow down the simulation time. Moreover, this 'seemingly unnecessary' population of cells may still engage in other important processes needed for the development of the whole population.
- Furthermore, as a result of this, the probability of the events is known in every iteration of the algorithm. Thus, giving a probabilistic interpretation of different emerging phenomena is more straightforward in a simulation done with the mRRM algorithm (see Sec. 4.7).
- Since the design of the RRM and mRRM algorithm is based on the classification of cells according to their free adjacent sites, including new classes to this method is straightforward. For example, one may include several cell types, or even extend the model incorporating the fact that the chemical substances $S_1, \ldots, S_\omega$ may influence the state of the cells. This could be achieved in the PDM method only in very inconvenient ways, if it is possible at all.

5

Cell culture simulation with explicit cell cycle length

5.1 Motivation

Cell cycle is a series of precisely controlled events that take place in a cell as it grows and divides. During the cycle, cells undergo many structural changes, each of which takes a considerable amount of time. As we will see later, this makes it much more difficult to use classical compartmental models, which assume that the time spent in each compartment is exponentially distributed. But what does actually happen during cell cycle?

Division of eukaryotic cells starts with the synthesis of the necessary RNA and protein structures during the so called *G1 phase*. Cells go through several checkpoints to ensure that the necessary number of molecules are present to start the replication of the genetic material. The last such checkpoint is the so called *no return point* after which the molecular machinery that is assembled in G1 begins to copy the DNA of the cell. This is called the *S phase*, where 'S' stands for the synthesis of the nuclear DNA. When the two DNA copies are ready for the two daughter cell, the mother cell enters the *G2 phase* in which continues protein and RNA synthesis and approximately doubles its size. Then, the mother cell goes through nuclear division (mitosis) and cell division (cytokinesis) producing two daughter cells during the *M phase*. After M, cells may go into the quiescent *G0 phase*, that may last hours or even the whole lifetime of the cells (depending on its type and the circumstances), or start the whole process over. For further details we refer to the excellent textbook by Nelson and Cox [38], the cell cycle is illustrated in Fig. 5.1.

Given a cell population, cells may be in different stages of the cell cycle. A population is synchronized, when the cells are in more or less the same state. In contrast, when most of the cells are in different state, the population is asynchronous. Naturally, it may be partially synchronized as well (c.f. Fig. 5.1).

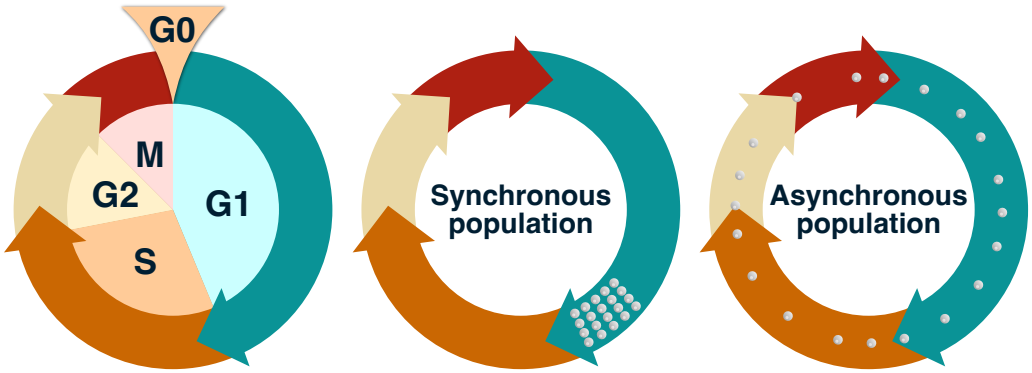

Figure 5.1: Cell cycle, the series of biochemical events that lead to cell division. The diagram on the left illustrates the order of the phases in the cycle: G1, S, G2, M. After mitosis cells may enter the quiescent G0 phase. The whole cycle (without G0) last for several hours, typically between 8-24 hours. The diagram in the middle illustrates a synchronous population, when most of the cells are in the same phase in their cycles. The one on the right illustrate an asynchronous population.

5.1.1 Mathematical models of cell cycle

Cell cycle complicates both mathematical modeling and simulating cell cultures. In fact, when cell cycle play important role in the behavior of the cell culture, exponentially distributed cell cycle length may lead to false conclusions as "the most probable time for a cell to divide is the current time" (Yates et al. [48]). Each phase of the cycle is of typical length, usually several hours. Thus, when the length of the cell cycle plays an important role, assuming exponentially distributed length is a poor choice.

Moreover, as Vitadello et al. point out, the classical exponential and logistic cell population models, and their variants inherently assume that the concerned population is asynchronous throughout the modeled period of time [46]. This may have important implications on various fields, including tissue development or even tumor therapy. The classical approach, with system of a small number of ODEs, is neither capable of capturing the true dynamics influenced by the length of the cell cycle, nor able to model synchronized population or spontaneous synchronization.

Based on 77 data sets for 16 cell types, Golubev found that the best fitting distribution for the experimentally measured cell cycle lengths was the exponentially modified gamma or exponentially modified Gaussian distribution [31]. To approximate this, using an integer shape parameter k to keep the Markovian property of the agent based stochastic framework, Yates et al. modeled the cell cycle with an exponentially modified Erlang distribution. They suggest a general multistage model of the cell cycle that consists of a chain of consecutive states, where the transition rate between $X_i \to X_{i+1}$ for $i \in \{1, \ldots, k-1\}$ is λ_1, cells spend an exponentially distributed waiting time in each state with mean λ_1^{-1}. Finally, cells leave the terminal state of the cycle and split to two daughter cells with rate λ_2 [48]:

$$X_1 \xrightarrow{\lambda_1} X_2 \xrightarrow{\lambda_1} \ldots \xrightarrow{\lambda_1} X_k \xrightarrow{\lambda_2} 2X_1.$$

Vitadello et al. applied this approach to explain the behavior of certain melanoma cell lines [46]. They used a cell cycle dependent fluorescent dye (FUCCI) to label cells with different colors according to three states in the cycle: G1, early S, and from late S till the end of the mitosis. They recorded the change of the number of cells in the states. The authors assumed that the time cells spend in these 3 stages are all of independent Erlang distributions with the corresponding typical parameters. They managed to achieve a remarkably good fit to experimental data at the cost of an extreme large number of intermediate stages. The number k of stages varies between 30-120, depending on the considered cell line [46].

We may therefore say that multistage models usually require a large number of artifact stages that make it problematic to investigate analytically and interpret biologically. Moreover, such models, relying on a Markovian process does not include the length of the cell cycle as an explicit parameter.

Our framework is highly motivated by the *transition probability model* (TPM) by Smith and Martin [42]. They observed that phase S and G2 are of characteristic length for the cell type and do not show significant variation in a given population. Thus, it is legitimate to think that the length of these phases is deterministic. On the other hand, the length of phase G1 shows high variability even in a homogeneous population. According to their TPM model, the life of the cell can be divided into two states, a proliferating state that lasts for a deterministic period of time (consists of phase S, G2 and M), and an inter division phase that we shall call motile state (consists of phase G0 and G1) that lasts for a random period of time, but the probability of leaving this state is always constant. The latter assumption about the constant transition rate implies that the length of the motile state is exponentially distributed [42].

5.1.2 The aim of this chapter

In this chapter, we take a step back and leave the spatial cell culture models for a short detour. First we include realistic cell cycle length to well stirred cell culture models and present a model that is able to incorporate cell cycle length of any distribution. Then we explain the spontaneous synchronization of cells that has been observed in the *in vitro* experiments by Vitadello et al. [46], but cannot be explained in the context of the presented multi stage model [46, 48]. Finally, we fit our model to the experimental data by Vitadello et al. [46].

5.2 The stochastic model

Now, we shall extend the well known logistic growth model $N' = rM\left(1 - \frac{M}{K}\right)$ with realistic cell cycle length. One possible way to implement the logistic growth model in

the stochastic simulation formalism is to incorporate two events: cell division and cell death. In this case the population size may show large fluctuation overshooting the carrying capacity K. Since, we want a direct comparison with the results by Vitadello et al. [46] and also with our lattice based framework, that includes volume exclusion, we choose the way we introduced in Sec 4.3.5.

Suppose that motile (M) cells commit to division and switch to the proliferating (P) phenotype in a contact inhibited reaction with proliferation rate r. Then ϑ time later, upon the cell cycle elapsed, P cells divide and switch back to the motile phenotype. To avoid overshooting and to keep the contact inhibited nature of the reaction, we assume that cells only switch phenotype, when they managed reserve space for the daughter cell. With carrying capacity K, the number of free spaces at time t is $K - M(t) - P(t) - Q(t)$, where $Q(t) = P(t)$ stands for the reserved spaces. Thus, in the well stirred model, we may obtain the following differential equation with delay:

$$\begin{aligned} M'(t) &= -rM(t)\frac{K - M(t) - P(t) - Q(t)}{K} + 2rM\frac{K - M(t-\vartheta) - P(t-\vartheta) - Q(t-\vartheta)}{K} \\ P'(t) &= rM(t)\frac{K - M(t) - P(t) - Q(t)}{K} - rM\frac{K - M(t-\vartheta) - P(t-\vartheta) - Q(t-\vartheta)}{K} \end{aligned} \tag{5.1}$$

In the last chapter of the thesis we will investigate this mean field model and present a detailed comparison with the lattice based model that is about to be introduced in the following chapter. Now, we focus on the simulation of the well stirred system.

Based on the above considerations we have two reactions: the phenotype switch $M \to P$, that is a random contact inhibited reaction with rate r/K between the M cells and the virtual $K - M - 2P$ vacant spaces, and the scheduled cell divisions $P \to 2M$. The propensity of the phenotype switching and the distribution of the corresponding waiting time is:

$$a = \frac{r}{K}M(K - M - 2P), \qquad P(\tau) = ae^{-a\tau}. \tag{5.2}$$

The deterministic cell cycle is of length ϑ, which may be constant, such as in the mean field model Eq. (5.1) or chosen from any distribution. Suppose that in a given state we have P number of proliferation cells and their division reactions are scheduled for time $\Theta_1, \ldots, \Theta_P$. From which the next delayed reaction will occur at time t_{R_D}, where

$$t_{R_D} = \min\{\Theta_i : i = 1, \ldots, P\} \tag{5.3}$$

Thus, we may summarize the transition between the states with the following flow chart:

$$M \xrightarrow{\frac{r}{K}(K-M-2P)} P \overset{\vartheta}{\Longrightarrow} 2M,$$

where $\Rightarrow$ stands for the delayed reaction. The required data structure is an integer valued variable, which we shall call M, to store the number of motile cells; and a list P

to store the the scheduled times of the division reactions. The algorithm is as follows:

Logistic growth with explicit cell cycle length SSA

1. **Initialization:**
   (i) set $t \leftarrow 0$, prescribe halting conditions H,
   (ii) initialize variables $M, \underline{P}$.
2. **Calculate propensity functions** a according to Eq. (5.2).
3. **Generate stochastic waiting time** τ according to Eq. (5.2).
4. **Obtain the termination time of the next deterministic reaction** t_{R_D} according to Eq. (5.3).
5. **Reaction selection and execution step**:
   - If $t + \tau < t_{R_D}$, the next reaction is a phenotype switching
     set $t \leftarrow t + \tau$, execute the reaction, update data structure.
   - Else, the next reaction R_D is the proliferation reaction of the cell for which Θ_i is minimal for $i \in \{1, \dots, N_P\}$,
     set $t \leftarrow t_{R_D}$, execute the reaction, update data structure.
6. **Halt if** $H = True$ **else continue the process with Step (2).**

5.3 Comparison with the mean field logistic growth

The parameters of the mean field logistic growth model are the proliferation rate r and the capacity K of the environment. It is inherently assumed that the inter division times are exponentially distributed and the cell cycle takes no time to complete. Thus, we may expect that the mean field logistic model gives a good approximation to the averaged output of the delayed logistic growth in case the length of the cell cycle is small.

To illustrate this, on Fig 5.2 we fixed $r = 1$ and $K = 10^4$ and chose initial condition $M(0) = 500, P(0) = 0$. The total number of cells is $N(t) = P(t) + M(t)$. Every figure shows the average of 20 realizations of the stochastic model with varying values of delay $\vartheta \in \{0.01, 1, 5\}$.

We may obtain a really good fit with $\vartheta = 0.01$ and it can be seen that almost all cells are in the inter division M state, as it is expected. For $\vartheta = 1$ the agreement between the logistic model and the stochastic simulation with cell cycle is very poor. The delay, caused by the cycle slows the development of the population shifting the curves to the right. It also creates a transient oscillation that we will investigate in the next section. A significant amount of proliferating cell appear due to the large cell cycle causing the $N(t)$ and $M(t)$ curves not to coincide. Finally, for $\vartheta = 5$, the non-monotonic behavior of the $P(t)$ curve becomes exaggerated making the transient oscillation last longer.

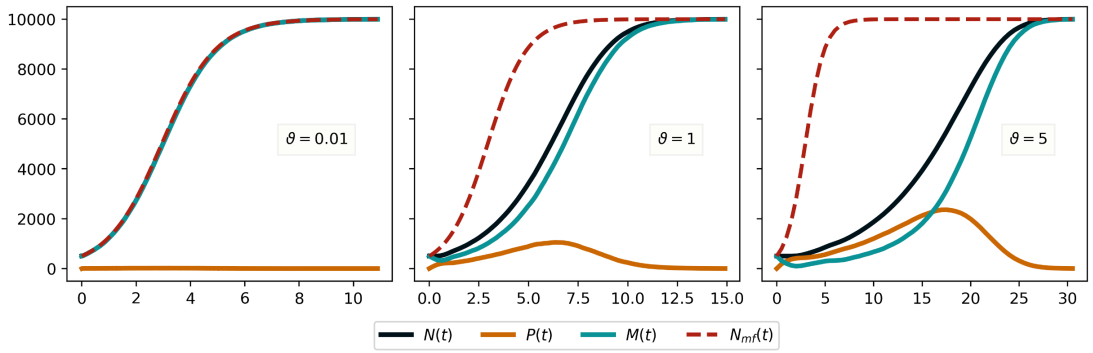

Figure 5.2: Logistic growth with explicit cell cycle length. The figure illustrates the effect of cell cycle on the dynamics. Each figure shows the averaged output of 20 independent runs with $r = 1, K = 10^4$ and initial condition $M(0) = 500, P(0) = 0$. The delay parameter, the length of the cell cycle varies on the figure: from left to right the value of ϑ is $0.01, 1, 5$, respectively. The legend on the bottom applies for all figures, $N_{mf}(t)$ stands for the mean field logistic curve.

5.4 Synchronicity

Synchronicity may occur when the time r^{-1} a cell spends in motile mode is small compared to the length of its cell cycle: $r^{-1} \ll \vartheta$. This causes the cells to enter the proliferating state more or less together resulting the unusual step-like growth dynamics that can be seen on the left part of Fig. 5.3. This special dynamics may last until the population reaches the capacity K in case of certain parameter configuration. We will spend more time investigating this in the very next chapter. Remarkably, cells are not coupled in any way, they do not influence each other's behavior. Thus, it is different from the synchronization of coupled oscillators.

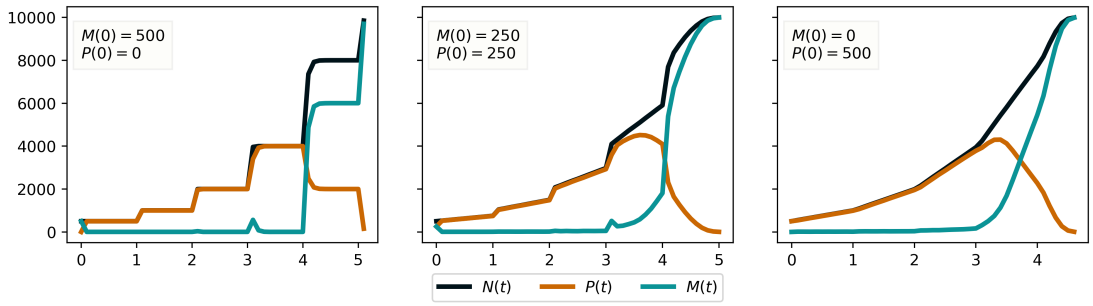

Figure 5.3: Step-wise synchronization of cells may last for a long time, but it is sensitive for the initial conditions. The figures show that the characteristic step-like dynamics may cease to exist if the initial cell population is in the proliferating state and their scheduled division events are scattered uniformly in the interval $[0, \vartheta]$.

Synchronization is very sensitive for initial conditions: on Fig. 5.3 we set $\vartheta = 1$ and $r = 100$ for all figures, then we plotted the averaged time series of 20 runs for 3 different initial conditions. In case of the leftmost figure all cells are motile initially, $M(0) = 500$. The high rate forces the cells to enter the proliferating state during a very short time interval. Then, the cells spend as long as $\vartheta = 1$ time in state P, resulting a horizontal line of length about ϑ. Upon the cell cycle elapsed, the population doubles its size in a short time.

On the middle plot of Fig. 5.3 the number of motile and proliferating cells are the

same: $M(0) = P(0) = 250$. As cells may be in infinitely many stages in their cell cycle we have to prescribe initial conditions for the cells in state P. We chose the scheduled times Θ for the delays from continuous uniform distribution $\Theta \sim \text{Uniform}(0, \vartheta)$ as initial condition. It is well known that on any finite interval $[a, b]$ among all the continuous distributions, for which the support is this interval, $\text{Uniform}(a, b)$ has the maximum entropy. Since the scheduled delay reactions are scattered uniformly in the interval $[0, \vartheta]$ the time between them is approximately the same, $\vartheta/250$. Making the cells to reenter state M with a constant rate, causing a slope in the straight line segments.

Finally, on the middle plot of Fig. 5.3 all cells are proliferating and their scheduled divisions are scattered uniformly in the interval $[0, \vartheta]$. This eliminates the synchronization.

5.5 Applying the model to experimental data

5.5.1 Experimental data

Now, we apply our model to the experimental data provided by Vitadello et al. in the supplementary material of their article *Mathematical models incorporating a multi-stage cell cycle replicate normally-hidden inherent synchronization in cell proliferation* [46].

The authors run several *in vitro* laboratory experiments on three melanoma cell lines: C8161, WM983C, and 1205Lu. The cells were transfected to express a cell cycle dependent fluorescent dye, FUCCI. This dye labels cells according to their state in the cell cycle. The nucleus of a cell in phase G1 fluoresces red, it fluoresces green in phase S, G2, and M, and in the early phase of S both colors appear and the nucleus fluoresces yellow. The population was monitored for 48 hours and photographed every 15 minutes under microscope. Then, from each image, the number of cells were counted and a multistage mathematical model were fit to the time series. The authors claim that they applied a careful preparation on the cell culture to prevent any induced synchronization. For further details we refer to the supplementary material of [46].

5.5.2 Fitting to the data

According to the multistage model [46] the cell cycle length is of exponentially modified Erlang distribution, where its shape parameter k corresponds to the number of stages. As we use delay to model the cell cycle we do not have to stick to the integer value of the shape parameter. We assume that the length of the cell cycle is of gamma distribution with shape parameter α and rate parameter λ to be fitted, $\vartheta \sim Gamma(\alpha, \lambda)$. The initial condition for the number of motile cells $M(0)$ and the number of proliferating cells $P(0)$ is given by the experiment. However, the initial distribution of cells in the proliferating state (the length of the remaining times of their cell cycle) have to be

determined. We assume that it is also of gamma distribution with parameters α_0 and λ_0. Finally, we have to determine the proliferation rate r, that is the parameter of the waiting time τ of exponential distribution: $\tau \sim Exp(r)$. Carrying capacity K was set to 10^5.

Thus, the vector of parameters to be determined is: $(r, \alpha, \lambda, \alpha_0, \lambda_0)$. We used the Nelder-Mead simplex optimization method implemented in the Scipy Python library to minimize the square of the difference between the simulated and the measured time series.

The parameter fitting was performed on one experiment each for cell lines C8161 and WM983C (Fig. 5.4). As shown in the figure, our model fits the data series remarkably well in case of cell line C8161 and WM983C. However, for cell line 1205Lu, the fitting did not show such good results, which could be due to the data series being too noisy or the one class delay scheme not being sufficient. Probably a model that splits the cell cycle into two phases is needed. This shall be a subject of further investigations.

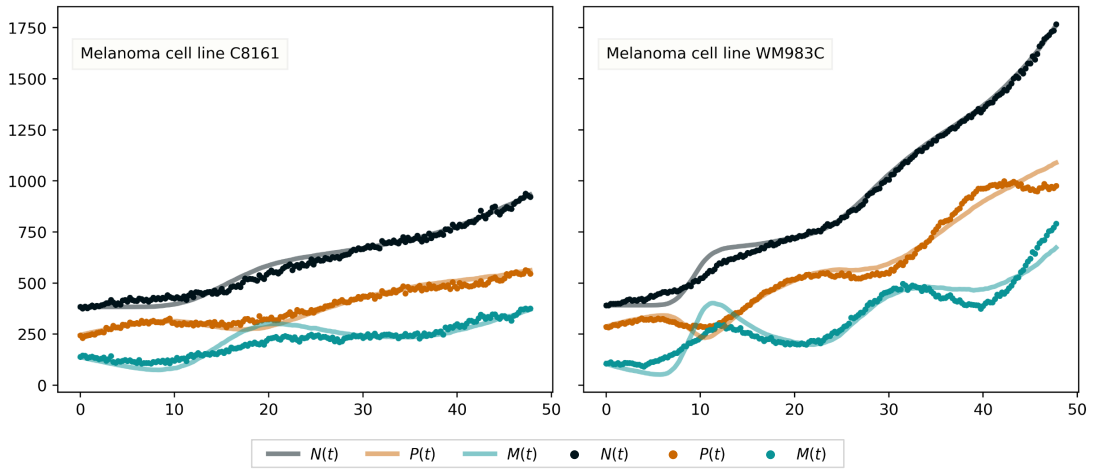

Figure 5.4: Cell cycle model with delay fitted to the experimental data by Vitadello et al. [46]. The scatter plot represents the measured data points, the solid lines are the simulation outputs with the fitted parameters.

5.6 Discussion

In this chapter, we have presented an exact stochastic simulation model, which is a feasible alternative to the multi-state models [46,48], we have presented in the beginning of the chapter. In agreement with the TPM [42] hypothesis, our model divides the cell cycle into two stages. One state from which cells leave with constant probability and one whose length is of any distribution required by the model.

Our approach includes a delay instead of several stages in a linear chain. One advantage of this approach is that we can model a cell cycle with length of arbitrary distribution. The cell cycle may even be of constant length, which is only possible in the limiting case of the multilevel model when the number of stages k tend to infinity.

Another advantage of our approach is low computational cost. Unlike in our model, in multistage models the system spends most of the simulation time to move cells between stages. Thus, with k number of stages the simulation engine has to perform at least k units of calculations (including waiting time and reaction selection and accounting the change in the state variables) to push one cell through the stages. In our model one cell goes through the cell cycle in only two steps requiring two units of calculations.

Third, our model has far less parameters than the proposed multistage model by Vitadello et al. Thus, fitting parameters and initial conditions is much easier. It is worth noting that in a multistage model the number k of stages is also a subject to parameter fitting. And k may vary between experiments even in case of one cell line. On the other hand, parameters in our method may be easily interpreted by measurable real life quantities.

Additionally, we use a version of a logistic model, which allows us to account for the limitation K of the living space. Also, for large K and small population we approximate exponential growth.

Finally, we have shown that our model allows for synchronicity of cells and fits well to real experimental data.

6

Go-or-Grow models with realistic cell cycle length

6.1 Motivation

Observation of certain tumor cells, the so called *Malignant Gliomas*, suggest that there is a trade-off between cell proliferation and migration [26]. The go-or-grow hypothesis emerged from this phenomenon.

This hypothesis asserts that the proliferation and the migration of certain cells spatiotemporarily exclude each other. Experiments suggest that, in the case of glioblastoma, one can distinguish two phenotypes of the malignant cells: the 'high proliferation–low migration' *proliferating type* and 'low proliferation–high migration' *motile type* [26]. The mechanisms of the phenotype switching is highly investigated in medical research, although there is no convincing explanation for the phenomena on the molecular level so far. However, it is suggested that cytoskeleton dynamics, which has a significant role in both migration and proliferation, is a limiting factor of the process [24].

The go-or-grow hypothesis was tested in the extensive study of Concoran et al. [14] and Garay et al. [23] and was not verified for *lung*, *mesothelioma* and *melanoma* cancer cell types. Nevertheless, Garay and colleagues [23] point out that "tumor cells from various organ systems may differently regulate migration and proliferation".

In contrast, according to the analysis in [18] there is a strong evidence that transplanted "glioma cells (...) migrated in a saltatory fashion with burst of migration separated by periods of immobility" (Farin et al. [18]). Using time-lapse video-microscopy, the authors observed that "cells proliferated *en route*, pausing for as short as an hour to divide before the daughter cells resumed migrating" (Farin et al. [18]).

6.1.1 The aim of this chapter

This chapter has two objectives. First, we extend the method we proposed in the previous chapter to model cells on a grid with a realistic cell cycle length. Then, we investigate the go-or-grow phenomenon by obtaining new appropriate spatial stochastic models with a delay that includes the cell cycle length as an explicit parameter. As we discussed in Chapter 5, Vitadello et al. experimentally showed that cells are able to synchronize even in an initially asynchronous population. They also provided a multistage model that fits well to experimental data.

6.2 The agent based model

Suppose that motile cells perform unbiased random walk on a two-dimensional lattice with motility rate r_M. With proliferation rate r_P motile cells switch to proliferating type and stop for a fixed period of time ϑ to complete cell cycle, upon which they attempt to divide and place a daughter cell on a randomly selected neighboring lattice site.

This brings forward the following problem: the randomly selected lattice site of the second daughter cell may be occupied. In other words, if the population consists of more than one cell, there is a positive probability that the neighboring sites of the mother cell would be occupied by other migrating or proliferating cells by the end of its cell cycle. It is even true when all the adjacent sites of a mother cell are vacant in the beginning of its cell cycle. This effect occurs because the memoryless property of the system is violated.

To overcome this problem, Yates et al. [48] made two assumptions: i) when cells are not able to divide they abort the whole process and return to mobile mode or ii) cells wait until a site opens up in their neighborhood. They applied their multistage models to these cases.

We made four assumptions that involve four essentially different agent-based stochastic models with delay. We hypothesize that cell division may occur according to one of the following four ways **(H1–H4)**. The first hypothesis **(H1)** corresponds to the model we proposed in the previous chapter. It is assumed that the mother cell attempts to reserve a free adjacent site in a contact-inhibited reaction. If this occurs, it switches to the proliferating phenotype and starts cell cycle. Upon which a mobile daughter cell is placed on the reserved site and the daughter cell situated at the site of the mother cell goes back to mobile mode.

In the other three hypotheses it is assumed that a motile cell switches to the proliferating phenotype in a spontaneous reaction. Then the mother cell:

- **(H2)** begins cell cycle, upon which it randomly selects an adjacent site.

If the site is empty, a motile daughter cell is placed there and the cell goes back to mobile mode. If the site is occupied, the reaction is rejected and the mother cell goes back to mobile mode.

- **(H3)** begins cell cycle, upon which it attempts to divide in a contact-inhibited reaction. Thus, the cell doesn't leave the cell cycle until space opens up for the 2nd daughter cell. We call the cells that are waiting to be able to divide non-motile *waiting type* cells. If the division is successful, a motile daughter cell is placed on a free neighboring site and the cell goes back to mobile mode.
- **(H4)** attempts to reserve a randomly selected free adjacent site. If succeeds, the mother cell begins cell cycle, upon which it places a motile daughter cell there and turns back to mobile mode. If the selected site is occupied, the mother cell turns to non-motile *waiting type* and waits until it can reserve a site to begin the cell cycle.

Naturally, a reserved site is considered to be occupied so neither the other cells can enter that site, nor they can reserve it. In this chapter we are focusing on the early stages of the development of the cell population, thus we do not include cell death into the models and chemical reactions are also not considered. In the following, we provide the Go-or-Grow Algorithms and discuss the models; then, we investigate the model predictions, such as cell synchronization and a 'step like' growth curve.

6.3 The Go-or-Grow Algorithms

In this section, a model is proposed for each hypothesis. We use the direct method combined with the prompt decision formulation to obtain the algorithms. According to the previous section, three cell types will be considered and we denote them with the characters *M, P, W*, standing for the motile, proliferating and waiting types. Naturally, the number of these cells at time t will be denoted by $M(t)$, $P(t)$, and $W(t)$, respectively. We will use the Prompt Decision Method approach that we introduced in Chapter 4 during the implementation.

6.3.1 Model 1: reserving site at first

In **(H1)**, the cell that goes into proliferation attempts to reserve one random neighboring site at first. If it is occupied, the proliferation is discarded and the cell remains *type M*. If it is free, the cell turns to the proliferating *type P*. After the cell cycle one *type M* daughter cell is placed at the reserved place and the cell turns back to *type M*.

Reactions in M1

- R_M – movement of *type M* cells with rate r_M.

- R_P – a randomly selected *type M* cell attempts to reserve a random neighboring site with rate r_P. If possible, it turns to *type P*, if not, the proliferation is discarded and the cell remains of *type M*.
- R_D – the *type P* cell with the least remaining time of its cell cycle finishes proliferation (turns back to *type M*) and places one *type M* daughter cell to the reserved neighboring site.

6.3.2 Model 2: rejecting the reaction

According to **(H2)**, right after completing the cell cycle, the proliferating cell goes back to mobile mode M and a randomly chosen neighboring site is occupied for one *type M* daughter cell if it is free. If it is occupied, then the proliferation will be discarded and no daughter cell will be created.

Reactions in M2

- R_M – movement of *type M* cells with rate r_M,
- R_P – a randomly selected *type M* cell undergoes proliferation and turns *type P* with rate r_P and stops moving for τ,
- R_D – the proliferating cell with the least remaining time of its cell cycle finishes proliferation, turns to *type M*. The cell randomly selects a neighboring site, if it is vacant, then places one *type M* cell there. If it is occupied, nothing happens.

6.3.3 Model 3: waiting for cell division

In **(H3)**, after the cell cycle elapsed, the proliferating cell attempts to place one *type M* daughter cell at a randomly selected neighboring site if it is free, then turns back to *type M*. If it is occupied, the proliferating cell turns to non-motile waiting type (*type W*) and, with activation rate r_W, attempts to place one *type M* daughter cell to one randomly chosen neighboring site if it succeeds then it turns back to *type M*, but it remains in waiting mode if it fails.

Reactions in M3

- R_M – movement of *Type M* cells with rate r_M.
- R_P – a randomly selected *Type M* cell undergoes proliferation and turns *Type P* with rate r_P and stops moving for τ.
- R_D – the *type P* cell with the least remaining time of its cell cycle finishes proliferation and attempts to put one *type M* cell to a random neighboring site. If it is possible the cell turns back to *type M*, if not, it turns to *type W*.
- R_W – a *type W* cell attempts to divide into a randomly selected neighboring site, if the neighbor is vacant it turns back to *type M* and leaves one daughter cell in

it. If the neighbor is occupied, the cell remains of *type W*.

6.3.4 Model 4: reserving site or waiting

In **(H4)** a cell that undergoes proliferation turns to *type P* and randomly reserves one neighboring site at first if possible. After the period of the cell cycle, one daughter cell is placed at the reserved site and the proliferating cell turns back to mobile mode. If the selected place is occupied, the proliferating cell turns to *type W*. W cells attempt to randomly reserve a neighboring site with activation rate r_W in order to turn to *type P* to complete the cell cycle.

Reactions in M4

- R_M – movement of *type M* cells with rate r_M.
- R_P – a randomly selected *type M* cell attempts to reserve a random neighboring site with rate r_P. If possible, it turns to *type P*, if not, it turns to *type W*.
- R_W – a *type W* cell attempts to reserve a randomly selected neighboring site. If succeeds, it turns to *type P* and starts proliferation, if it does not succeed, nothing happens.
- R_D – the *type P* cell with the least remaining time of its cell cycle finishes proliferation (turns back to *type M*) and puts one *type M* cell to the reserved neighboring site.

6.3.5 State space and data structure

Let $t \in [0, \infty)$ represent time, and $N_M(t), N_P(t), N_W(t) \in \mathbb{N}_0$ denote the number of motile, proliferating and waiting cells in the population at time t, respectively. The total number of cells in the population at t is $N = N(t) = N_M(t) + N_P(t) + N_W(t)$.

For *type M, P*, and *W* cells, we introduce the state descriptors:

$$\begin{aligned} \underline{\mathrm{M}}(t) &= (M_1(t), \ldots, M_{N_M}(t)) \\ \underline{\mathrm{P}}(t) &= (P_1(t), \ldots, P_{N_P}(t)), \\ \underline{\mathrm{W}}(t) &= (W_1(t), \ldots, W_{N_W}(t)), \end{aligned} \tag{6.1}$$

where $M_m(t), W_w(t)$ is the coordinates of the motile or waiting cell with index $m \in \{1, \ldots, N_M\}, w \in \{1, \ldots, N_W\}$ at time t on the lattice. In case of the proliferating cells, the 3 entries of vector $P_p(t) = (x_p, y_p, \Theta_p)$ stand for the x, y spatial coordinate and the scheduled time Θ of the cell division of the cell with index p for $p \in \{1, \ldots, N_P\}$.

A convenient data structure for the above state space (in Python) consists of three lists: $\underline{\mathrm{M}}, \underline{\mathrm{P}}, \underline{\mathrm{W}}$. We also keep the underlying lattice L to check if a particular lattice site

is occupied or empty. In the implementation, L is a matrix $L \in \mathbb{N}^{n_1 \times n_2}$, where entry $L[x_*, y_*]$ is 1 if a cell occupies the corresponding site and 0 if it is empty.

6.3.6 Propensities and probabilities

To facilitate the description of the algorithms, we assume that reactions R_M, R_P, and R_W are present in all four models, but in case of M1 and M2 (where there is no waiting type cell) the activation rate $r_W = 0$. Thus, for reaction R_M, R_P, and R_W we define the propensities in order:

$$a_M = r_M N_M, \quad a_P = r_P N_M, \quad a_W = r_W N_W, \quad a = a_M + a_P + a_W \tag{6.2}$$

The possible waiting time and the next possible reaction, generated by the stochastic part of the system is obtained from:

$$\mathbb{P}(\tau) = ae^{-a\tau}, \quad \mathbb{P}(\mu|\tau) = \frac{a_\mu}{a}, \quad \mu \in \{M, P, W\} \tag{6.3}$$

The deterministic reactions are scheduled for time $\Theta_1, \ldots, \Theta_{N_P}$. From which the next delayed reaction will occur at time t_{R_D}, where

$$t_{R_D} = \min\{\Theta_i : i = 1, \ldots, N_P\} \tag{6.4}$$

General script of the Go or Grow Algorithm

1. **Initialization:**
   (i) set $t \leftarrow 0$, prescribe halting conditions H,
   (ii) initialize the lattice L with size $n_1 \times n_2$
   place $N_M(0) + N_P(0) + N_W(0) \leq n_1 \cdot n_2$ cells on the lattice according to essay,
   (iii) initialize lists $\underline{\mathrm{M}}, \underline{\mathrm{P}}, \underline{\mathrm{W}}$
   store location of M, P, W cells in the corresponding list.
2. **Calculate propensity functions** a_M, a_P, a_W and a according to Eq. (6.2). Note that $r_W = 0$ in case of model 1 and model 2.
3. **Generate stochastic waiting time** τ according to $\mathbb{P}(\tau)$ from Eq. (6.3).
4. **Obtain the termination time of the next deterministic reaction** t_{R_D} according to Eq. (6.4).
5. **Reaction selection step**:
   - If $t + \tau < t_{R_D}$ select the next reaction from $\mathbb{P}(\mu|\tau)$ (Eq. (6.3)),
     set $t \leftarrow t + \tau$.
   - Else, the next reaction R_D is the proliferation reaction of the cell for which Θ_i is minimal for $i \in \{1, \ldots, N_P\}$,
     set $t \leftarrow t_{R_D}$

6. **Decision step:** execute the selected reaction if possible. Reject the reaction if it is not possible. Update data structure.
7. **Halt if $H = True$ else continue the process with Step (2).**

6.4 Results

6.4.1 Typical dynamics and snapshots

In this section, we present the typical behavior of the above models. The experiments shown in the figures were performed on a 100×100 grid ($K = 10^4$), with $N(0) = M(0) = 500$ initial moving cells. In each experiment the mobility rate was unity, $r_M = 1$.

Each figure illustrates the behavior of the corresponding model with two simulations. The top rows correspond to a simulation where the proliferation rate is $r_P = 2$ and the cell cycle length is $\vartheta = 0.05$. In the bottom row, the proliferation rate is $r_P = 100$ and the cell cycle length is $\vartheta = 2$.

The inverse r_P^{-1} of the proliferation rate is the expected value of the time a cell in a sparse population spends in mobile mode. Thus, in each experiment, the cells corresponding to the top row corresponds are expected to spend $r_P^{-1} = 0.5$ time on average in mobile mode, while those in the second row spend $r_P^{-1} = 0.01$ time moving on the lattice before attempting to proliferate. This suggests that the upper rows represent the dynamics of cells that spend most of their life in a mobile state, while the simulations in the lower rows correspond to cells that are primarily proliferating.

The graphs in the center of each row show the temporal dynamics from $t = 0$ until time t for which $N(t) = 0.98K$. The snapshots on the left were taken when the total number of cells reached 2500, and the snapshots on the right corresponds to the instance when the number of cells reached 7500.

The graphs in the top row show the usual logistic-like population growth, on the one in the bottom rows we may observe the synchronization of cells, that we saw in the previous chapter.

Model 1 In this model, cells switch to the proliferating phenotype and reserve a site in a contact-inhibited reaction. Thus, most of the proliferating cells appear on the boundary of the patches (c.f. bottom row of Fig. 6.1) and the number of proliferating cells matches the number of of reserved sites at every instance. In case of large proliferation rate, we may observe a proliferating front on the contour of the patches.

As the transition to the proliferating mode is contact inhibited, the experiments suggest that all cells would be of the motile type after sufficiently long time. In fact, this observation will be proved in the next chapter. This is the only model of the four where the large proliferation rate coupled with a long cell cycle does not lead to

longer time for the cells to populate the lattice, and this is due to the contact-inhibited phenotype switching.

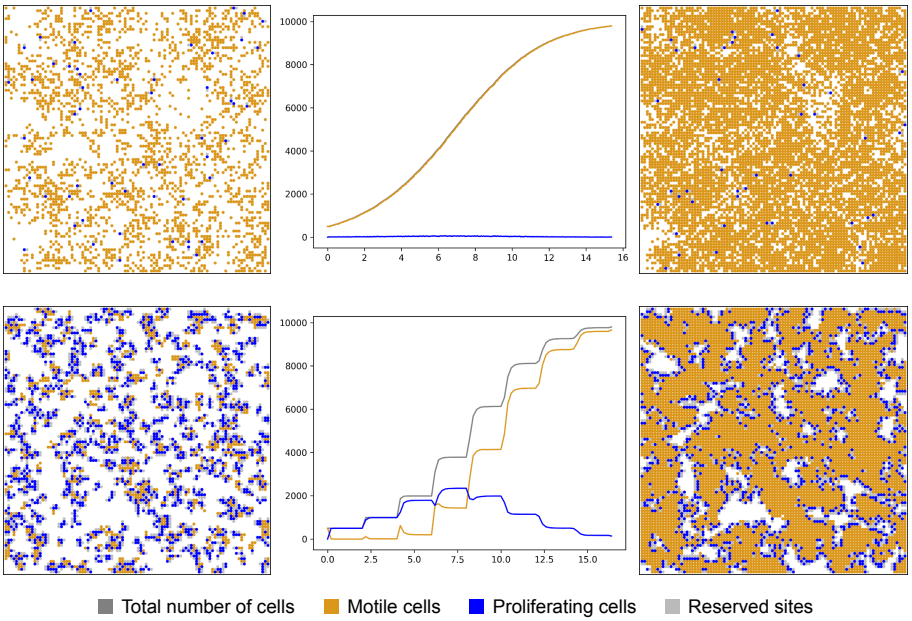

Figure 6.1: M1

Model 2 In this model, cells change to the proliferating phenotype in a spontaneous reaction without any constrains about their neighborhood. In case the proliferation rate is large (bottom row of Fig. 6.2), we find most of the cells in proliferating state. Thus, almost all cells are blue in the bottom row and this corresponds with the dynamics in the middle of the bottom row. On the contrary, in case of small proliferation rate we may find almost all cells in mobile mode (see the dynamics in the first row).

Remarkably, this does not necessarily correlate positively with the growth speed of the population as the length of the cell cycle is also limiting. In Model 2, in case of large proliferation rate and short cell cycle, the population needs more time to reach 98% of the capacity of the lattice, compared to small proliferation rate and long cell cycle. We may observe the same in case of Models 3 and 4 as well. The graph in the bottom row of Fig. 6.2) clearly shows the step-like growth of the population and also shows that the "length of the stairs" coincides with the length of the cell cycle.

Model 3 In case of this model, the cells also switch to the proliferating type in a spontaneous way, but unlike in M2, the reversion is contact inhibited. If it is successful (the mother cell finds a site to place the daughter cell into) the mother cell reverts. However, in case it fails, the mother cell turns to a waiting type cell (red dots on Fig. 6.3), that also attempts to revert in a contact inhibited way. Because of this, we may observe red patches bordered by blue contours. Inside the patches there are cells frozen

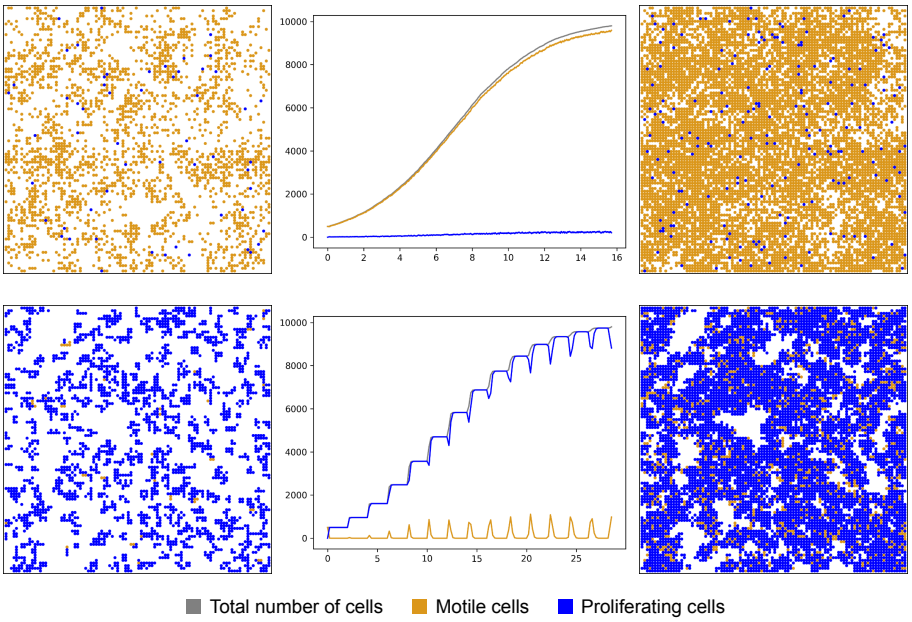

Figure 6.2: Model 2

in the waiting state because they have no place to divide into. The graphs of both experiments suggest that at the end all the cells will be of waiting type. The graph in the second row of Fig. 6.3 clearly shows that the horizontal part of the growth curve $N(t)$ coincides with the length of the cell cycle.

Model 4 In this approach, the phenotype switching is still spontaneous, but unlike in the previous cases, the cell only begins proliferation if there is place for the daughter cell. Otherwise, the cell turns waiting type and attempts to reserve a site in contact inhibited reaction.

The dynamics (Fig. 6.4) are very similar to the dynamics of M3, we may expect that all the cells turn waiting type after a sufficiently long time. In case of the bottom row we may observe that the red patches are surrounded by blue and gray cells, where the gray ones represent the reserved sites.

6.5 The role of the model hypotheses

In this section, we attempt to find measurable quantities that may help us distinguish the models. First, we investigate which domain of parameters (ϑ, r_P^{-1}) is suitable for the synchronization of the population. As it will turn out, this measure is unfortunately not informative. However, the average growth velocity of the population may be useful to this purpose.

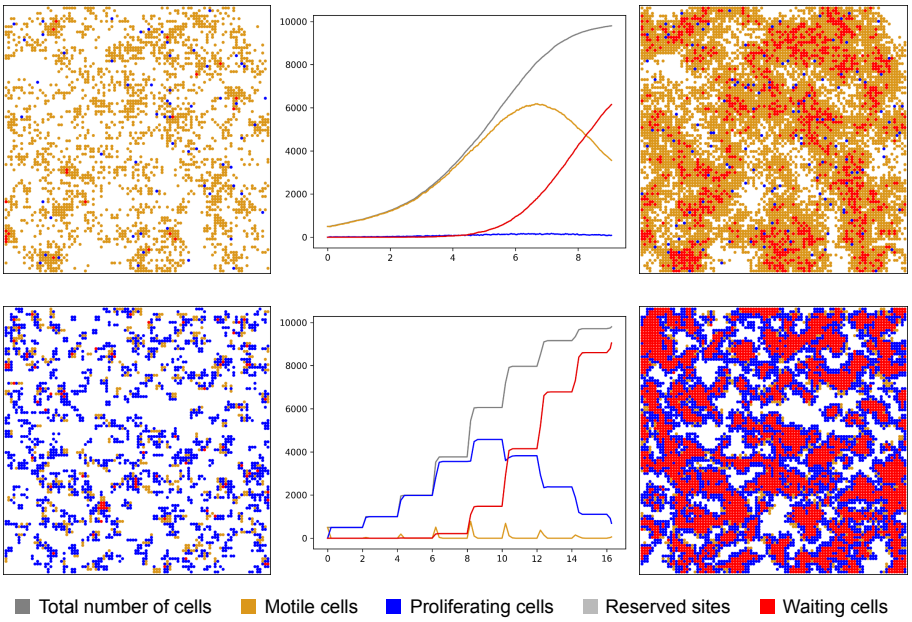

Figure 6.3: M3

6.5.1 Synchronicity landscape

Synchronization may be a hidden behavior of real cell populations and classical population models cannot predict it [46]. From Figures 6.1-6.4, it is clear that our four lattice-based models predict that cells may remain synchronous.

In this section, we discuss the parameter configuration for which synchronicity may be preserved. A large number of simulations were run on a biologically feasible part of the parameter space with an initial number of $N(0) = M(0) = 500$ randomly seeded motile cells on a 100×100 lattice, with capacity $K = 10^4$.

The motility rate and the activation rate are fixed in every simulation: $r_M = 1$ and $r_W = 400$. We sampled the (ϑ, r_P^{-1}) parameter space as follows. Including the endpoints, 64 points are chosen from the interval $[0.01, 0.5]$ so that the distances between consecutive points are the same, let J denote the set of these points. Then we ran and averaged out 5 simulations for all pairs (ϑ, r_P^{-1}) from

$$\left\{(\vartheta, r_P^{-1}) : \vartheta \in \{0.05, 0.1, ..., 2\}, r_P^{-1} \in J\right\}.$$

The separation of logistic and step-like growth is a rather challenging task, since the change of the parameters results a continuous shift from one type of growth to the other. Nevertheless, the increase of the total number of cells ('the size of the steps') between $[\vartheta, \vartheta + \frac{\vartheta}{10}]$ seems to be an effective indicator. We measured these increases from $t = 0$ to the time the population reaches $K/2$. Above a 30% increase in the number of cells between $[\vartheta, \vartheta + \frac{\vartheta}{10}]$, 150 cells in our experiments, we consider the cells to be synchronized. Below 15% (75 cells) we consider the population asynchronous. Between

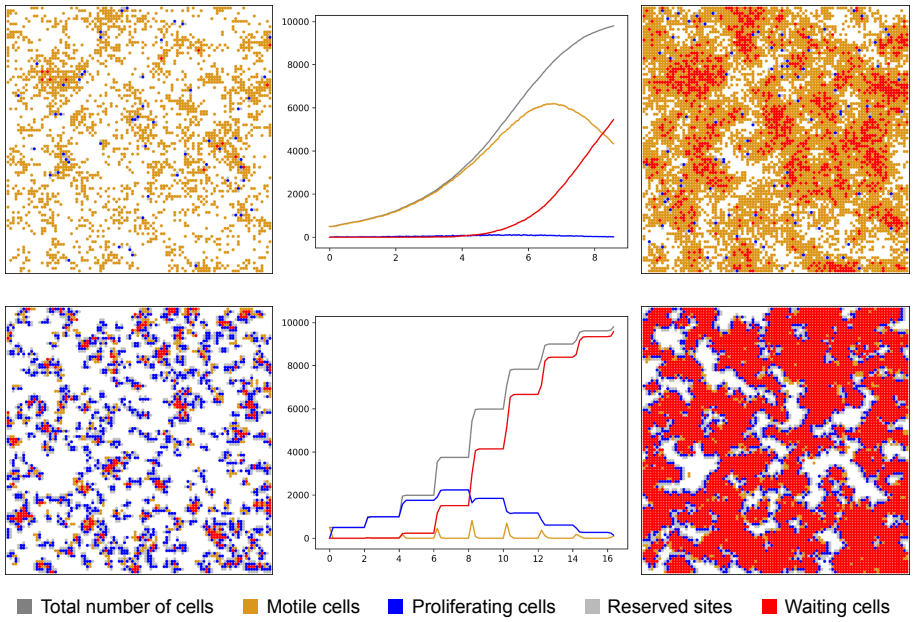

Figure 6.4: M4

these two values we consider the growth mixed: after a short transient synchronization with steep ascents, the growth curves become logistic-like. It is worth noting that we experimented with several other characterizations, but the output was always roughly the same.

On Figure 6.5 we plotted the output of the classification. One rectangle with coordinates (ϑ, r_P^{-1}) represents the averaged output of 5 simulations. The domain of the parameter space which leads asynchronous population is colored burgundy, the domain where synchronicity is preserved is gray, the middle white domain represents transient synchronicity. The characterization of Model 1 can be seen in the top row of Figure 6.5 and the characterization of Model 2, 3 and 4 is shown in the bottom row. The plots look quite similar, making it impossible to tell the models apart based on these outputs.

6.5.2 Average growth velocity

The average speed of the population growth may be measured by the time the population needs to reach a certain number of cells. We choose this number 98% of the capacity (9800 cells). From the results of the simulations described in section 6.5.1 we plotted these time values as heatmaps on Fig. 6.6 for our four models. Some contour lines are shown on every heatmap for orientation. In general, we may note the not so surprising fact that increasing the length of the cell cycle and decreasing the proliferation rate slows down the growth of the population in each model. Models 3 and 4 both show a similar behavior: for a given pair of parameters (ϑ, r_P^{-1}) we expect our virtual populations to reach $N = 0.98K$ in more or less the same amount of time. However,

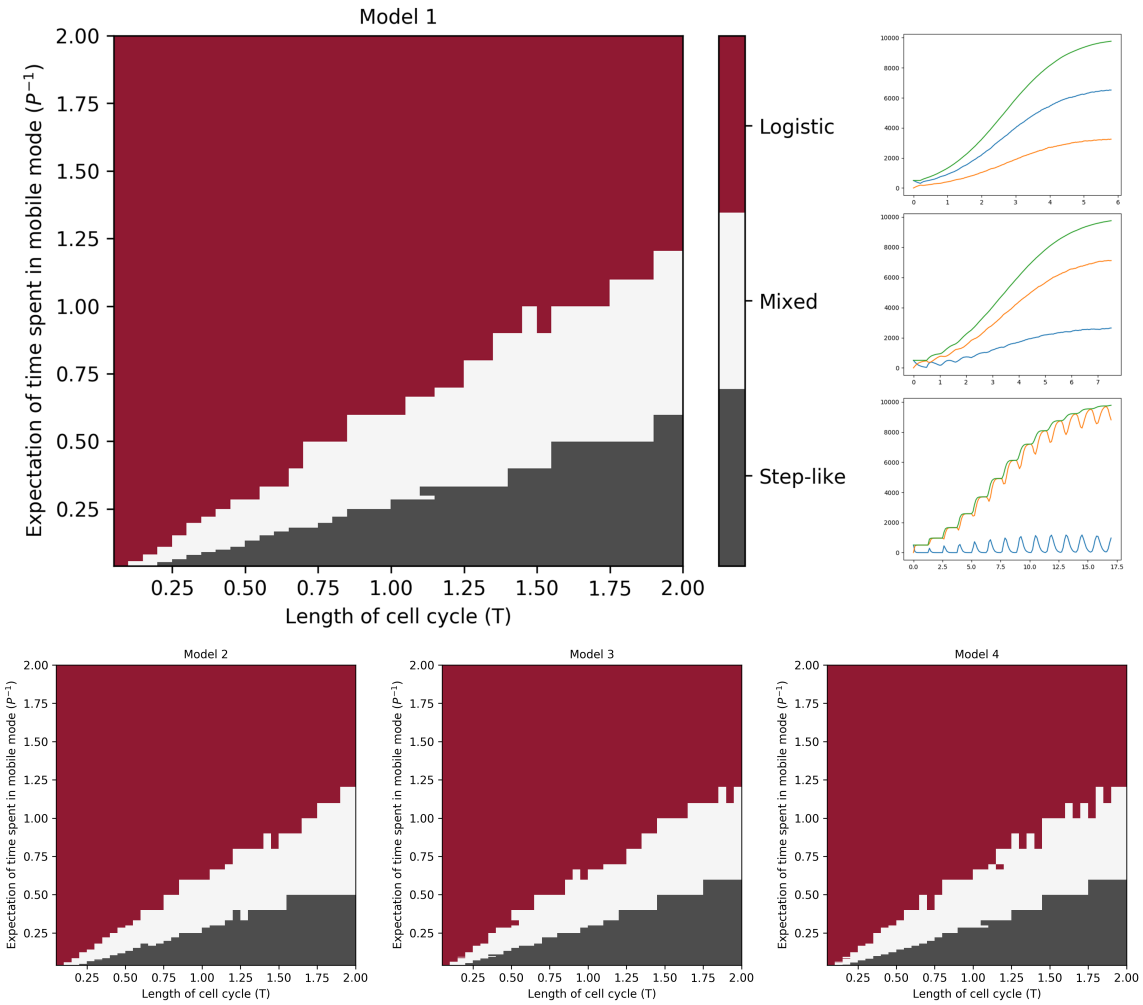

Figure 6.5: Synchronicity landscape. The top left figure shows the domain of the parameter space (ϑ, r_P^{-1}) for which the cell cultures synchronize in case of Model 1. The bottom row shows the same in case of Model 2, 3, and 4 in order on the same parameter space. See the legend on the middle for color code. The right column in the top row shows the averaged dynamics of 5 stochastic realizations for parameters $(\tau, p) \in \{(1.25, 15), (0.5, 5.5), (0.2, 2.5)\}$, from top to bottom. The colors on the time series plots are: orange - $M(t)$, blue - $P(t)$, gray - $N(t)$. The legend applies to all figures.

in the case of Model 2, we can observe a broader time scale as the corresponding time values can reach up to 34.

The heatmap of Model 2 suggests non-monotonic behavior of the average growth velocity as some contour lines seem to fold back (see contour lines 20, 22 and 24 on the top right heatmap of Fig. 6.6).

On Fig. 6.7 we fixed $\vartheta = 1.65$ (see the vertical gray line on the heatmap) and plotted the average velocity values as the function of r_P^{-1}. The curve has a minima at around $r_P^{-1} \approx 6.25 \cdot 10^{-2}$ suggesting that the optimal proliferation rate to reach the maximal velocity in the population growth is $r_P \approx 16$ in for the fixed $\vartheta = 1.65$.

In other words, in the case of a fixed cell cycle length, there may be an optimal proliferation rate when the growth of the population is maximal. A plausible interpretation of this is that cells form dense local clusters reaching the local carrying capacity quickly. Thus, they are not able to reach out from these plaques to invade other parts

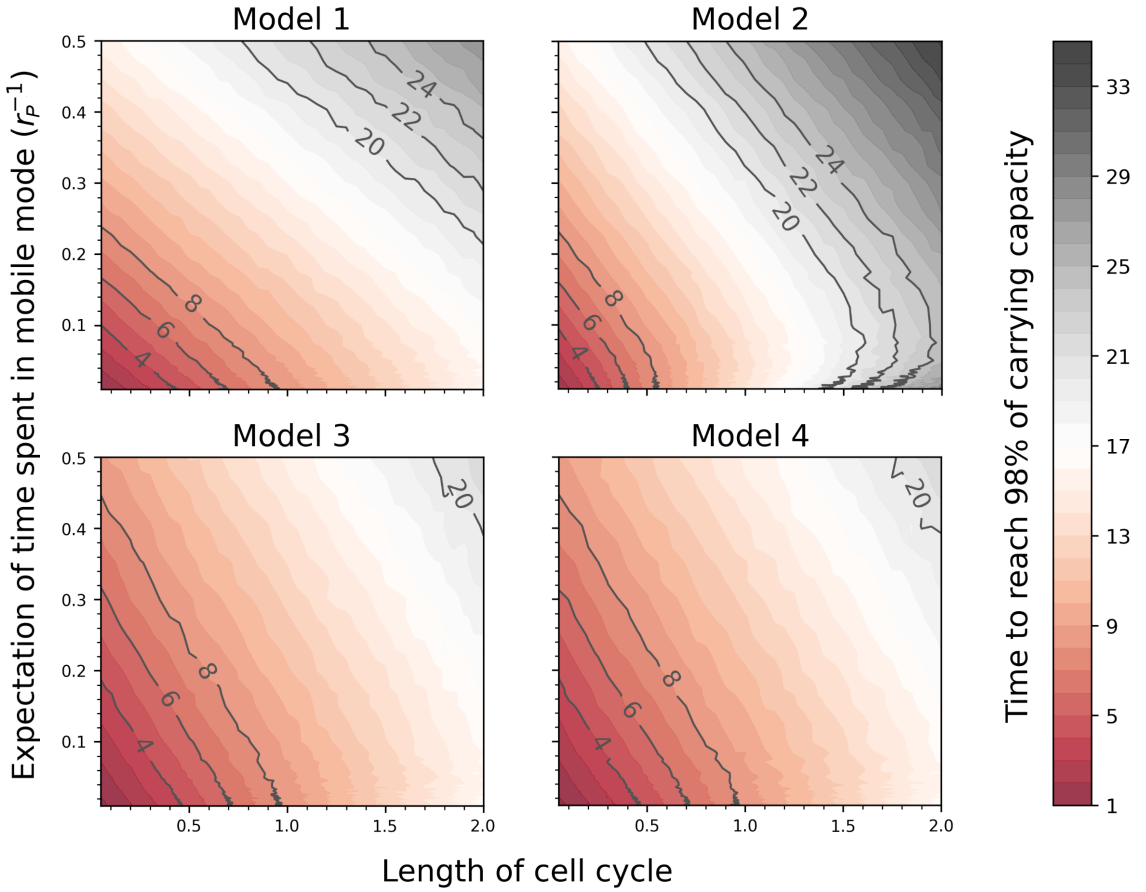

Figure 6.6: In case of every (cell cycle length, proliferation rate) parameter pair we measured the time it takes to reach 98% of the carrying capacity. These time values can be seen on the heatmaps of this figure for our four models: M1 and M2 in the top row, M3 and M4 in the bottom row.

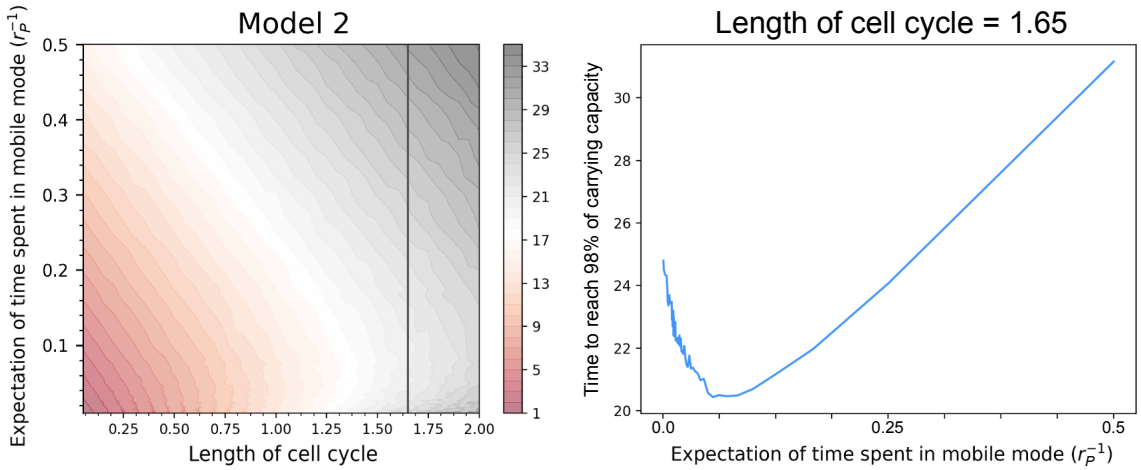

Figure 6.7: Figure shows the non-monotonic behavior of Model 2. Decrease in the proliferation rate at a fixed cell cycle length increases the time it takes to reach 98% of the carrying capacity.

of the lattice. Thus, unlike in the Markovian approaches [6, 35], in these delay models higher motility rate rate does not necessarily imply a reduction in local correlations.

This may imply surprising therapeutic consequences. In certain parameter config-

urations one possible intervention is to increase proliferation rate of the cells (equivalently decrease r_P^{-1}) to slow down the growth of the cell culture. Conversely, chemotherapeutic drugs that aim to suppress proliferation of cancer cells may have a harmful effect in case of unfortunate parameter combinations.

6.6 Discussion

In this section we showed that the time delay, caused by the duration of the cell cycle, raises new modeling questions in lattice-based models. This delay breaks the Markovian property: it is not always guaranteed that cells would be able to complete the initiated cell division, as their adjacent sites may turn occupied by the end of the cycle.

We proposed four new models to tackle this problem. Although all of them are based on very different assumptions, they show similar dynamics. Remarkably, the parameter domain for which synchronicity is preserved, is almost the same in all of them (see Fig. 6.5), suggesting that this family of models is robust for this preserving property. Thus, when only synchronicity is concerned, a wrong model selection due to our possibly insufficient knowledge about the system does not seem to elicit problematic behavior.

This may have serious therapeutic implications. Most of antitumor chemotherapeutics work by impairing a single cellular function that occurs during a well defined and possibly short time interval in the cell cycle.

For example, cell division can be restricted by inhibiting the mitotic spindle. While in the G1 phase, topoisomerase enzyme inhibitors and alkylating agents restrict the preparations to create copy DNA. In phase S, the formation of functional DNA can be inhibited by administering antimetabolites, by inhibiting the formation of intrinsic nucleotides, or by inhibiting the DNA polymerase.

Instinctively, when planning antitumor chemotherapy, the main goal is to administer these drugs in the lowest possible effective dose, in order to avoid off-target effects on other affected cell types in the body. The synchronicity landscape showed in Sec. 6.5.1 (Fig 6.8) may help in planning such therapeutic regimes.

A possible therapeutic approach would be the synchronization of the cell cycle in the cell population. Assume that the system with parameters $(r_{P_1}^{-1}, \vartheta_1)$ does not preserve synchronicity. A possible application of certain compounds may change the parameters to a new pair $(r_{P_2}^{-1}, \vartheta_2)$, which may enable the system to sustain the synchronicity of the population.

It must be noted that the synchronicity suggested here would not happen in a spontaneous manner, e. g., due to the self-synchronization of the population. Instead, a set of favorable circumstances is needed, since the model does not account for any coupling between the cells. Furthermore, cells may not be synchronous at the beginning.

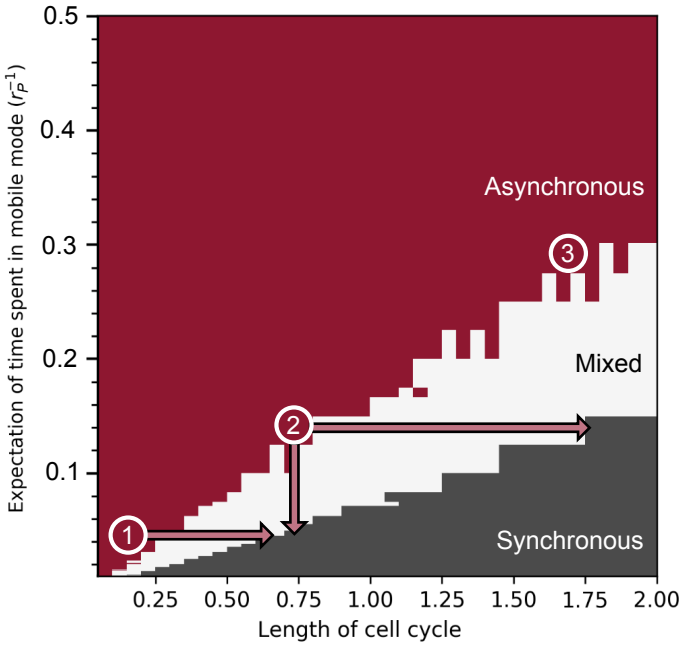

Figure 6.8: The synchronicity landscape may help us to plan tumor therapies which involves manipulation of the cell cycle length or proliferation rate of cells.

Thus, the population must be synchronized first in order to maximally exploit the cell cylce-dependent drug effects.

Therefore, an essential issue arises: a viable and safe procedure is needed to reach the domain in the parameter plane where the system can preserve synchronicity. Fig 6.8 illustrates three possibilities.

In the case of the type 1 hypothetical tumor cell, characterized by a fast proliferation rate and a short cell cycle, prolonging the cell cycle may be a viable approach. On the figure, this would mean a rightward shift. The type 2 hypothetical tumor cell is characterized by moderate proliferation rate and moderate cell cycle length. This leads to a difficult problem to solve, since either the cell cycle or the proliferation rate needs to be increased in order to achieve synchronization. On one hand, the increased proliferation rate may likely pose an extensive risk in the case of anti-tumor therapy. On the other hand, even though the prolonged cell cycle would not pose such a high risk, the target length of the cell cycle is rather far away from the baseline value, suggesting that achieving this target may not be feasible. The type 3 hypothetical tumor cell population would be the least favorable example, as the parameter domain that would enable synchronization of its population is too far away from the baseline parameters that characterize this cell line, therefore the transition would be very costly and possibly even risky to achieve.

Therefore, if an actual cell line can be modeled using its corresponding, experimentally verified parameters, our model could provide valuable information regarding the optimization of possible therapeutic approaches in antitumor chemotherapy.

7

A mean field model

7.1 Convergence of solutions in a mean-field model of go-or-grow type with reservation of sites for proliferation and cell cycle delay

Cell proliferation and motility are key processes that govern cancer invasion or wound healing. The go-or-grow hypothesis postulates that proliferation and migration spatiotemporally exclude each other. This has been acknowledged, for example, for glioblastoma [26]. In general, two phenotypes that can be of particular importance to progression of aggressive cancers are 'high proliferation-low migration' and 'low proliferation-high migration', and the mechanisms governing this switching are of great interest in current medical research [39]. Here we consider a strong simplification of this phenomenon by assuming that (differently from [24]) motile cells stop for a fixed period of time to complete cell division, upon which they immediately switch back into the migratory phenotype. We study the mathematical properties of a mean-field approximation of an individual based model describing this process, and this note complements our other ongoing works [4] where we investigate in detail a range of biological hypotheses with the corresponding individual-based as well as mean-field, analytically tractable, models.

7.2 The model

Assume that agents (representing biological cells) move and proliferate on an n - dimensional square lattice with length ℓ (in each direction), so that $K = \ell^n$ is an integer describing the number of lattice sites. We divide our agent population into two subpopulations, motile and proliferative, with the condition that a proliferative agent has to be attached to an adjacent site which is reserved until the end of proliferation. As a result, sites can either contain a motile agent, a proliferating agent, be reserved for the daughter agent of an attached proliferative agent, or be empty. At each time step, each motile agent can attempt to move into an adjacent lattice site or proliferate at

its current site. However, if a motile agent attempts to move into a site that is already occupied or reserved, the movement event is aborted. Similarly, if a motile agent attempts to begin proliferation by reserving a site that is already occupied, then the proliferation event is aborted. Agents attempt to convert from being motile to proliferative at constant rate r, and the proliferative phase has length ϑ, upon which two motile daughter agents appear, one on the proliferating site, and one on the reserved site.

Based on the above, tracking the rate of change of the number of motile agents, $m(t)$, proliferative agents, $p(t)$, and reserved sites, $q(t)$, in time, and following the arguments and derivation in [6], we obtain the following mean-field approximation:

$$m'(t) = -rm(t)\frac{K-m(t)-p(t)-q(t)}{K} + 2rm(t-\vartheta)\frac{K-m(t-\vartheta)-p(t-\vartheta)-q(t-\vartheta)}{K},$$

$$p'(t) = rm(t)\frac{K-m(t)-p(t)-q(t)}{K} - rm(t-\vartheta)\frac{K-m(t-\vartheta)-p(t-\vartheta)-q(t-\vartheta)}{K},$$

$$q'(t) = rm(t)\frac{K-m(t)-p(t)-q(t)}{K} - rm(t-\vartheta)\frac{K-m(t-\vartheta)-p(t-\vartheta)-q(t-\vartheta)}{K}.$$

Here, the first term in the m equation expresses that m-cells attempt to proliferate with rate r, but proliferation starts only if the randomly selected target site is empty at time t, which has probability $(K-m(t)-p(t)-q(t))/K$, if we assume that there is no spatial correlation. The other terms can be interpreted in a similar way. Using the variable $u = K-m-p-q$ that accounts for empty sites, we can write

$$\begin{aligned}
m'(t) &= -rK^{-1}m(t)u(t) + 2rK^{-1}m(t-\vartheta)u(t-\vartheta), && (7.1)\\
p'(t) &= rK^{-1}m(t)u(t) - rK^{-1}m(t-\vartheta)u(t-\vartheta), && (7.2)\\
q'(t) &= rK^{-1}m(t)u(t) - rK^{-1}m(t-\vartheta)u(t-\vartheta), && (7.3)\\
u'(t) &= -rK^{-1}m(t)u(t). && (7.4)
\end{aligned}$$

7.3 Long-term behaviour

The usual phase space for Eqs. (7.1)-(7.4) is $C = C([-\vartheta, 0], R^4)$, the Banach space of continuous function from the interval $[-\vartheta, 0]$ to R^4 equipped with the supremum norm. With the notation $x(t) = (m(t), p(t), q(t), u(t))$, our system is of the form $x'(t) = f(x_t)$ where $x_t \in C$ is defined by the relation $x_t(\theta) = x(t+\theta)$ for $\theta \in [-\vartheta, 0]$ and $f : C \to R^4$ is defined by the right-hand side of Eqs. (7.1)-(7.4). The standard results for delay differential equations provide existence and uniqueness of solutions from initial data $x_0 = \phi \in C$ (see, for example [33]).

Given the biological motivation, we are interested only in non-negative solutions,

for which $p(t) = q(t) = rK^{-1}\int_{t-\vartheta}^{t} m(s)u(s)\mathrm{d}s$ holds, meaning that proliferative cells at a given time t are exactly those who started the proliferation process in the time interval $[t-\vartheta, t]$, and the reserved sites correspond to them. With this compatibility condition and the balance law $K = m(t) + p(t) + q(t) + u(t)$, we define the feasible phase space

$$\Omega := \left\{ \phi \in C : \phi_j(\theta) \geq 0 \text{ for all } \theta \in [-\vartheta, 0], j = 1,2,3,4; \right.$$
$$\left. \sum_{j=1}^{4} \phi_j(0) = K; \quad \phi_2(0) = \phi_3(0) = rK^{-1}\int_{-\vartheta}^{0} \phi_1(s)\phi_4(s)\mathrm{d}s \right\}. \quad (7.5)$$

Lemma 7.1. *The set Ω is forward invariant, that is for any solution $x(t)$ with $x_0 \in \Omega$, $x_t \in \Omega$ for all $t \geq 0$.*

Proof. Integrate Eq. (7.2) from 0 to t to obtain (similarly for $q(t)$)

$$p(t) - p(0) = rK^{-1}\int_{t-\vartheta}^{t} m(s)u(s)\mathrm{d}s - rK^{-1}\int_{-\vartheta}^{0} m(s)u(s)\mathrm{d}s.$$

From $x_0 \in \Omega$ we have $p(0) = q(0) = rK^{-1}\int_{-\vartheta}^{0} m(s)u(s)\mathrm{d}s$, hence

$$p(t) = q(t) = rK^{-1}\int_{t-\vartheta}^{t} m(s)u(s)\mathrm{d}s, \quad (7.6)$$

thus the third condition in the definiton of Ω is preserved. The second trivially follows from summing up the equations to see $(m(t) + p(t) + q(t) + u(t))' = 0$, so $K = m(t) + p(t) + q(t) + u(t)$ is preserved. To confirm nonnegativity, note that $u(t) = u(0)\exp(-rK^{-1}\int_0^t m(s)\mathrm{d}s) \geq 0$. Assuming that $m(t) \geq 0$ for $t \leq t_0$, we have $m(t-\vartheta)u(t-\vartheta) \geq 0$ for $t \leq t_0+\vartheta$, and consequently $m(t) \geq m(t_0)\exp(-rK^{-1}\int_{t_0}^t u(s)\mathrm{d}s)$ holds on $[t_0, t_0+\vartheta]$. Hence, by the method of steps we obtain non-negativity of $m(t)$ for all t. Then the non-negativity of $p(t)$ and $q(t)$ follow from Eq. (7.6). ■

Note that since solutions starting from Ω stay in this bounded set, they exist globally. Following [33], we say that a continuous functional $V : C \to R$ is a Lyapunov functional on the set Ω in C for Eqs. (7.1)-(7.4), if it is continuous on the closure of Ω, and $\dot{V} \leq 0$ on Ω. Here, $\dot{V}$ denotes the derivative of V along solutions. In our case Ω is itself closed. We also define $E := \{\phi \in \Omega : \dot{V} = 0\}$ and $M :=$ the largest set in E which is invariant with respect to Eqs. (7.1)-(7.4)).

Theorem 7.2. *If $m(0) > 0$, then $\lim_{t\to\infty}(m(t), p(t), q(t), u(t)) = (K, 0, 0, 0)$.*

Proof. Consider the functional $V(\phi) = \phi_4(0)$. Then $\dot{V} = -rK^{-1}m(t)u(t) \leq 0$ for solutions in Ω, and by LaSalle's invariance principle (cf. Thm. 2.5.3 in [33]), the

limit set of any solution is in M, thus on the limit set of any solution, $mu \equiv 0$ holds. Since for any solution u is always zero or always positive, we have either $m \equiv 0$ or $u \equiv 0$. In both cases, $p = q \equiv 0$ follows. Hence, the limit set can only be composed of the two equilibria $(K, 0, 0, 0)$ or $(0, 0, 0, K)$. Finally, we show that if $m(0) > 0$, then $m(t)$ can not converge to 0. Since $u(t)$ is monotone decreasing, for such a solution $m(t) + p(t) + q(t) = K - u(t) \geq K - u(0) > 0$ should hold. If $m(t) \to 0$ as $t \to \infty$, then from Eq. (7.6), also $p(t) = q(t) \to 0$. This contradicts $m(t) + p(t) + q(t) \geq K - u(0) > 0$ and so we can exclude $(0, 0, 0, K)$ from the limit set. Therefore $\lim_{t\to\infty}(m(t), p(t), q(t), u(t)) = (K, 0, 0, 0)$. ■

Remark: if $m(0) = 0$, then also $p(0) = q(0) = 0$, so $u(0) = K$ and we are on the empty lattice having the trivial solution $(0, 0, 0, K)$.

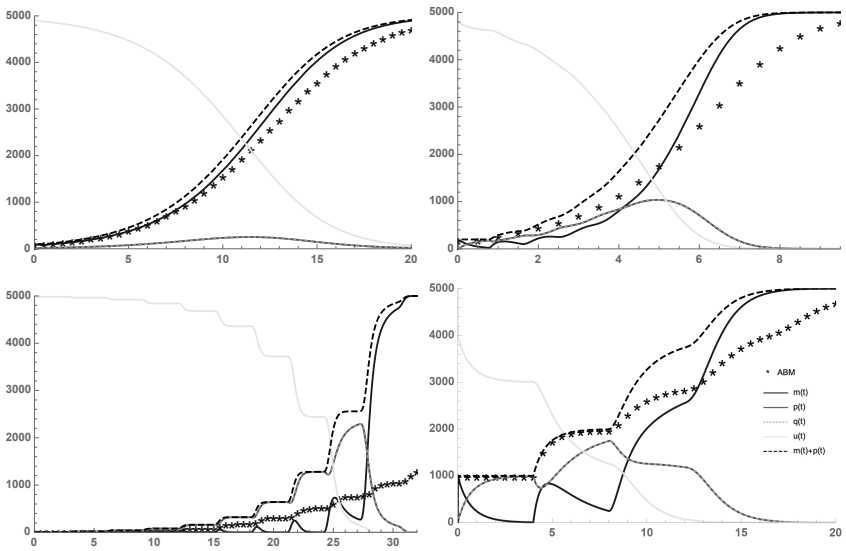

Figure 7.1: Four numerical simulations, where in each realization, the initial function ϕ_1 on $[-\vartheta, 0]$ is given by $aH(\theta)$, where $a > 0$ is the initial number of cells, $\phi_2 = \phi_3 = 0$ and $\phi_4 = K - aH(\theta)$. This choice of the initial function models an *in vitro* experiment where motile cells are added to the plate at $t = 0$. The parameters are the following: *Top Left* – $r = 0.5, a = 100, \vartheta = 0.5$; *Top Right* – $r = 3, a = 200, \vartheta = 0.8$; *Bottom Left* – $r = 10.5, a = 10, \vartheta = 3$; *Bottom Right* – $r = 2, a = 1000, \vartheta = 4$; and $K = 5000$ in each case. The mean-field equations are compared to the output of averaged stochastic simulations of a corresponding agent-based model (ABM) on a 50×100 square lattice with cell motility rate 2. The legend in the bottom right figure applies to each.

7.4 Simulations and conclusion

According to the choice of the initial functions, different *in vitro* experiments can be modelled with Eqs. (7.1)-(7.4). One approach is to add a number of motile cells all at once at $t = 0$ to the empty cell space (e.g. a Petri dish). In this experiment the initial function ϕ_1 is given by $\phi_1(\theta) = aH(\theta)$ for $\theta \in [-\vartheta, 0]$, where a stands for the number of introduced cells at $t = 0$, and $H(\theta)$ is the right-continuous Heaviside-function, i.e. $H(\theta) = 0$ for $\theta < 0$ and $H(\theta) = 1$ for $\theta \geq 0$. In this setting, we take $\phi_2(\theta) = \phi_3(\theta) = 0$ and $\phi_4(\theta) = K - aH(\theta)$. While such initial data is not from C, they satisfy Eq. (7.5) and generate a continuous solution for $t > 0$. Some of such simulations are shown in Figure 7.1.

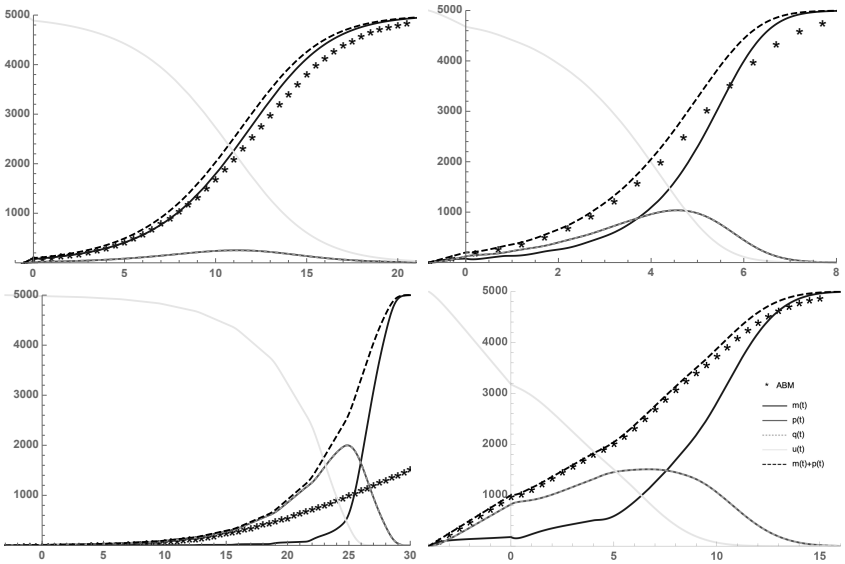

Figure 7.2: Four simulations, where in each realization motile cells are added with rate a in the initial interval. The parameters are the following: *Top Left* – $r = 0.5, a = 200, \vartheta = 0.5$; *Top Right* – $r = 3, a = 250, \vartheta = 0.8$; *Bottom Left* – $r = 10.5, a = 3.3, \vartheta = 3$; *Bottom Right* – $r = 2, a = 250, \vartheta = 4$ and $K = 5000$ in each case. The ABM is the same as in Figure 1, but the initial cell distribution has also been simulated here. The legend in the bottom right figure applies to each.

A more elaborate *in vitro* experiment is the following. Instead of adding motile cells all at once, we add them in to the assay with a constant rate a for a time interval of length ϑ. After this, we leave the cell population intact. The initial data corresponding to this experiment can be obtained by solving a modification of Eqs. (7.1)-(7.4) with an additive forcing term $+a$ to the m-equation (and $-a$ to the u-equation), representing the gradual addition of m-cells, on an interval of length ϑ, starting from the state $(0, 0, 0, K)$. Then we start solutions of Eqs. (7.1)-(7.4) with such initial functions, which satisfy Eq. (7.5). Four realizations of this experimental setting are shown in Figure 7.2.

The point of considering these two setups is that in the first we have only motile cells at $t = 0$, while in the second at $t = 0$ we have a distribution of cells in different phases of the cell cycle. This has a profound impact on the behaviour of solutions. For the sake of easier comparison, in both experimental settings the number of cells at $t = 0$ is exactly the same in the corresponding simulations whenever the parameter r and ϑ are the same. For both experimental settings, we can observe that the smaller the proliferation rate, the better the agreement between the mean-field model and the output of the agent based model (ABM). This is intuitively clear, as for smaller proliferation rate the cells in the ABM has more time to move around between proliferation events, hence the cell population becomes more well-mixed. While in Section 3 we proved that all solutions settle eventually at the state $(K, 0, 0, 0)$, there are distinctive features of solutions in different scenarios. Figure 1 shows that when the cell cycle delay is small, the solutions resemble logistic growth. In contrast, when the delay is large relative to the average time between individual cells attempting enter the proliferative

state, the initially motile cells enter the proliferative state more or less together, and hence complete cell division more or less together too, resulting in a step-function-style growth curve in the total cell count. The sudden switching between phenotypes causes non-monotonic behaviours in $m(t)$ and $p(t)$ also. When we add motile cells continuously rather than adding them all at once, the solutions are much more similar to the expected logistic growth curve, and a different characteristic can be observed only for high proliferation rates or large numbers of initially added cells. In conclusion, an intermittent growth of a cell population can be an indication that the cell cycle length is relatively large (relative to inter-proliferation times), while its variance is small.

8

Summary

The advantage of stochastic simulations is that they can be used to obtain information not only on the time course of processes, but also on the time evolution of the distribution of state variables. These methods can even be used when standard mathematical tools become intractable due to the high complexity of the system. In this thesis, we have introduced several algorithms that allow exact stochastic simulation of cell populations.

In Chapter 3, Preliminaries, we introduced a novel approach to incorporate possible cellular reactions into stochastic models of cells. To the best of our knowledge, these reaction types have not been taken into account in lattice-based stochastic simulations of cell cultures. These new types of reactions may be easily applied to complicated systems, enabling the generation of biologically feasible stochastic cell culture simulations.

- We treat the reactions systematically and categorized them according to how they depend on the local environment of the cells.
- We introduced contact-inhibited, contact-promoted, and spontaneous reactions.

In Chapter 4 we introduced one of the main results of the thesis.

- We significantly extended the lattice based cellular simulation algorithm by Baker and Simpson [6] with the reactions we defined in Chapter 3. We call this algorithm the Prompt Decision Method (PDM).
- Then we proposed and derived from first principles our exact method, the Reduced Rate Method (RRM).
- We proposed the equivalent but simplified version of the RRM algorithm, the marginal Reduced Rate Method.
- We showed that the PDM algorithm and our RRM algorithm are mathematically equivalent.
- We showed that it is substantially quicker than the PDM for large number of

cells.

In Chapter 5 we incorporated realistic cell cycle length to our models.

- We presented a new model and a new algorithm that is able to model the evolution of cell populations with realistic cell cycle length of any distribution.
- We also fit our model to *in vitro* experimental data measured by Vitadello et al. [46].

In Chapter 6, we put together the tools we have developed to propose the Go or Grow algorithms.

- We proposed four new models and algorithms to overcome the difficulties that cell cycle causes in a spatial population model.
- We showed the surprising effects the cell cycle takes on the dynamics and highlighted the possible therapeutic implications.

Finally, in Chapter 7, we introduced the mean field delay differential equation that is obtained from the stochastic cell cycle model we introduced in Chapter 5.

- We investigated the model with analytical methods and compared the model predictions with the predictions of the stochastic model.

The dissertation is based on the following two published papers

- R. E. Baker, P. Boldog, and G. Röst,
  Convergence of solutions in a mean- field model of go-or-grow type with reservation of sites for proliferation and cell cycle delay;
  Progress in Industrial Mathematics at ECMI 2018 (Cham, 2019), I. Faragó, F. Izsák, and P. L. Simon, Eds., Springer International Publishing, pp. 381–387.
- P. Boldog, N. Bogya, and Z. Vizi,
  Propensity matrix method for age dependent stochastic infectious disease models *Trends in Biomathematics: Stability and Oscillations in Environmental, Social, and Biological Models*, R. P. Mondaini (ed.), Springer, https://doi.org/10.1007/978-3-031-12515-7_17 (2021);

and the preprint (submitted to a journal and under review)

- P. Boldog,
  Exact lattice-based stochastic cell culture simulation algorithms incorporating spontaneous and contact-dependent reactions.
  https://arxiv.org/pdf/2208.04774.pdf

Further publications of the author:

- P. Boldog, T. Tekeli, Z. Vizi, A. Dénes, F. A. Bartha, G. Röst,

Risk Assessment of Novel Coronavirus COVID-19 Outbreaks Outside China
Journal of Clinical Medicine, 9, 2, 571 (2020), MDPI, https://www.mdpi.com/2077-0383/9/2/571

- G. Röst, F.A. Bartha, N. Bogya, P. Boldog, A. Dénes, T. Ferenci, K.J. Horváth, A. Juhász, C. Nagy, T. Tekeli, Z. Vizi, B. Oroszi
  Early Phase of the COVID-19 Outbreak in Hungary and Post-Lockdown Scenarios, *Viruses*. 2020; 12(7):708. https://doi.org/10.3390/v12070708
- P. Boldog, K. Hajdu, M. Magyar, É. Hideg, K. Hernádi, E. Horváth, A. Magrez, K. Nagy, G. Váró, L. Forró and L. Nagy (2013), Carbon nanotubes quench singlet oxygen generated by photosynthetic reaction centers. *Phys. Status Solidi B*, 250: 2539-2543. https://doi.org/10.1002/pssb.201300074
- F.A. Bartha, P. Boldog, A. Dénes, T. Tekeli, Z. Vizi, G. Röst, Potential severity, mitigation, and control of Omicron waves depending on pre-existing immunity and immune evasion *Trends in Biomathematics: Stability and Oscillations in Environmental, Social, and Biological Models*, Springer R. P. Mondaini (ed.), https://doi.org/10.1007/978-3-031-12515-7_22
- D. Meszéna, A. Barlay, P. Boldog, D. Cserpán, L. Wittner, I. Ulbert, and Z. Somogyvári, Reconstruction of spatio-temporal membrane potential and resistive current distribution from paired intra- and extracellular recordings
  The Journal of Physiology, Wiley (accepted)

Bibliography

[1] Allen, L. J. A primer on stochastic epidemic models: Formulation, numerical simulation, and analysis. *Infectious Disease Modelling 2*, 2 (2017), 128–142.

[2] Allen, L. J. S. *An Introduction to Mathematical Biology*. Pearson, 2006.

[3] Allen, L. J. S. *An Introduction to Stochastic Processes with Applications to Biology*. Chapman and Hall/CRC, 2010.

[4] Baker, R., and Röst, G. Global dynamics of a novel delayed logistic equation arising from cell biology. *Journal of Nonlinear Science 30* (2020), 397–418.

[5] Baker, R. E., Boldog, P., and Röst, G. Convergence of solutions in a mean-field model of go-or-grow type with reservation of sites for proliferation and cell cycle delay. In *Progress in Industrial Mathematics at ECMI 2018* (Cham, 2019), I. Faragó, F. Izsák, and P. L. Simon, Eds., Springer International Publishing, pp. 381–387.

[6] Baker, R. E., and Simpson, M. J. Correcting mean-field approximations for birth-death-movement processes. *Phys. Rev. E 82* (10 2010), 041905.

[7] Boldog, P., Bogya, N., and Vizi, Z. *Propensity matrix method for age dependent stochastic infectious disease models*. Springer, 2021.

[8] Bonabeau, E. Agent-based modeling: Methods and techniques for simulating human systems. *Proceedings of the National Academy of Sciences 99*, suppl_3 (2002), 7280–7287.

[9] Cao, Y., Gillespie, D. T., and Petzold, L. R. Avoiding negative populations in explicit poisson tau-leaping. *The Journal of Chemical Physics 123*, 5 (2005), 054104.

[10] Cao, Y., Gillespie, D. T., and Petzold, L. R. The slow-scale stochastic simulation algorithm. *The Journal of Chemical Physics 122*, 1 (2005), 014116.

[11] Cao, Y., Gillespie, D. T., and Petzold, L. R. Efficient step size selection for the tau-leaping simulation method. *The Journal of Chemical Physics 124*, 4 (2006), 044109.

[12] Cao, Y., Li, H., and Petzold, L. Efficient formulation of the stochastic simulation algorithm for chemically reacting systems. *The Journal of Chemical Physics 121*, 9 (2004), 4059–4067.

[13] Charlebois, Daniel A.; Balázsi, G. Modeling cell population dynamics. *In Silico Biology 13*, 1-2 (2019), 21 – 39.

[14] Corcoran, A., and Del Maestro, R. F. Testing the "go or grow" hypothesis in human medulloblastoma cell lines in two and three dimensions. *Neurosurgery 53*, 1 (2003), 174–185.

[15] DeLong, J. P., and Gibert, J. P. Gillespie eco-evolutionary models (gems) reveal the role of heritable trait variation in eco-evolutionary dynamics. *Ecology and Evolution 6*, 4, 935–945.

[16] Dyson, L., Maini, P. K., and Baker, R. E. Macroscopic limits of individual-based models for motile cell populations with volume exclusion. *Physical Review E 86*, 3 (2012), 031903.

[17] Erban, R., and Chapman, S. J. *Stochastic Modelling of Reaction–Diffusion Processes*. Cambridge Texts in Applied Mathematics. Cambridge University Press, 2020.

[18] Farin, A., Suzuki, S. O., Weiker, M., Goldman, J. E., Bruce, J. N., and Canoll, P. Transplanted glioma cells migrate and proliferate on host brain vasculature: A dynamic analysis. *Glia 53*, 8 (2006), 799–808.

[19] Fedoroff, N., and Fontana, W. Small numbers of big molecules. *Science 297*, 5584 (2002), 1129–1131.

[20] Fowler, A. C. Atto-foxes and other minutiae. *Bull Math Biol 83* (8 2021), 041905.

[21] Franz, B., and Erban, R. Hybrid modelling of individual movement and collective behaviour. In *Lecture Notes in Mathematics, Dispersal, Individual Movement and Spatial Ecology, A Mathematical Perspective*, M. A. Lewis, P. K. Maini, and S. V. Petrovskii, Eds. Springer Berlin, Heidelberg, 2013, pp. 129–158.

[22] Fredrickson, A., Ramkrishna, D., and Tsuchiya, H. M. Statistics and dynamics of procaryotic cell populations. *Bellman Prize in Mathematical Biosciences 1* (1967), 327–374.

[23] Garay, T., Juhász, E., Molnár, E., Eisenbauer, M., Czirók, A., Dekan, B., László, V., Hoda, M. A., Döme, B., Tímár, J., Klepetko, W.,

Berger, W., and Hegedűs, B. Cell migration or cytokinesis and proliferation? – revisiting the "go or grow" hypothesis in cancer cells in vitro. *Experimental Cell Research 319*, 20 (2013), 3094–3103.

[24] Gerlee, P., and Nelander, S. The impact of phenotypic switching on glioblastoma growth and invasion. *PLOS Computational Biology 8*, 6 (06 2012), 1–12.

[25] Gibson, M. A., and Bruck, J. Efficient exact stochastic simulation of chemical systems with many species and many channels. *The Journal of Physical Chemistry A 104*, 9 (2000), 1876–1889.

[26] Giese, A., Bjerkvig, R., Berens, M., and Westphal, M. Cost of migration: Invasion of malignant gliomas and implications for treatment. *Journal of Clinical Oncology 21*, 8 (2003), 1624–1636. PMID: 12697889.

[27] Gillespie, D. T. A general method for numerically simulating the stochastic time evolution of coupled chemical reactions. *Journal Of Computational Physics 22* (1976), 403–434.

[28] Gillespie, D. T. Exact stochastic simulation of coupled chemical reactions. *The Journal of Physical Chemistry 81*, 25 (1977), 2340–2361.

[29] Gillespie, D. T. Approximate accelerated stochastic simulation of chemically reacting systems. *The Journal of Chemical Physics 115*, 4 (2001), 1716–1733.

[30] Gillespie, D. T., Hellander, A., and Petzold, L. R. Perspective: Stochastic algorithms for chemical kinetics. *The Journal of Chemical Physics 138*, 17 (2013), 170901.

[31] Golubev, A. Applications and implications of the exponentially modified gamma distribution as a model for time variabilities related to cell proliferation and gene expression. *Journal of Theoretical Biology 393* (2016), 203–217.

[32] Johannes Müller, C. K. *Methods and Models in Mathematical Biology*. Lecture Notes on Mathematical Modelling in the Life Sciences. Springer Berlin, Heidelberg, 2020.

[33] Kuang, Y. *Delay differential equations: with applications in population dynamics*. Academic press, 1993.

[34] Maini, P. K., and Baker, R. E. Modelling collective cell motion in biology. In *Advances in Applied Mathematics* (Cham, 2014), A. R. Ansari, Ed., Springer International Publishing, pp. 1–11.

[35] Markham, D. C., Simpson, M. J., Maini, P. K., Gaffney, E. A., and Baker, R. E. Incorporating spatial correlations into multispecies mean-field models. *Phys. Rev. E 88* (11 2013), 052713.

[36] McAdams, H. H., and Arkin, A. Stochastic mechanisms in gene exprexpression. *Proceedings of the National Academy of Sciences 94*, 3 (1997), 814–819.

[37] Méhes, E., and Vicsek, T. Collective motion of cells: from experiments to models. *Integrative Biology 6*, 9 (07 2014), 831–854.

[38] Nelson, D. L., and Cox, M. M. *Lehninger principles of biochemistry*. W.H. Freeman, 2017.

[39] Noren, D. P., Wesley, C., Lee, S. H., Qutub, A. A., Warmflash, A., Wagner, D. S., Popel, A. S., and Levchenko, A. Endothelial cells decode vegf-mediated Ca2+ signaling patterns to produce distinct functional responses. *Science Signaling 9*, 416 (2016), ra20–ra20.

[40] Renshaw, E. *Stochastic population processes: analysis, approximations, simulations*. OUP Oxford, 2011.

[41] Schwehm, M. Fast stochastic simulation of metabolic networks. 2001.

[42] Smith, J. A., and Martin, L. Do cells cycle? *Proceedings of the National Academy of Sciences 70*, 4 (1973), 1263–1267.

[43] Thomopoulos, N. T. *ReactionStatistical Distributions*. Springer Cham, 2017.

[44] Tóth, J., Nagy, A. L., and Papp, D. *Reaction Kinetics: Exercises, Programs and Theorems*. Springer New York, NY, 2018.

[45] Van Dyke Parunak, H., Savit, R., and Riolo, R. L. Agent-based modeling vs. equation-based modeling: A case study and users' guide. In *Multi-Agent Systems and Agent-Based Simulation* (Berlin, Heidelberg, 1998), J. S. Sichman, R. Conte, and N. Gilbert, Eds., Springer Berlin, Heidelberg, pp. 10–25.

[46] Vittadello, S. T., McCue, S. W., Gunasingh, G., Haass, N. K., and Simpson, M. J. Mathematical models incorporating a multi-stage cell cycle replicate normally-hidden inherent synchronization in cell proliferation. *Journal of The Royal Society Interface 16*, 157 (2019), 20190382.

[47] Wilensky, U., and Rand, W. *An Introducton to Agent Based Modeleing*. MIT Press, 2015.

[48] Yates, C., Ford, M., and Mort, R. A multi-stage representation of cell proliferation as a Markov process. *Bulletin of Mathematical Biology 79* (2017), 2905–2928.