Statistical Properties of Microbial Phenotypes and Colony Growth

Fluctuations in Transfection, Phenotypic Switching and Range Expansion

Jan-Timm Eike Dieter Maximilian Kuhr

München 2011

Statistical Properties of Microbial Phenotypes and Colony Growth

Fluctuations in Transfection, Phenotypic Switching and Range Expansion

Jan-Timm Eike Dieter Maximilian Kuhr

Dissertation
an der Fakultät für Physik
der Ludwig–Maximilians–Universität
München

vorgelegt von
Jan-Timm Eike Dieter Maximilian Kuhr
aus Starnberg

München, den 31. Oktober 2011

Erstgutachter: Prof. E. Frey
Zweitgutachter: Prof. J. O. Rädler

Tag der mündlichen Prüfung: 19. Januar 2012

Zusammenfassung

Zellen sind die fundamentalen Einheiten aus denen alle Lebewesen aufgebaut sind. Will man verstehen wie grundlegende Aspekte des Lebens auf elementarer Ebene organisiert sind, liegt es also nahe zunächst einzelne Zellen zu untersuchen. Moderne Mikroskopietechniken machen es möglich sowohl subzelluläre Abläufe als auch kollektives Verhalten vieler Mikroben zu untersuchen. In diesem mikroskopischen Regime kommt Fluktuationen eine enorme Bedeutung zu. Da dieses Rauschen eine inhärente Eigenschaft solcher Systeme ist, hat das Leben robuste Systeme entwickelt, welche trotz großer Schwankungen sehr effektiv arbeiten. Nicht nur das: die Fluktuationen werden teilweise sogar ausgenutzt um vorteilhafte Funktionen zu verwirklichen. Bei der Modellierung in diesem Forschungsbereich helfen Konzepte aus der statistischer Mechanik und der Analyse stochastischer Prozesse weiter, welche die Fluktuationen beschreiben können.

Grob gliedert sich diese Arbeit in zwei Teile, in denen auch in die Hintergründe, Konzepte und die Literatur zu den entsprechenden Themenkomplexen eingeführt wird.

Der erste Teil setzt sich mit der Modellierung intrazellulärer Prozesse auseinander, für die molekulares Rauschen bedeutsam ist. Dazu werden zwei Veröffentlichungen, die aus Kollaborationen mit experimentellen Biophysikern erwachsen sind, diskutiert:

Im Rahmen von gentherapeutischen Behandlungen wird externes Erbmaterial an Zellen verabreicht um fehlerhaftes Verhalten zu beheben. Um diesen Prozess zu charakterisieren, wurden Plasmide, welche Fluorophore kodieren, mittels zweier chemischer Transfektionsmethoden in eukaryotische Zellen eingebracht. Die Verteilung der Expressionsniveaus lässt sich durch mehrere stark stochastische Schritte während der Transfektion und nachgeschaltete, quasi-deterministische Expression erklären.

Die zweite Kollaboration beschäftigt sich mit dem Umschaltverhalten zwischen verschiedenen Phänotypen in Bakterien. Im hier untersuchten Fall wird das Auftreten von “Kompetenz” in *B. subtilis* betrachtet. Diese Fähigkeit, genetisches Material aus dem extrazellulären Medium aufzunehmen, ist durch ein Netzwerk wechselwirkender Gene stark reguliert. Während sich die verschiedenen Phänotypen auf stabile Fixpunkte in nicht-linearer Differentialgleichungen zurückführen lassen, wird das Umschalten zwischen den Phänotypen durch Fluktuationen in der kleinen Anzahl von mRNA-Molekülen ausgelöst.

Der zweite Teil der Arbeit geht auf kollektives, stochastisches Wachstum vieler Zellen in einer expandierenden Kolonie ein. Im dazugehörigen Manuskript wird ein theoretisches Modell mit Methoden der statistischen Mechanik analysiert.

Das Wachstums mikrobieller Kolonien wird als Modellsystem für Kolonisationsprozesse gesehen. Inspiriert durch Experimente wird ein stochastischer Oberflächenwachstums-Prozess, beschrieben durch ein verallgemeinertes Eden-Modell, analysiert. Das Modell berücksichtigt

explizit Selektion zwischen zwei Stämmen, irreversible Mutationen sowie die Rauheit der voranschreitenden Front der Kolonie. Der asymmetrische Charakter der Mutationen bringt einen absorbierenden Zustand mit sich, in dem sich nur mehr der Mutantenstamm an der Front befindet. Dadurch verbindet das Modell zwei interessante Teilgebiete aus der statistischen Mechanik des Nicht-Gleichgewichts: Phasenübergänge zu absorbierenden Zuständen und dynamische Oberflächenaufrauhung. Wie für diese Prozesse üblich, lassen sich in der Nähe des Phasenübergangs kritische Exponenten definieren, die das Divergenzverhalten physikalischer Observablen beschreiben und eine Einteilung in Universalitätsklassen ermöglichen. Es stellt sich heraus, dass die Kopplung von Oberflächen- und Populationsdynamik qualitative Veränderungen mit sich bringt, sodass sich das Modell keiner bisher bekannten Universalitätsklasse zuordnen lässt.

Abstract

Cells are the fundamental units of which all life forms are composed. To understand the elementary organization of life, it is therefore meaningful to start the investigation on the single cell level. Modern microscopy permits the examination of both subcellular processes and collective microbial behavior. In this microscopic regime, fluctuations are of eminent importance. As this noise is an inherent property of such systems, life evolved robust systems, which work effectively in spite of severe fluctuations. Moreover, life also makes use of these fluctuations for its benefit. For modeling purposes in this field of research, concepts from statistical mechanics and from the analysis of stochastic processes can be applied to account for the fluctuations.

This work is roughly divided into two parts, which also address the background, concepts and literature of the corresponding topics.

The first part is concerned with the modeling of intracellular processes, for which noise is important. In this context two publications, which arose from collaborations with experimental biophysicists, are discussed:

In gene therapy external genetic material is injected into cells to remedy deficient behavior. To characterize this process, fluorophore encoding plasmids were administered to eukaryotic cells by means of two chemical transfection methods. The distribution of expression levels is explained by several strongly stochastic steps during transfection and subsequent quasi-deterministic gene expression.

The second collaboration addresses the switching kinetics between different phenotypes in bacteria. In the case at hand, the emergence of "competence" in *B. subtilis* is studied. This ability (to take up genetic material from the extracellular medium) is strongly regulated by a network of interacting genes. While the different phenotypes are associated with stable fixed points of non-linear differential equations, switching between phenotypes relies on fluctuations in the small number of mRNA molecules.

The second part of this work elaborates on collective, stochastic growth of many cells in an expanding colony. The corresponding manuscript analyzes a theoretical model with methods from statistical mechanics.

Microbial colony growth is sometimes seen as a model system for range expansion or colonization processes. Inspired by experiments, a stochastic surface growth process, in the form of a generalized Eden model, is set up and analyzed. The model explicitly takes into account selection between two strains, irreversible mutations, and the roughness of the propagating colony front. The asymmetric character of mutations implies the existence of an absorbing state, where only the mutant strain is at the front of the expanding population. Hence, the model combines two interesting branches of non-equilibrium statistical mechanics: phase

transitions to absorbing states and dynamic surface roughening. As usual for these processes, one can define critical exponents, which describe the divergence of observables near the phase transition, and admit organization of models into universality classes. It turns out that the coupling between roughening dynamics and population dynamics induces qualitative different behavior. As a consequence, the model cannot be assigned to any universality class currently known.

Contents

1 Introduction

1.1 Biological Systems and the Physical Approach

Why is it that more and more physicists, and to a lesser degree mathematicians, are getting involved in the general research field of life sciences? The answer to this is probably two-fold: it can be related to recent developments in these sciences and to how their subject, biology in the widest sense, is organized on a fundamental level.

1.1.1 Quantitative Data

Research in many disciplines such as neuroscience, microbiology, genomics, ecology, developmental and evolutionary biology, and so on, is becoming more and more quantitative, thanks to better microscopes, miniaturization and automatization of experimental setups. Where in the past biologists had been able to state that there is a correlation between quantities of interest, while neither the causality nor the mechanisms of interaction were accessible, nowadays high precision, real-time measurements of the relevant quantities are often possible, at least in principle. This brings about an enormous amount of data, which needs to be analyzed and interpreted. To the traditional biologist, this is quite a challenge, since his/her education would usually more focus on a multitudes of (more or less independent) facts. Modern, quantitative biologists, more educated in general principles and their connections, are well capable of handling this data, but physicists have had to deal with complicated data since a long time. This gives physicist a head start in handling complicated biological data, as it does in analyzing data from other fields, e.g., the financial markets, where the field of econophysics prospers.

1.1.2 Biological Systems are Complex Systems

Economy and biology have other things in common: everything is basically connected to everything else, at least indirectly. In both economics and biology, the dynamics are driven, by the flow of money and energy, respectively, and are therefore non-equilibrium in nature. Typically, systems exhibit behavior which emerges from the collective dynamics of the constituents, which are usually very many interacting entities which mainly obey simple rules. Fluctuations are omnipresent, stock courses and biological data are inherently stochastic. Small causes can trigger large scale events with possibly catastrophic consequences.

All of these are hallmarks of complex systems. Again, this is an advantage for a physicist, especially one with experience in the fields of non-linear dynamics, stochastic processes and

statistical mechanics. In these fields complex systems, which show such characteristic behavior, are commonplace. Experience with this kind of intricate problems helps to determine which aspects of a system are relevant and which aspects can, without oversimplification, be neglected to make an analysis feasible.

1.2 Fluctuations in Biological Systems

Biological data is often noisy, sometimes as a result of unavoidable inaccuracy in the measuring process, but often because the measured quantities are inherently distinguished by strong fluctuations. Copy numbers of biomolecules in bacteria are a prominent example [1, 2]. Copy numbers are usually of order 10–100, but occasionally even of order 1. In this number range the discreteness of integer numbers can have marked impact on the dynamics, as can be seen for example in burst-like expression of proteins [3, 4, 5].

To a physicist, fluctuations are very common: on the most fundamental level they arise from the principles of quantum dynamics, which also imply that due to the uncertainty principle, time and place of a reaction between two molecules are as unpredictable as the time when a specific radioactive atom decays. But even if nature was deterministic, it would be virtually impossible to measure the positions and momenta of all atoms and molecules in any given sample of biological matter with a sufficient accuracy for Laplace's demon to predict the dynamics over any significant period of time. Consequently, already on a fundamental level, fluctuations and randomness are unavoidable features of biological systems.

More importantly, thermal fluctuations have significant impact on living systems, since thermal energies are often comparable to energy scales in cellular processes. Thermal fluctuations are also the origin of Brownian motion, or diffusion, which is the most widespread transport mechanism on short length scales. This has first been analytically described by physicists like Einstein, Langevin and Smoluchowski [6, 7, 8, 9].

Often in biology this stochasticity is further coupled to non-linear processes and hence fluctuations may be enhanced to an extent where non-stochastic descriptions turn out to be flat wrong. An impressive example of this is found in autocatalytic reactions [10].

Stochastic modeling of biological processes has the potential to give answers to fundamental questions, as is evident for example in the seminal work of Luria (a biologist) and Delbrück (a theoretical physicist), which showed that mutations in bacteria appear by chance and not in response to an environmental stimulus [11]. This was a strong argument in favor of Darwinian evolution as opposed to the theory put forward by Lamarck.

1.3 Biological Systems in Space

1.3.1 Biological Relevant Length Scales

Classical research fields for a traditional physicist are atomic physics and physical chemistry, which investigates how molecules are formed from the atoms under the influence of quantum

mechanical interactions. Right at this molecular scale biology starts. Every living cell is composed of a myriad of different molecules which build up a living whole, which constantly converts energy to drive all necessary processes. Many subsystems on this scale are composed of identical units: membranes are composed of phospholipids, filaments of monomers and so on. Statistical mechanics is well suited to address such systems, e.g., length regulation of filaments [12], the organization of the genetic material in the nucleus [13], or the interactions between genes in gene regulatory networks and the expression of genes [14, 15, 16, 17, 18]. On the length scale of the cell, spatial aspects become more and more important, and many problems can be addressed by non-linear partial differential equations — for example, organization of cell polarity and cell division [19]. At a broader scale, one might be interested in the interaction of cells, be it in tissue growth and morphogenesis [20] or in the growth dynamics of microbial colonies composed of different phenotypes [21]. On yet larger scales one might study the interaction between species, and how populations interact and organize spatially in their environment. This is part of the fields of ecology and population dynamics and happens on very large length scales, many times the size of the individuals which compose the population.

1.3.2 Spatial Dimension of Biological Systems

Depending on the problem at hand, it might be important to take spatial aspects into account. Reaction kinetics in a single cell are often well described by a zero-dimensional model, as diffusion is very fast and such small volumes can therefore be considered well-mixed. However, if the spatial structure is important itself (as for the organization of the cytoskeleton) one, of course, has to take at least the linear extension of a single filament into account. In the context of population dynamics, especially for the problem of cooperation, it is known that the organization in groups, e.g., mediated through spatial distance in two dimensions, has massive impact on the dynamics and often changes the outcome qualitatively [22]. In some problems, such as the Min oscillations which appear in *E. coli* cell division, it turns out that even though all reactions happen on the $2d$ membrane, the properties of the extension and topology of the $3d$ cytosol may have qualitative impact on the performance of the system [19].

The fact that a spatial setting changes the properties of a system in a fundamental way is well-known from many areas of physics. In statistical mechanics the paradigmatic Ising model exemplifies this: While in a $1d$ setting the Ising model has no phase transition, a phase transition is found in two and more dimensions. Though it is not trivial to describe the phase behavior in two and three dimensions analytically, in more than four dimensions one finds that it is well-described by a mean-field theory, which ignores the spatial correlations.

1.3.3 Non-linearities in Stochastic Spatial Systems

To model biological systems which are explicitly spatial, one often works with stochastic partial differential equations. In this work we shortly introduce three important examples of such equations, the Fisher-Kolmogorov equation (5.1), a coarse-grained field equation for directed percolation (6.17), and the Kardar-Parisi-Zhang equation (6.39). The Kardar-Parisi-Zhang equation differs from the others in the non-linear term and it is conceivable that its

solutions are therefore quite distinct from the other equations. The coarse-grained description of directed percolation is only distinct from the stochastic version of the Fisher-Kolmogorov equation in the shape of the noise term, but still their solutions differ widely.

This is characteristic for non-linear processes, where short wavelength excitations, which are created by the noise, couple into the longer wavelength modes as a result of the non-linearity. Hence, it is not only the deterministic parts of the equations which are of importance, but also the characteristics of the noise.

1.4 Scope of this Work

This work is split up into two parts, one addressing intracellular processes, the other collective growth of heterogeneous microbial cultures.

In the first part we follow the most important steps that led to an understanding of molecular biology and in parallel introduce central concepts of this ever growing discipline along with observation techniques (Chapter 2). In this context we also discuss the importance of fluctuations in more detail, and introduce some concepts relevant for modeling biomolecular systems. These include non-linear differential equations, which describe reaction networks and the master equation, and related concepts, which takes care of randomness resulting from stochastic reactions and discrete molecular copy numbers.

We then turn towards more specific problems and start with an experiment involving artificial gene transfer followed by gene expression, where small number fluctuations lead to a heterogeneity across the members of a cellular population (Chapter 3).

In Chapter 4 we discuss an experiment which reveals how bacteria are able to use fluctuations and non-linear feedback to improve their chances in the struggle to survive. Here, stochastic switching between distinct phenotypes in an isogenic population is described on the single cell level.

The second part of this work does not analyze experiments. However, it has been inspired by experimental work on the growth of bacterial cultures which has been carried out by the author. We first discuss the spread of populations on large scales, and then concentrate on setups with microbial organisms, which are easier to observe (Chapter 5).

In Chapter 6 we introduce two classes of processes from non-equilibrium statistical mechanics: (i) models that undergo phase transitions to absorbing states and (ii) kinetic surface roughening models. Both exhibit properties that can also be found in range expansion experiments carried out with microbial populations. A model that couples these two classes of processes is introduced and analyzed. It is found that the coupling alters the characteristic properties of the individual processes qualitatively.

Part I

Gene Expression and Regulation in Single Cells

2 Molecular Biology and its Modeling

All forms of autonomous life that we know of are composed of cells. From unicellular organisms like most bacteria, archaea, and many eukaryotes to multicellular life forms, all rely on this basic building block of life. The adjective "basic", however, should not be mistaken as a synonym for simple: Even the single-celled bacterium *Escherichia coli*, the workhorse of microbiological research since decades, is far from understood in all its ever extending complexity. Clearly, all evolved mechanisms serve the ultimate aim to better adapt to the environment, thereby enhancing reproduction success and outcompeting competitors. How this is biologically implemented, however, is far from clear. To unravel how molecular constituents are physically organized to process matter, (free) energy and information to live, grow and reproduce is a formidable task from the viewpoint of both biology and physics.

In this chapter we will shortly review some of the important insights in the machinery of the single cell, that have been made during the last century. As knowledge in this subdiscipline of biology is not only stunningly vast and detailed on many levels, but also growing exponentially fast, the here introduced aspects will only cover areas needed to understand the research articles contained in Chapters 3 and 4. To the interested reader the compendium "Molecular Biology of the Cell" by Alberts *et al.* [23] and the references therein are recommended, on which, beside the given references, the biological parts of the following introduction heavily draw.

2.1 Storage and Flow of Information — A Primer to the Genetic Code and Gene Expression

Roughly half of the dry weight of a cell is made up of proteins, each of which is comprised of one or more polypeptide chains of variable length. Proteins are the ubiquitous workers of the cell, which actually *do* most of the numerous tasks a cell performs: Metabolizing nutrients, sensing chemical gradients and other stimuli, propelling the cell through the extracellular medium, the majority of enzymatic activity, to name but a few. Proteins were recognized as a distinct class of biomolecules as early as the 18^{th} century and scientifically studied already in the first half of the 19^{th} [24]. Nonetheless, how proteins received their remarkable range of properties, remained obscure until the role of another biomolecule, the macromolecular nucleic acid DNA (deoxyribonucleic acid), was clarified in the 20^{th} century.

Though DNA had been isolated in the late 19^{th} century [25], and its chemical constituents had been specified in the early 20^{th} [26], its eminent role as the carrier of hereditary information only became clear by the middle of the century. Schrödinger picked up the ideas of the modern evolutionary synthesis [27] when he theorized about a "crystalline" carrier of genetic

information [28], on which Mendel's laws, known since the famous experiments and thorough bookkeeping of Mendel [29], and Darwin's evolution [30], could work. Further Beadle and Tatum had stated the "one gene — one enzyme" hypothesis, which connected the Mendelian genes to proteins [31]. More importantly, a number of ingenious experiments shed light on the physical counterparts of the Mendelian genes and made clear that DNA was the relevant ingredient that gave a virus[1] its virulence, and not its protein content [32, 33]. Further the double-strandedness of DNA, which, through base-pairing, allows for duplication of information, while providing chemical stability needed for storage, was hinted at by the result that pyrimidine bases (cytosine and thymine) and purine bases (adenine and guanine) are found in equal parts in the DNA [34].

In 1953, X-ray diffraction patterns of DNA by Franklin *et al.* [35] and Wilkins *et al.* [36], enabled Watson and Crick to reveal the double helix structure of the DNA [37, 38]. After the physical manifestation of genes had conclusively been identified as nucleic acids, Crick set out to incorporate proteins into the picture by publishing the "sequence hypothesis" and the "central dogma of molecular biology"[2] [40], both of which we will now briefly discuss.

2.1.1 The Genetic Code

The sequence hypothesis simply states that there is a decisive mapping between the one-dimensional sequence of bases on the DNA and the equally one-dimensional sequence (when ignoring folding, disulfide bonds and other subsequent modifications) of amino acids in proteins. The exact mapping was not known by then, but by inducing frameshift mutations of variable length it was soon figured out that three consecutive nucleotides, called a codon, correspond to one amino acid[3] [42]. The remaining task of deciphering the exact mapping (cf. Table 2.1) started with single-nucleotide polymers which were found to produce single-amino-acid poly-peptides [43, 44, 45, 46] and continued by the complete deciphering of the genetic code by use of more complicated artificial nucleotide templates [47, 48].

2.1.2 The Flow of Information

Cricks other proposition, the central dogma of molecular biology, which he detailed further in 1970 [49], expressed that information flows from nucleic acids to polypeptides, but not the other way around, out of proteins and into nucleic acids. By then the intermediary role of

[1] In some viruses the genetic information is stored in RNA instead of DNA. The important point though is that nucleic acids and not proteins inherit "the message".

[2] In his autobiography "What Mad Pursuit" [39] Crick later pointed out:

> I called this idea the central dogma, ... As it turned out, the use of the word dogma caused almost more trouble than it was worth. ... Many years later Jacques Monod pointed out to me that I did not appear to understand the correct use of the word dogma, which is a belief that cannot be doubted.

[3] The first to hypothesize about this 3:1 relationship between the four nucleic acids and twenty amino acids was the theoretical physicist and cosmologist George Gamow [41]. Though his tentative mapping was easily proofed wrong, his simple mathematical reasoning for a minimal codon length of three base pairs ($4^2 < 20 < 4^3$) turned out to be right.

Table 2.1: The genetic code. A codon, i.e. three nucleic acids, contained in the DNA/RNA corresponds to an amino acid, the building block from which any protein is composed. Nucleic acids are given in bold, capital, and single letters (**U**: uracil, **T**: thymine, **C**: cytosine, **A**: adenine, **G**: guanine); where the DNA contains a **T** the mRNA transcript has a **U** at the corresponding position. Amino acids are denoted by common three letters abbreviations. Translation usually starts at the first **ATG**-RNA codon while it is terminated at codons denoted by "Stop".

1st position	2nd position				3rd position
	T/U	**C**	**A**	**G**	
T/U	Phe	Ser	Tyr	Cys	**T/U**
	Phe	Ser	Tyr	Cys	**C**
	Leu	Ser	Stop	Stop	**A**
	Leu	Ser	Stop	Trp	**G**
C	Leu	Pro	His	Arg	**T/U**
	Leu	Pro	His	Arg	**C**
	Leu	Pro	His	Arg	**A**
	Leu	Pro	His	Arg	**G**
A	Ile	Thr	Asn	Ser	**T/U**
	Ile	Thr	Asn	Ser	**C**
	Ile	Thr	Lys	Arg	**A**
	Met (or Start)	Thr	Lys	Arg	**G**
G	Val	Ala	Asp	Gly	**T/U**
	Val	Ala	Asp	Gly	**C**
	Val	Ala	Glu	Gly	**A**
	Val	Ala	Glu	Gly	**G**

RNA (ribonucleic acid), in the form of messenger RNA (mRNA), had been elucidated, and it stood equal to DNA and proteins as a prime carrier of information. In more detail the central dogma classified all possible channels of information flow between the key players, DNA, RNA, and proteins, into three classes: general, special, and unknown, cf. Table 2.2. While one can

Table 2.2: The central dogma of molecular biology. Information flow between DNA, RNA and proteins can be categorized with respect to its prevalence in biology.

general	special	unknown
DNA → DNA	RNA → DNA	proteins → DNA
DNA → RNA	RNA → RNA	proteins → RNA
RNA → proteins	DNA → proteins	proteins → proteins

argue about the exact definition of information flow and about overlaps between the three classes, the general notion is still at the heart of modern molecular biology. Gene expression from DNA, which is transcribed by polymerases to RNA, which in turn is translated by ribosomes to proteins, is one prominent example of information flow of the general class.

That gene expression is still an important topic today, can bee seen upon flicking through any recent issue of a journal concerned with cell biology. The other well known example is the replication of DNA during reproduction, and, less well-know, in the course of horizontal gene transfer (see Chapter 3). The manuscripts introduced in Chapters 3 and 4 discuss and analyze experimental settings in which both horizontal gene transfer and gene expression are of importance.

2.1.3 Complications

With the genetic code and the machinery of gene expression clarified, the sequence of amino acids produced in translation is straightforward. It is now common knowledge, that this linear chain of amino acids is not the functional form of a protein inside a cell. Instead, the protein folds up into a three-dimensional shape: Already during translation first domains form through interactions between neighboring amino acids along the polypeptide backbone. As translation is completed these subunits find their relative positions through electrostatic, especially hydrophobic, interactions and the creation of disulfide and hydrogen bonds or other chemical reactions. This process is often supported by specialized proteins called chaperones. In many cases the folded amino acid chains then ensemble into protein complexes, which in concert perform a given cellular task.

Though vivid, the strict division between nucleic acids as information storage and proteins as cellular tool kit, is, while usually accurate, not true in general — there are contradictions to the dogma. Especially the notion of nucleic acids as the sole source for information had to be reconsidered in the light of epigenetics[4] [51], and, to a lesser extend, prions [52].

Furthermore, without the enzymes polymerase and ribosome no gene expression would be possible, though their blueprint is encoded in DNA as well. This leads to a paradoxical statement: no proteins without DNA, no DNA without proteins. An escape route to this chicken or egg dilemma may be the hermaphroditic character of the RNA, which can both store information and act as an enzyme or regulator. This is the hypothesis of an "RNA world" which may have existed before today's DNA-protein-based biosphere [53]. In this scenario, RNA molecules catalyze their own reproduction. While the notion of an RNA world still leaves the question of how it came into existence unanswered, it is well conceivable that from it the less simple, but more stable DNA-protein world has emerged, which then outperformed its predecessor.

The flow of information is formally restricted to end up in proteins, if neglecting special cases like prions. However, already epigenetic mechanisms in many cases work through proteins which modify the extend to which genes are expressed. This establishes a feedback, which acts on a time scales of a few reproduction times. Furthermore, regulation of gene expression through proteins is ubiquitous and works down to much shorter times. The concept of gene regulation is of so eminent importance that we will discuss it in greater detail below.

[4] Here, the term "epigenetics" is understood as heritable information that is not encoded in the DNA sequence. A contrasting meaning of epigenetics exists in developmental biology, where the term refers to the establishment of the phenotype of a cell from an omnipotent stem cell [50].

2.2 Gene Regulation

It is remarkable what a wide range of organisms exists in the biosphere. In all of them information storage and gene expression work essentially through the mechanisms introduced in the previous section. The beauty of these concepts lies in the mechanistic simplicity of the genetic code and of the mechanisms of protein production.

With that said, it comes as a little bit of a shock, that so-called higher, or more complicated organisms do not necessarily have a bigger genome, or number of genes, to carry all that substantial information. However, there is a reassuring analogy to brain size and intelligence: not the animal with the biggest brain is the most intelligent, but rather the cerebral performance is to a large part dependent on the connections between the brain's neurons. Similarly, not the total length of the genome or the total number of genes is solely relevant, but also the interconnections between the genes do matter a lot.

2.2.1 The lac Operon - Transcription Regulation

The research in gene regulation started with the *lac* operon, which is a segment in the genome of *E. coli*, which was studied by Monod and Jacob [54]. They found out how the microbe adapts its metabolism if its most common food source, glucose, is replaced by lactose[5]. The key realization was, that under this nutrient condition, a specific protein (the *lac* repressor) which inhibits the polymerase from binding to the DNA, detaches from the DNA. Now polymerase can attach to this site and transcription of genes can begin. This is facilitated by another protein (the catabolite activator protein), which attaches nearby and enhances binding of polymerase. The transcription of three genes is initiated, followed by expression of the corresponding proteins, which make consumption of lactose possible. Thus, Jacob and Monod where abel to categorize different regions in the DNA sequence: "structural genes", which encode proteins necessary for lactose digestion, and "regulatory genes", like the promoter (where the polymerase binds), the sites of the described regulation (the operator), and the terminator (where transcription stops). Together they form the *lac* operon.

The regulation principles of the *lac* operon can be extended to many other operons hosting other genes, and are probably present, to some extend, in all forms of life[6]. It turns out, that many different proteins act as activators or repressors, often on a number of operons including their own. This can create non-trivial, extended gene regulatory networks, which perform tasks similar to engineered electronic circuits, including feedback, signal integration and signal processing [55].

[5] Of course, other factors are also involved in the regulation of the *lac* operon, but, for the sake of simplicity, we focus on the most dominant ones.

[6] Or, as Monod reportedly put it,

> What is true for *E. coli* is true for an elephant, only more so.

2.2.2 Other Regulation Mechanisms

Transcription regulation is by far not the only way through which expression of different proteins is coupled; regulatory mechanisms exist on every conceivable level and display an enormous variability. Some of them are:

- Transcription initiation, which can not only work through transcription factors, but also through alternations at other parts of the apparatus necessary for the polymerase to start transcription.

- Post-transcriptional modification, which occurs only in eukaryotes, includes, e.g., splicing of mRNA and addition of a poly(A) tail to it. The latter influences export from the nucleus, translation and degradation of mRNA.

- Translation regulation, which refers to binding of proteins or RNA to an mRNA template or the ribosome, which results in regulations of translation efficiency.

- RNA interference, where small RNA molecules, called siRNAs enhance the degradation of other single stranded RNAs through formation of double-stranded RNA, which are rapidly degraded.

- Epigenetics, where proteins organize the arrangement of the chromatin structure in the nucleus. This mechanism has been argued to be important for cell differentiation in multicellular organisms.

2.2.3 Some Important Examples of Regulation

The multitude of regulatory mechanisms lets cells adapt to changes in their environment with unparalleled versatility, manifest in such prominent examples as the heat shock reaction [56], chemotaxis [57], and quorum sensing [58]. Astonishingly, by means of the same regulatory modules in different combination, complex processes can robustly be implemented, e.g., circadian rhythms [59] and the cell cycle.

Generally, regulatory networks may constitute bi-, or multistable, excitable or more complicated dynamical systems in the space of gene expression levels. This implies that even though organisms may have identical genes, i.e., are clonal or have the same genotype, they can still be very different with respect to the proteins being produced, resulting in different physical appearance and features, also known as phenotype. An example for phenotypic regulation will be discussed in Chapter 4. In multicellular organisms similar mechanisms lead to cellular differentiation during development and morphogenesis, thus creating all kinds of tissue of the organism [60, 61].

2.3 Observation on the Single Cell Level — Fluctuations in Biological Systems

2.3.1 Diffusion and Fluctuations in Chemical Reactions

Most of the important discoveries discussed so far in this chapter, and many more historic results, have been found through measurements on large numbers of cells or viruses. This sampling of a whole population gives robust results, when interested in the mean behavior of biological systems. However, biological systems are prone to fluctuations, they live in an environment very different to the one we experience. At the relevant length scales ($\sim \mu m$), friction dominates over inertia and the medium is effectively of very high viscosity [62]. Furthermore thermal fluctuations ($\sim k_B T$) are prevalent since the typical energy scales are of similar magnitude ($\sim pN \cdot nm$).

While of importance to biology, these effects are observed in any mesoscopic system. The most prominent example is diffusion, also known as Brownian motion. This jittery movement, which small pieces of matter perform in a liquid or gas, can easily be observed under a microscope, and even by bare eye. Also first mentioned (and interpreted quite correctly) by Lucretius more than two millennia ago [63], it took great minds like Einstein [6], Langevin [7] and Smoluchowski [8] to work out the theoretical details of this seemingly simple problem in the 20^{th} century. While many aspects of diffusion have been clarified since then, it continues to be a field of active research to the present day especially in the context of biological physics [9].

Diffusive movement is even more relevant for molecules, resulting in random encounters between reactive components. Consequently it is neither possible to predict which molecules will react next, nor when and where[7]. In the large volumes used in chemical laboratories the discreteness of chemical reactions is averaged out through the large number of reactions happening almost simultaneously. Under these conditions descriptions in terms of concentrations suffice, which neglect fluctuations. This is not the case if copy numbers of molecules are small. Delbrück demonstrated this as one of the first, by analyzing a system exhibiting autocatalytic reactions [10], where averages alone were not sufficient any more. Instead probabilistic descriptions are necessary to capture the essential features. This is the result of small number fluctuations being amplified through the autocatalytic reaction.

2.3.2 Single Cell vs. Population-Wide Studies

In single cells volumes are small ($\sim 10^{-15}$–$10^{-12} l$) and instead of concentrations one has to turn to number of molecules; in a bacterium the number of molecules of a given kind is

[7] Erwin Schrödinger phrased this as [28]:

> [...] when a chemist handles a very complicated molecule *in vitro* he is always faced with an enormous number of like molecules. To them his laws apply. He might tell you, for example, that one minute after he has started some particular reaction half of the molecules will have reacted, and after a second minute three-quarters of them will have done so. But whether a particular molecule, supposing you could follow its course, will be among those which have reacted or among those which are still untouched, he could not predict. That is a matter of pure chance.

often of order 1 (e.g. for the DNA) to 1000 (e.g. for many proteins). Here, randomness in the reactions, known as "intrinsic noise" [64] can be of eminent importance, as was demonstrated, e.g., in a seminal work by Arkin *et al.* [65]. They showed that fluctuations can determine whether a host cell is rapidly exploited and killed by a virus, or instead used as a carrier of the viral DNA for an extended period of time.

Work like this emphasizes the need to observe cells not only on a population wide level; much more information is hidden in the cell-to-cell variability [1, 3]. During the last decades experimental setups were refined to go down to observations of single cells, and lately even to the single molecules in them. It increasingly becomes clear that fluctuations should not be seen as an inevitable vexation, but are instead necessary and useful ingredients to cellular function, or functional noise [66]. Many single-cell observation techniques, which are used to elucidate the impact of fluctuations, rely heavily on the use of fluorescent marker proteins, which can be expressed *in vivo*.

2.3.3 Fluorescent Proteins — Reporters of Gene Expression

The most prominent example of this class is the "green fluorescent protein" (GFP), naturally produced in the jellyfish *Aequorea victoria*, which emits green light after excitation with blue or ultraviolet light. Thirty years after identification of the protein [67], it was first expressed endogenously in model organisms [68], where its potential as a reporter for gene expression was readily recognized. With the availability of derivatives (produced through mutations in the corresponding DNA sequence) with different color [69, 70] and enhanced fluorescence properties [71], it became possible to monitor different cellular processes simultaneously in real time and *in vivo*. The spectacular properties of GFP and its derivatives are a direct result of its folding into an exceptional three dimensional shape, during which the chromophore forms by spontaneous oxidation at a specific position along the amino acid sequence [69, 70]. Fluorescent proteins can be used to study a variety of biological systems, since it is possible to construct DNA where expression is under the control of a promoter corresponding to an altogether different protein. This way, the expression patterns of any given protein can in principle be observed. Furthermore, it is possible to express fusion proteins, made from the fluorescent protein and any other protein. If the other protein is located at a specific organelle or specific position of the cell, then so will be the fluorescence.

2.4 Mathematical Treatment

In the last section we already discussed the importance of incorporating the discreteness of molecule copy numbers and stochasticity when modeling single cell processes. Before considering the effects of these complications we will shortly discuss the case where they can be neglected.

2.4.1 Averaged Description by Differential Equations

If copy numbers are large and reactions fast compared to consecutive observations, and the volume where the reactions happen can be considered well-mixed so that spatial aspects can be neglected, the change of concentrations can be inferred directly from chemical equations. These typically have the form

$$\text{reactants} \xrightarrow{\text{rate}} \text{products}\,. \tag{2.1}$$

As reactants can take part in different reactions and products might themselves be reactants of different reactions one often has to consider a set of chemical reactions. The change in the concentration c_i, of a given molecular species i, can then be written as a differential equation,

$$\dot{c}_i = f_i(c_1, c_2, \ldots)\,. \tag{2.2}$$

The reaction term f_i on the-right hand side is usually composed of a sum of terms, each of which corresponds to a reaction of type (2.1), in which the number of chemical species i changes. Since the number of reacting molecules taking part in reactions is often larger than one, or the reaction rates depend on concentrations, reaction terms are non-linear in all but the simplest cases. The non-linearities often render analytical derivation of the temporal evolution of concentrations impossible, but they generate interesting behavior which can be analyzed by methods of non-linear dynamics [72, 73], with many examples in biological modeling [74]. They are also the main reason for the richness of modules encountered in gene regulatory networks (as discussed in Section 2.2), where the reactive molecules are interacting genes, mRNAs and proteins.

To reduce the number of coupled equations one often assumes that some reactions, which are fast compared to all others, are in chemical equilibrium, known as adiabatic approximation. This is especially useful for intermediary bound complexes, as in the case of the famous Michaelis-Menten reaction, for substrate-enzyme-complexes.

2.4.2 Stochastic Description by the Master Equation

When the typical number of molecules becomes small, fluctuations in chemical reactions are so important that a description in terms of concentration is not meaningful anymore and one has to resort to a stochastic description [75, 76].

If states are discrete, as for molecular numbers, the master equation[8] is the most basic and general description. Instead of concentrations we now have to consider probabilities $p(s)$ of finding the system in state vector or simply state $s = (X_1, X_2, \ldots)$, where X_i is the random number of molecules (or complexes) of species i. The master equation is actually a set of differential equations which describe the temporal change of probabilities of all possible states:

$$\dot{p}(s) = \sum_{s'} w_{s' \to s} p(s') - \sum_{s'} w_{s \to s'} p(s) \tag{2.3}$$

[8] Due to its importance in the context of chemical reactions it is often termed the *chemical* master equation.

Here $w_{s' \to s}$ is the transition rate for the system to go from state s' to state s. As there typically is a vast number of coupled equations, the master equation is exactly solvable only in the simplest cases. In these cases the set of equations is mostly transformed into a (usually non-linear) partial-differential equation by use of generating functions [77]. The partial-differential equation then has to be solved and subsequently back-transformed, where the last step is the hardest and can often only be achieved through some approximation.

2.4.3 Gillespie's Stochastic Algorithm

Before we consider some analytical approximations to the master equation, we discuss its direct numerical solution through Monte Carlo simulations. The idea here is that instead of solving the master equation numerically by discretization in time, one simulates an ensemble of possible realizations of the system under consideration. The most common approach to this is the stochastic first reaction method or simply Gillespie algorithm [78, 79]. A specific realization is given by the sequence of reactions that alter the state of the system and the corresponding times of each reaction, both of which are sequences of random numbers. All processes we consider here are Markov processes, which means that though random, the evolution of the system under consideration only depends on its current state. To simulate a realization it is therefore necessary to infer the probability distributions of time and type of the next reaction depending on which state the system is in.

The total rate $\lambda(s)$ at with the system leaves its state s, is given by the sum over all transition rates out to this state:

$$\lambda(s) = \sum_{s'} w_{s \to s'} \tag{2.4}$$

It is easy to show that the distribution $p(\tau)$ of waiting times τ until the next reaction, disregarding which one, is given by

$$p(\tau) = \lambda e^{-\lambda \tau} \ . \tag{2.5}$$

Which of all possible reactions happens next is a matter of chance. However, the probability $p(s \to s')$, for the reaction that takes the system front state s to state s', is given by the ration of the respective reaction rate to the total reaction rate:

$$p(s \to s') = \frac{w_{s \to s'}}{\lambda(s)} \tag{2.6}$$

With the probability distributions (2.5) and (2.6) derived, the algorithm is straightforward:

1. Set values for the rates $w_{s \to s'}$ and the initial state s of the system at time $t = 0$.
2. Calculate $\lambda(s)$ according to (2.4)
3. Determine type $s \to s'$ and time τ of next reaction according to the probability distributions (2.6) and (2.5).
4. Set the system to its new state s' and update time $t \to t + \tau$.

5. Return to step 2.

The above procedure is repeated sufficiently often to sample a representative proportion of all possible realization of the process under consideration. All quantities of interest can then be derived for any time t and up to any desired accuracy, only limited by available computing time.

2.4.4 Analytical Approximations of the Master Equation

While simulations can, in principle, be employed to analyze arbitrary complex systems, they rarely grant access to the more general mechanism, which are encoded in the underlying organization of biochemical reaction networks. To elucidate these evolved design principles, one often simplifies the system under consideration as far as possible, without changing it qualitatively, and then tries to approximatively solve the master equation analytically.

A very common approach is the expansion of the master equation to derive a Fokker-Planck equation [75, 80], which is a second order partial differential equation in probability space, of the general form

$$\dot{p}(s) = -\nabla \mathbf{A}(s)p(s) + \frac{1}{2}\nabla^2 \mathbf{B}(s)p(s) \ , \tag{2.7}$$

where ∇ is the nabla operator in the space of molecular copy numbers X_i and $\mathbf{A}(s)$ and $\mathbf{B}(s)$ are state dependent drift vector and diffusion coefficient matrix, respectively. The first term is usually associated with the well-mixed mean field differential equations (2.2), while the second term describes the influence of stochastic fluctuations. The Fokker-Planck equation can then be tackled with the analytical tools of calculus of partial differential equations.

To obtain a Fokker-Planck equation there exist at least two different methods, which are distinct with respect to the quantity in which the expansion is performed. In what is known as the Kramers-Moyal expansion [75, 80], one assumes that the change in molecular copy number ΔX_i in any reaction is much smaller than the number of molecules X_i present. Therefore, one can perform a continuum limit and expand both the rates and the probabilities in the master equation. To obtain a Fokker-Planck equation one then neglects all terms containing derivatives of order three or higher. Alternatively, for the van-Kampen-expansion [76], one argues that the system size is large and one can therefore perform a systematic expansion in the inverse system size. Similar to perturbation theory one receives a hierarchical set of Fokker-Planck equations, which can then be solved consecutively. Though more complicated, the van-Kampen-expansion has the advantage that it can systematically treat processes where the system size is not fixed.

The Fokker-Planck equation (2.7), is in many cases equivalent to the Langevin equation, another partial differential equation, which treats fluctuations in a different way. It still has a drift term closely related to the the mean field differential equations (2.2), but models the fluctuations as a additional random field. It has the general form

$$\dot{s} = \mathbf{A}(s) + \sqrt{\mathbf{B}(s)}\eta(s,t) \tag{2.8a}$$

where $\eta(s,t)$ is Gaussian white noise,

$$\langle \eta(s,t) \rangle = 0 \,, \tag{2.8b}$$

$$\langle \eta(s,t)\eta(s',t') \rangle = \delta(s-s')\delta(t-t') \,. \tag{2.8c}$$

Here $\delta(s-s')$ is shorthand for $\prod_i \delta(X_i - X_i')$. To go from a Fokker-Planck equation to the corresponding Langevin equation, one usually diagonalizes the diffusion coefficient matrix $\mathbf{B}(s)$, which enables computation of the square root.

There exist several other approaches to the problem of solving the master equation, and a number of methods to further treat the derived Langevin and/or Fokker-Planck equations. However, these techniques are more involved, and their use very much depends on the problem at hand. Since they are not relevant to this work, the reader is referred to one of the standard text books [75, 76].

3 Artificial Gene Transfer and Transgene Expression

As a cell divides, its genetic information is duplicated and passed on to both daughter cells. Sometimes, during DNA duplication, errors happen (a major source of mutations) which alter the genetic information. Though changed, the genes are closely related to that of the previous generation, so that it is in principle possible to set up a "family tree", which illustrates the descent and degree of relatedness of genes. Such a family tree can also be derived for sexual species, since any specific gene can be linked directly to either father or mother, though the situation is more involved as meiosis and recombination mix up the parental genomes. By tracking back where a gene originated, it is possible to set up real family trees. This even facilities the reconstruction of interconnections between species, i.e. the phylogeny or "tree of life". However, as usually in biology, if one takes a closer look, things turn out not to be quite so simple.

3.1 Horizontal Gene Transfer

Not all genetic information is passed on by means of cellular division, but instead *horizontal gene transfer* between two contemporaries is possible [81]. Astonishingly, this occasionally even happens between individuals of completely different species [82, 83].

3.1.1 Modes of Horizontal Gene Transfer

One may distinguish between different modes of horizontal gene transfer, though definitions sometimes overlap:

- **Conjugation** is the direct transfer of genetic material between two bacteria through direct contact of the cell membranes. Conjugation is the bacterial equivalent of sex and seen as the main reason for the spread of resistance against antibiotic agents and other traits between different strains of bacteria [84].

- **Transformation** is the uptake of free exogenous DNA by bacteria from the extracellular environment, first described by Griffith [85]. This may be induced artificially, but is a natural property of some bacteria, which are then called competent. In Chapter 4 we address switching to the competent phenotype of *B. subtilis*, which is capable of transformation, in more detail.

- **Transduction** works through bacteriophages, which are viruses with bacterial hosts [86]. Upon infection the viral genome is either expressed excessively and new viral particles are released as the bacterium is lysed (i.e. the bacterium bursts open) or the virus' genome is integrated into the bacterium's and (possibly) expressed much later, which is called lysogeny. The latter may result in horizontal gene transfer, since neither integration into the genome nor packing of the viral DNA into the new viral particles are error-free. Thus, genes may travel from bacterium to bacterium by means of a viral vector (a vector is a vehicle which transports genetic material). Less often viruses also transport genetic material between different species.

- **Transfection**, in distinction to the other modes of transfer, is the artificial and deliberate transfer of nucleic acids into cells. Mostly it refers to non-viral transfer of genetic material into eukaryotic cells. Transfection is an important tool of molecular biology to clarify intracellular processes. A wealth of different methods has been developed during the past decades, and we will discuss them in greater detail in the next section.

The natural occurrence of horizontal gene transfer posts some problems to phylogenetics, the study of relatedness between species. It implies that species that may have separated a long time ago, and consequently have very different genomes, may still share some very similar genes, which have been horizontally transferred more recently. It is in general not a simple task to infer which genes have been inherited in a direct line from ancestors and which have been "shared", by the above mechanisms, between contemporaries at some intermediary time.

3.1.2 The Endosymbiont Theory

Besides the horizontal transfer of just a part of the genetic content, as discussed above, it is also possible that two individuals "join" their genomes by forming a new organism: The widely accepted *endosymbiont theory* [87, 88], which conjectures that some organelles (usually mitochondria and plastids) of modern eukaryotes have in fact in the far past been independent bacteria, which have been engulfed through phagocytosis by an eukaryotic ancestor. This led to a symbiotic relation between host and endosymbiont, in the course of which specific cellular functions were reallocated through exchange of genetic material. This introduces yet another level of complication in resolving a complete tree of life, but is well beyond the scope of this work and is mentioned here mainly for completeness and to demonstrate on the extraordinary paths nature takes.

3.2 Transfection

While we have just seen that horizontal gene transfer may happen through different natural mechanisms, the one best suitable for laboratory research is the artificial variant, transfection, which is achieved in various ways.

3.2.1 Transfection Methods

Methods of artificial horizontal gene transfer can be roughly distinguished by the way entry to the cytosol is achieved. Uptake either relies on the chemical properties of the used vector, or on other physical effects. Some non-chemical methods are:

- Gene electrotransfer or electroporation: A transient voltage potential is applied to the cell membrane, which opens pores through which genetic material can enter the cell [89].
- Sonoporation: Acoustic waves, most often at ultrasonic frequencies, are used to partially rupture the cell membrane [90].
- Optical transfection: A short, focussed, and intense laser pulse is used to burn a hole into the membrane of a single cell [91].
- Impalefection: Cells are pressed to surfaces covered with sub-micrometer sized needles or fibers to which plasmids are attached. As a result of the mechanical pressure the cells are pierced and the genetic material enters [92].
- Microinjection: A micropipette is used to inject the genetic material into an individual cell [93].
- Biolistic particle delivery, also known as gene gun: Metal particles are covered with genetic material and then shot into the cell [94].
- Magnetofection: Magnetic nanoparticles are covered with genetic material. A magnetic field is applied to bring particles and cells into contact and facilitate cellular uptake [95].

Chemical methods rely on the interaction of *complexes*, composed of DNA and other chemical compounds, with the cell wall:

- Calcium phosphate: Through a precipitation reaction fine grains of calcium phosphate are formed which bind DNA to their surface. These complexes are brought into contact with cells and taken up by processes which are not fully resolved so far [96].
- Lipofection: Liposomes are artificial spherical structures composed from (usually cationic) phospholipid bilayers (i.e. liposomes are vesicles). As cell membranes are also composed of lipid bilayers, liposomes can fuse with them and thereby deliver genetic material, contained in their interior, to the cytosol [97].
- Cationic polymers: Positively charged polymers and negatively charged nucleic acids form complexes, which are taken up by cells via endocytosis [98].

After uptake of the genetic material to the cytosol, the next step is the import of the plasmids into the nucleus. In many cases one finds correlations of import probability with the cell cycle, since during mitosis the nuclear envelope breaks down, which facilitates uptake [99, 100, 101]. Therefore, transfection is more efficient if cell cycle and administration of genetic material to the cells are properly timed.

3.2.2 Applications of Transfection

Transfection experiments are used to investigate in the mechanisms of gene transfer and expression, but also increasingly for specific applications. Two important fields are medical use in the form of *gene therapy* and genetic alteration of organisms to fabricate chemical compounds or food sources, create enhanced materials, or even process information, known as *genetic engineering*.

Genetic Engineering

After the mechanisms of gene expression and regulation had been elucidated (cf. Chapter 2), the mechanistic way in which they work made human alterations of cellular functions conceivable. In the 1970s both bacterial [102, 103] and eukaryotic [104] organisms were first altered by genetical engineering, using techniques of artificial horizontal gene transfer. Since then genetic engineering has been applied in various fields [105]. Prominent examples include:

- Transgenic expression of proteins for medical use, such as insulin, human growth hormone and antibiotics.
- Providing crops with resistance to herbicides to increase yield in agriculture.
- Enabling microorganisms to produce larger quantities of biofuel from a wider range of biological crude materials.

One of the younger subdisciplines of genetic engineering, *synthetic biology*, goes one step further by assembling the genetic material to be introduced from scratch, instead of simply using an intact segment of genetic material from an existing species. The first examples for this were simple motives of artificial gene regulatory networks, like the repressilator [106] or the toggle switch [107]. Since then this technique attracted more and more attention due to the diverse applications of the approach [108]. At its extreme, synthetic biology is used to create "new" organisms by assembly of a complete genome and subsequent injection in a cell deprived of its own genome [109, 110].

The moral and ethical implications of genetic engineering are a matter of social and political debate and will (hopefully) be as long as there is progress in this field. From the viewpoint of science we stand at the beginning of an epoch, which will probably change production, agriculture, engineering and medicine in a similar fashion as chemical engineering did from the beginning of the 20^{th} century onwards.

Gene Therapy

Some human diseases stem from absence, under- or overexpression or other malfunctions of certain genes. In these cases one does not want to add or change the genetic functionality by adding non-human genes, but instead aims to restore the function to what is considered healthy. If this is done by bringing genetic material into the cells, one speaks of gene therapy.

All transfection methods described above, alongside with viral and hybrid approaches, are being used in the growing field of gene therapy. Here, the aim is to ameliorate a disease by either inserting useful genes, silencing or switching off of harmful genes, or altering the expression of malfunctioning genes. The aforementioned techniques all have their advantages and disadvantages, and it is important to compare the effectiveness of treatment, relative to each other. To this end single-cell experiments can be performed, which compare the percentage of successfully transfected cells. Moreover, it is not only important if whether or not a gene in question has successfully been transfected to a given cell, but also at which gene copy number: As in more traditional therapeutic approaches, one does want to treat a as great as possible percentage of all cells, while not overdosing too many.

3.2.3 Stochasticity in Transfection

On a subcellular level, the processes involved in transfection are very complicated [111, 112, 113, 114]. The genetic material, mostly in form of DNA rings known as *plasmids*, has to be brought in contact with the cell membrane, pass it, possibly escape from an endosome, advance to the nucleus, enter the nucleus, and finally be expressed by the cellular machinery.[1] Each of these steps is composed of a number of substeps itself and each may be of more or less importance for successful transfection, depending on the method used.

All of these steps are chemical reactions, during which the concentration of surviving plasmids is strongly diluted. For this reason usually only a fraction of all cells in a culture is successfully transfected. This is of course the result of the inherent strong fluctuations in all processes involving small numbers of molecules (cf. Section 2.3). These fluctuations also lead to a distribution of gene copy numbers in the subset of cells that have been transfected and show transgenic expression. Consequently, the level of proteins expressed from the plasmids shows a wide distribution across the cell population. The variability is composed from the influence of both variance in gene copy number and stochasticity in gene expression. By analyzing models of transfection and subsequent expression it is possible to infer information about the efficiency and mechanisms of transfection from the distribution of expressed proteins.

In the manuscript entitled “Predictive Modeling of Non-Viral Gene Transfer” by G. Schwake, S. Youssef, J.-T. Kuhr, S. Gude, M. P. David, E. Mendoza, E. Frey, and J. O. Rädler single-cell gene expression data, after transfection by means of lipofection and cationic polymers, is analyzed. The distribution of expressed proteins is mainly determined by the gene copy number distribution and thus carries the signature of the stochastic processes relevant during transfection.

[1] It should be noted that it is also possible to transfect RNA or proteins, in which case uptake to the nucleus is of no importance.

ARTICLE

BIOTECHNOLOGY and BIOENGINEERING

Predictive Modeling of Non-Viral Gene Transfer

Gerlinde Schwake,[1] Simon Youssef,[1,2] Jan-Timm Kuhr,[1,2,3] Sebastian Gude,[1] Maria Pamela David,[1] Eduardo Mendoza,[1,2,4] Erwin Frey,[1,2,3] Joachim O. Rädler[1,2]

[1]Fakultät für Physik, Ludwig-Maximilians-Universität, Geschwister-Scholl-Platz 1, D-80539 München, Germany; telephone: +49-(0)89-2180-2438; fax: +49-(0)89-2180-3182; e-mail: gerlinde.schwake@physik.lmu.de

[2]Center for NanoScience (CeNS), Geschwister-Scholl-Platz 1, D-80539 München, Germany

[3]Arnold Sommerfeld Center for Theoretical Physics, Geschwister-Scholl-Platz 1, D-80539 München, Germany

[4]Institute of Mathematics, University of the Philippines, Diliman, Quezon City 1101, Philippines

Received 16 June 2009; revision received 17 September 2009; accepted 21 October 2009

Published online 1 December 2009 in Wiley InterScience (www.interscience.wiley.com). DOI 10.1002/bit.22604

ABSTRACT: In non-viral gene delivery, the variance of transgenic expression stems from the low number of plasmids successfully transferred. Here, we experimentally determine Lipofectamine- and PEI-mediated exogenous gene expression distributions from single cell time-lapse analysis. Broad Poisson-like distributions of steady state expression are observed for both transfection agents, when used with synchronized cell lines. At the same time, co-transfection analysis with YFP- and CFP-coding plasmids shows that multiple plasmids are simultaneously expressed, suggesting that plasmids are delivered in correlated units (complexes). We present a mathematical model of transfection, where a stochastic, two-step process is assumed, with the first being the low-probability entry step of complexes into the nucleus, followed by the subsequent release and activation of a small number of plasmids from a delivered complex. This conceptually simple model consistently predicts the observed fraction of transfected cells, the cotransfection ratio and the expression level distribution. It yields the number of efficient plasmids per complex and elucidates the origin of the associated noise, consequently providing a platform for evaluating and improving non-viral vectors.

Biotechnol. Bioeng. 2010;105: 805–813.

KEYWORDS: mathematical modeling; gene transfer; single cell; transfection/gene expression

Introduction

Non-viral gene delivery systems have evolved over the last decade into widely used vectors for exogenous DNA delivery to eukaryotic cells. Synthetic cationic lipids and polymers, in particular, are used in molecular biology for transgene expression, and are being further refined for use in DNA-based therapies (Ferber, 2001; Patil et al., 2005; Roth and Sundaram, 2004). Despite considerable progress in the efficiency and characterization of vectors, important aspects of the delivery pathway and transfer kinetics remain poorly understood, including how artificial vectors are taken up, transported to the nucleus, and how these factors collectively influence the expression characteristics of a cell population. Current understanding from intracellular studies of transgene delivery includes the following steps: DNA–vector complex uptake via the endosomal pathway, followed by endosomal escape and cytoplasmic transport, nuclear entry, vector unpacking and transcription initiation (de Bruin et al., 2007; Kircheis et al., 2001; Lechardeur et al., 2005; Roth and Sundaram, 2004; Safinya, 2001; Suh et al., 2003). These processes are accompanied by a huge loss of material and temporal delays. It is therefore not surprising that transfected cells in a culture respond very heterogeneously over time, notably in terms of the expression onset time (t_{on}) and the maximum expression levels attained. It is generally accepted that the expression behavior of a single transfected cell is stochastic, yet cell culture averaged expression levels are reliable indicators of gene transfer efficiency.

Flow cytometry is commonly used to measure fluorescence distributions over a population at a rate of up to 10,000 cells/second (Longo and Hasty, 2006). High-content single cell assays, in contrast, are particularly

Gerlinde Schwake, Simon Youssef, and Jan-Timm Kuhr contributed equally to this work.

Correspondence to: J.O. Rädler

Contract grant sponsor: Deutsche Forschungsgemeinschaft

Contract grant number: SFB486-B10; SFB TR12

Additional Supporting Information may be found in the online version of this article.

suitable for investigating the dynamics and heterogeneity of clonal cell populations, since individual cells can be followed with a high temporal resolution. In addition, quantitative image analysis has been successfully improved to reliably convert fluorescence intensities into copies of molecules, hence paving the path to follow ''gene expression by numbers'' (Rosenfeld et al., 2005).

In this article, we analyze gene expression following non-viral gene delivery, with focus on the variance of expression levels. The expression of genes exhibits all-or-nothing characteristics (Hume, 2000) and additional stochasticity exists in transcriptional regulation (McAdams and Arkin, 1999; Rao et al., 2002). Elowitz et al. (2002) have analyzed noise in bacterial gene expression and elucidated the distinction between extrinsic and intrinsic noise, that is, the contribution of fluctuations in cellular components and inherent stochasticity of the biochemical processes during gene expression. The extrinsic variance of gene expression within a clonal population of eukaryotic cells has been investigated in the light of stochastic theories (Blake et al., 2003; McAdams and Arkin, 1997; Raser and O'Shea, 2004; Volfson et al., 2006). It was only recently that attempts were made to generate models for transgene expression following non-viral gene delivery (Dinh et al., 2007; Varga et al., 2000, 2001; Zhou et al., 2007). Computational modeling might greatly enhance our understanding of gene transfer and aid in elucidating the nature of the underlying transport barriers. Many of the issues regarding cell entry and intracellular transport are shared with attempts to model viral infection (Douglas, 2008; Varga et al., 2005).

In this article, we used quantitative single cell time-lapse microscopy combined with mathematical modeling to analyze the variability in transgene expression (Fig. 1). From the synthetic delivery agents currently being evaluated for therapeutic use, we chose polyethyleneimine (PEI) (Boussif et al., 1995) and the commercial Lipofectamine 2000, as cationic polymer and lipid model systems, respectively. Both synthetic vectors are able to condense plasmid DNA into DNA-nano particles, denoted as cationic lipid (cationic polymer)–DNA complexes or just ''complexes.'' Distributions of the expression onset times and expression steady state levels were evaluated for both vectors.

Data are well described by a stochastic delivery model, which is based on the assumption that in a decisive step, only a small number of complexes enter the nucleus through a stochastic process. Out of these complexes, only a fraction of the plasmid load is expressed (Fig. 1). The theoretical model is further corroborated by a cotransfection analysis, that is, the case of the simultaneous transfection using two distinguishable plasmids encoding for CFP and YFP. It is shown that this model consistently describes the fraction of transfected cells and the observed expression level distribution. As a consequence the effective size of a stochastically delivered unit of plasmids (complex) can be determined.

Materials and Methods

Cell Culture

A human bronchial epithelial cell line (BEAS-2B, ATCC) was grown in Earle's MEM supplemented with 10% FBS at 37°C in a humidified atmosphere, 5% CO_2 level. Transfection was performed on both non-synchronized and synchronized cultures. A thymidine kinase double-block was performed to synchronize cells.

Transfection

BEAS-2B cells were grown to 80% confluence from an initial seeding density of 1×10^5 cells/well in six-well plates 24 h before transfection. Cells were washed and the medium is replaced with 1 mL OptiMEM/well immediately before transfection. Optimized transfection procedures were performed using either 2% (v/v) Lipofectamine^{TM}2000/OptiMEM or PEI (N/P = 8)/HBS; 1 μg of pEGFP-N1 or pd2EGFP-N1, is used for transfecting each batch of cells. The transfection medium was prepared either by adding the Lipofectamine or the PEI solution to the plasmid solution. After the transfection media were allowed to stand for 20 min the cells were incubated with 200 μL/well Lipofectamine or PEI transfection medium for 3 h at 37°C, 5% CO_2 level. After 3 h of incubation the medium was removed, and cells were washed with PBS. Cells were reincubated with Leibovitz's L-15 Medium with 10% FBS prior to EGFP expression monitoring.

Cotransfection

Cotransfection was performed with two kinds of preparations containing the same molar amount per plasmid. For one preparation, ECFP/Lipofectamine, EYFP/Lipofectamine, ECFP/PEI and EYFP/PEI were complexed separately. For the other, a mixture of ECFP and EYFP hetero-complexes were complexed with Lipofectamine or PEI. Transfection using either the hetero-complexes (pre-mixed) or a mixture of homo-complexes (post-mixed) was performed as previously described. Cells were reincubated in growth medium and CFP and YFP expression was monitored by fluorescence microscopy after 24 h.

Data Acquisition and Quantitative Image Analysis

Images were taken at 10× magnification, with a constant exposure time of 1 s, at 10 min intervals for at least 30 h post-transfection. Fluorescence images are consolidated into single image sequence files. Negative control images were taken to assess lamp threshold values and autofluorescence, and were subtracted from corresponding image sequence files to eliminate autofluorescence effects. To capture cell

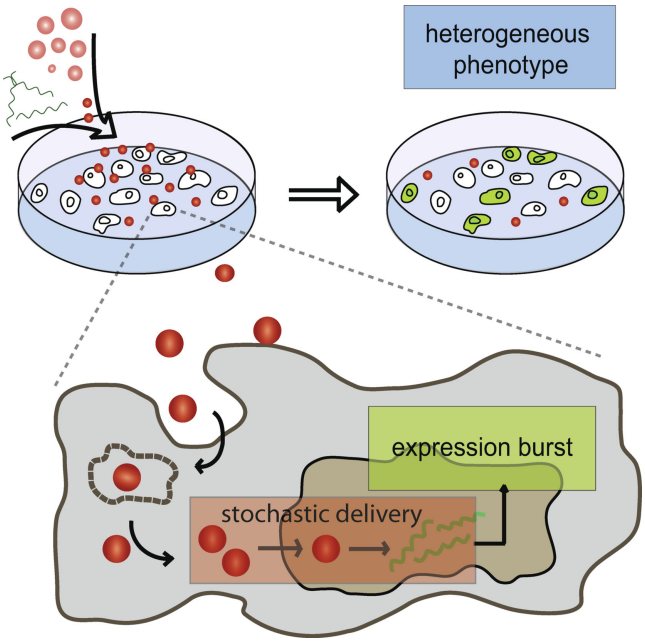


Figure 1. Experimental setup for single cell transfection experiments (**upper part**) and key elements of the theoretical model (**lower part**). EGFP-encoding plasmids and cationic agents form complexes, which are administered to eukaryotic cell cultures. Automated single cell microscopy yields statistics on phenotypic expression of EGFP. For the delivery of plasmids to the nucleus, stochastic effects are important, while the following expression of fluorescent proteins can be described in a deterministic fashion.

fluorescence over the entire sequence, regions of interest (ROIs) were manually defined around each cell (Fig. 2). Changes in total gray measurements in individual ROIs were determined for each time point.

Results

Time Lapse Microscopy and Single Cell EGFP Expression

A cell line of lung epithelium cells was transfected with a plasmid encoding for the green fluorescent protein (EGFP). Transfection protocols for PEI- and Lipofectamine-mediated delivery followed standard procedures and are described in detail in the Supplementary Online Information. We denote the time of the gene vector administration to the cell culture as $t = 0$ h. Transfection medium was removed and cell growth medium added at $t = 3$ h. Single-cell EGFP expression was monitored by automatically taking sequences of fluorescence micrographs from 25 view fields at 10-min intervals. Figure 2a shows a representative sequence from a Lipofectamine transfection experiment, with the initial bright field image, as well as the EGFP fluorescence at $t = 4$, 8, and 12 h post-transfection. These images demonstrate heterogeneity in both the expression onset times and levels of exogenous gene expression. It is observed that the number of fluorescent cells increases with time; at the late stage (~30 h), the ratio of fluorescent to non-fluorescent cells is about 23% and 30% for PEI- and Lipofectamine-mediated transfection, respectively. A total of 500–1,500 cells were monitored in parallel within one time-lapse experiment. Individual time courses of the total fluorescence per cell were evaluated by image processing from data stacks as shown in Figure 2b and described in the Supplementary Online Information. Figure 2c shows a series of representative time traces from one transfection experiment, illustrating the significant variance in both the expression onset time and EGFP expression level. The typical sigmoidal shape of the time courses is well described by the phenomenological function

$$I(t) = \frac{I_{\max}}{2}\left[1 + \tanh\left(\frac{t - t_{1/2}}{t_{\text{rise}}}\right)\right] \qquad (1)$$

which allows the determination of the maximal fluorescence plateau value ($I_{\max}$) the time of half-maximum ($t_{1/2}$) and the characteristic rise time, t_{rise}. The fluorescence intensities were converted into molecular units using EGFP standard beads for calibration (see Supplementary Online Information). In the remainder of the text we will give the fluorescence intensity $I_{\max}$ in units of EGFP numbers G. Equation (1) proved to be robust for automated data analysis, facilitating the accumulation of statistics for a large number of individual cells. In order to determine the time points of expression onset, t_{on}, we use $t_{1/2}$–t_{rise} as an approximation due to the lack of a well-defined point of onset as shown in Figure 2d. Since the timing of gene expression is expected to be dependent on the cell cycle, we investigated the distribution of t_{on} for synchronized and non-synchronized cells. To this end, cells were arrested at the G1/S-Phase transition using a thymidine kinase double-block (Merrill, 1998).

Distribution of Expression Onset Times

Figure 3 summarizes the measured distribution functions of t_{on} and $I_{\max}$ for GFP expression after transfection of non-synchronized and synchronized cultures with PEI- and Lipofectamine-based complexes respectively. Expression onset times range from 5 and 25 h for both PEI- and Lipofectamine, indicating the existence of a time window during which plasmids are successfully transcribed in the nucleus. The distribution for Lipofectamine clearly peaks at an earlier time (~8 h) compared to PEI (~16 h). On cell cycle synchronization, the distribution of onset times sharpen and become peaked at about $t = 15$ h for both PEI and Lipofectamine. Bright field images reveal that most synchronized cells divide 12 h after transfection. This is consistent with the fact that transfection was carried out in mid-S-phase, 3 h after release from the thymidine kinase

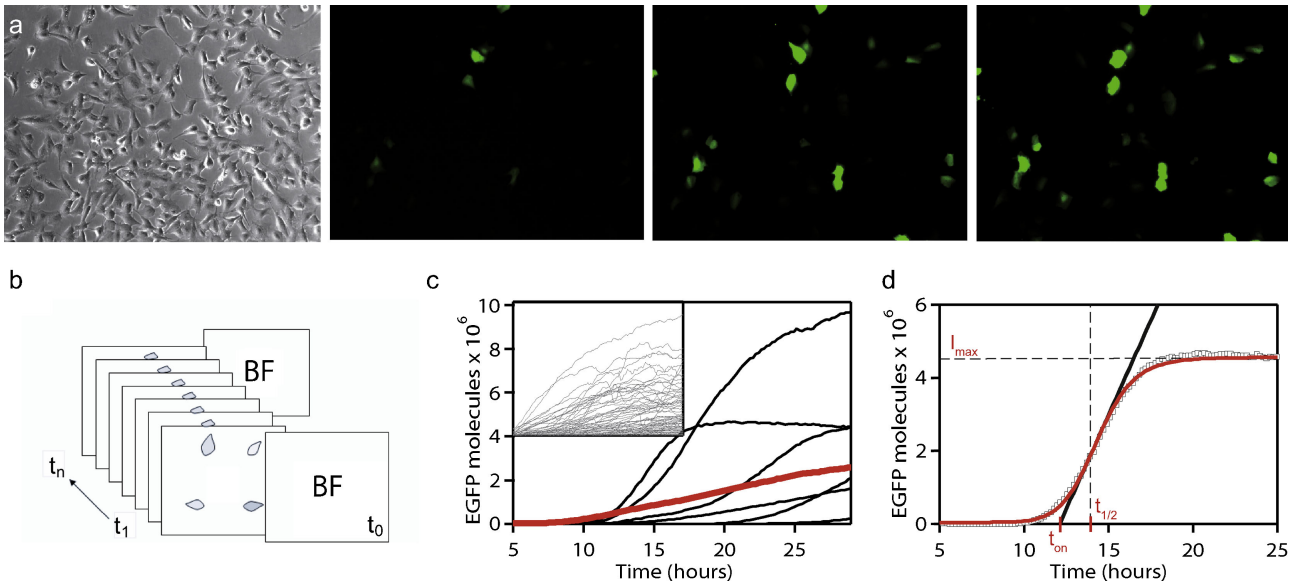


Figure 2. Acquisition of single cell time series. **a**: Microscopy view fields from a Lipofectamine transfection experiment. The first frame is a bright field (BF) control image. Fluorescence image sequences are taken automatically at 10-min intervals for at least 30 h. **b**: Definition of regions of interest (ROIs), total gray value measurement and conversion to the number of EGFP molecules. **c**: Representative time-courses of EGFP expression in individual cells following PEI-transfection. The population average (red) is plotted to demonstrate its linear increase in contrast to the sigmoidal shape of the individual traces. **d**: Characteristic parameters of expression are obtained by fitting the heuristic function 1 (red) to the recorded fluorescence time course (black). The time of expression onset, t_{on}, is calculated from the time of half-maximal expression $t_{1/2}$ and the slope at that point.

double-block. This shows that plasmid activation occurs about 3 h after the M-phase. Furthermore, since cell cycle synchronization suppresses expression at earlier times, there is evidence that the delivery process depends on the cell cycle dependent breakdown of the nuclear membrane. This is consistent with previous studies claiming that mitosis enhances transgene nuclear translocation in cationic lipid gene delivery (Brunner et al., 2000; Mortimer et al., 1999; Tseng et al., 1999). We also find that synchronization leads to a twofold higher steady state expression for PEI-mediated (from 2.6×10^6 to 4.3×10^6 average EGFP molecules per cell) and Lipofectamine-mediated (from 2.9×10^6 to 5.1×10^6 average EGFP molecules per cell) transfection, consistent with earlier observations on ensemble averaged data (Brunner et al., 2000).

Modeling Steady State Gene Expression

In order to analyze the distribution of expression steady states, we introduce a mathematical model that describes EGFP expression after transfer and nuclear translocation of complexes containing exogenous EGFP plasmids. Stochasticity due to nuclear translocation of the plasmid complexes and the intra-nuclear activation will give rise to a probability distribution $P(X)$ for X successfully expressed plasmids (see Fig. 4a). In the first stage of analysis, we describe the expression of EGFP from a single activated plasmid in a linear deterministic model and neglect any cell–cell variability. Based on the biochemical reactions describing transcription, translation and maturation as shown in Figure 4b we denote the ensuing rate equations. Here, R denotes the number of RNA molecules, U the number of unfolded polypeptide chains, and G the number of folded EGFP proteins:

$$\dot{R} = s_A X - \delta_R R \tag{2}$$

$$\dot{U} = s_p R - (k_M + \delta_U) U \tag{3}$$

$$\dot{G} = k_M U - \delta_G G \tag{4}$$

s_A, s_P, and k_M denote the rate constants for transcription, translation and EGFP maturation, δ_R, δ_U, and δ_G denote the degradation constants of each product, respectively. The degradation rates of folded (δ_G) and unfolded protein (δ_U) are assumed to be equal, since the same proteases are involved (Leveau and Lindow, 2001). A plasmid degradation term was omitted, since its occurrence is predicted to be negligible within the time frame considered (Subramanian and Srienc, 1996). Literature values for the individual kinetic rates are summarized in Table I. Equations (2)–(4) can be solved analytically. For the steady state value, a linear relation

$$I_{max} = G(t \to \infty) = \frac{k_M s_P s_A}{\delta_G \delta_R (k_M + \delta_G)} X \tag{5}$$

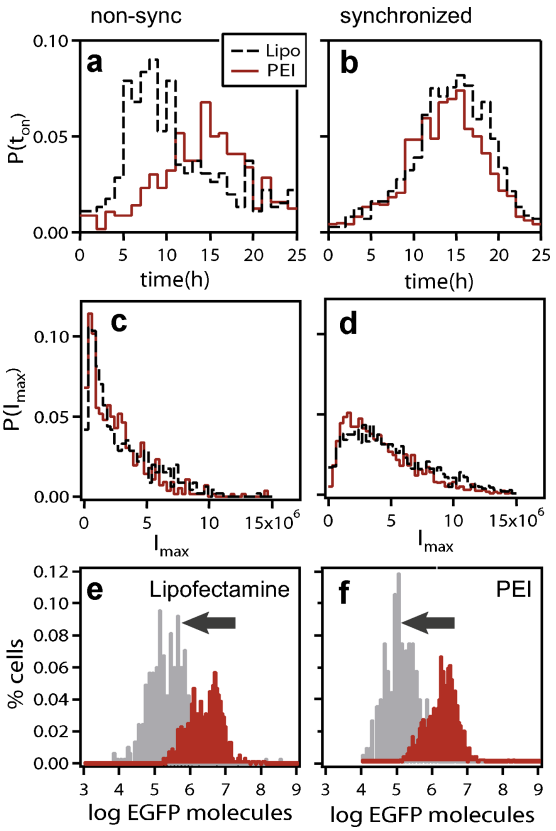


Figure 3. EGFP expression statistics for PEI- and Lipofectamine-mediated transfection. Distributions of expression onset times t_{on} (**a** and **b**) and maximal expression values I_{max} (**c** and **d**), for PEI-mediated (red) and Lipofectamine-mediated (dashed black) transfection depict strong variability within the cell cultures. The total number of expressing cells was 23% out of 560 for PEI and 30% out of 502 in the case of Lipofectamine. **b** and **d**: Thymidine kinase-synchronized cultures with 40% out of 1981 and 30% out of 1797 cells expressing EGFP for PEI and Lipofectamine, respectively. For synchronized cells, expression onset time distributions coincide for Lipofectamine and PEI, indicating that transfection is more likely to happen in specific phases of the cell cycle. Distributions for I_{max} (given in units of EGFP molecules) cannot be explained by post-transfectional sources of fluctuations alone. **e** and **f**: Effect of the altered expression rates on the distribution of maximal expression levels I_{max}. Distributions for d2EGFP (gray) and EGFP (red) transfected with Lipofectamine (e) or PEI (f) are shown. d2EGFP, which has a higher degradation rate, exhibits a systematic shift of the I_{max} distribution compared to EGFP, independent of the vector used. Besides this shift, a change in the number of proteins per active plasmid, k_{exp}, preserves the shape of the distribution. This suggests that the shape is determined during plasmid delivery prior to expression.

between the expression level $I_{max} = [GFP]$ and the number of expressed plasmids $X = [plasmids]$ is obtained:

$$[GFP] = k_{exp}[plasmids] \tag{6}$$

Here k_{exp} denotes an effective expression factor, corresponding to the number of proteins expressed per transcribed plasmid in the steady state. With the values given in Table I, we find $k_{exp} \approx 4 \times 10^6$ molecules/plasmid, which compared to the experimental number of molecules $(1\text{–}15 \times 10^6)$, results in the remarkable finding that the number of plasmids X is of order 1. This implies that most of the variance in expression level originates from stochastic variations in the small number of plasmids, such that the distribution of GFP expression is determined by the distribution of successfully delivered plasmids, $P(G) \sim P(X)$.

To further substantiate this conclusion, we designed an experiment where the expression factor k_{exp} is deliberately modified through the use of destabilized EGFP. It has a 14-fold higher degradation rate (δ_{desG}) due to an additional amino acid sequence (PEST), which makes it more susceptible to proteolysis (Kain, 1999). Figure 3e and f display the shift in the steady state distribution of I_{max}, shown in a logarithmic scale. As predicted above, the shape of the distribution function is almost unchanged for both PEI- and Lipofectamine-mediated transfection. In addition, the peak positions shifted by a factor 12.5, which is close to the value 14.3 predicted from Equation (5).

Modeling Transfection Noise

Unlike in chromosomal DNA, which contains a fixed number of genes, the transfection experiments discussed here result in the delivery of a variable number of genes per vector. We model gene delivery as a two-step stochastic process as shown in Figure 4a. As we will argue in the following, a two-step model is the simplest model that is in accordance with the experimental data. The model consists of (i) the nuclear translocation with probability μ of complexes containing an average of m plasmids and (ii) intra-nuclear activation of plasmids, with probability q. Whereas the probability q subsumes all phenomena promoting or interfering with transcription such as DNA methylation or complexation. We assume that the first process, the delivery of complexes to the nucleus, is rare and statistically independent, yielding a Poisson distribution for the number of delivered complexes C:

$$P(C) = \frac{\mu^C}{C!}\, e^{-\mu} \tag{7}$$

characterized by its mean value μ. Secondly, the independent activation of a plasmid in the nucleus is described by a Bernoulli process with success probability q. The concatenation of both processes results in an expression for $P(X)$ which retains the characteristics of a Poissonian. Mathematical details of its derivation can be found in the Supplementary Data. Figure 5 shows the calculated distribution of activated plasmids, $P(X)$ (red bars), to the measured experimental protein distribution, $P_{exp}(G)$ (green bars). In addition a theoretical protein distribution is shown as black lines. $P_{theo}(G)$ is obtained from $P(X)$ by additionally accounting for noise in gene expression, where we have used a relative magnitude of 0.3 for post-transfectional noise from the literature (see Supplementary Data). Note that the x-axis of the distributions $P(G)$ is rescaled by the factor k_{exp} according to Equation (6). The agreement between

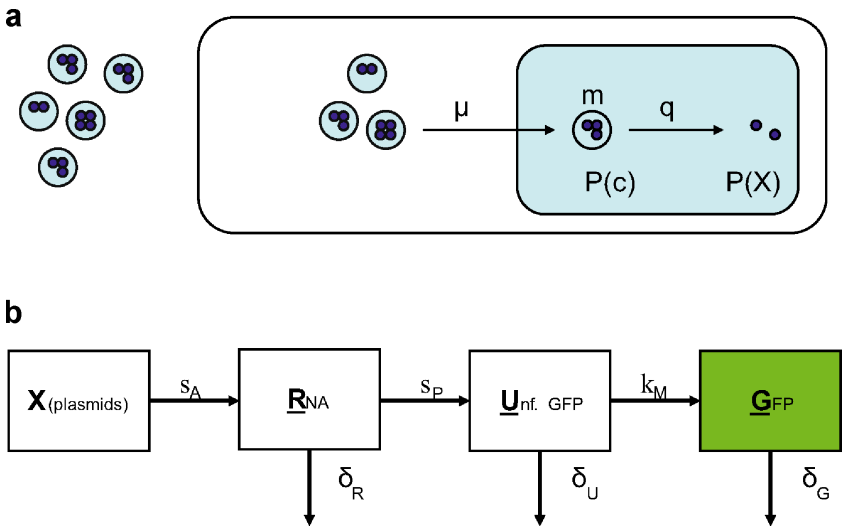


Figure 4. Theoretical model for transfection and gene expression. **a**: Our model of plasmid delivery consists of several stochastic components. The number of complexes *C* delivered per cell is Poisson-distributed, with mean μ. Each complex carries a random number of plasmids, described by a Poisson distribution with mean *m*. Finally, each plasmid has an activation probability *q*, resulting in a Binomial distribution of active plasmids *X* out of the total number of delivered plasmids. With this approach, the overall distribution, *P*(*X*), of actively expressing plasmids can be derived. **b**: Deterministic model of EGFP expression including transcription (s_A), translation (s_P) and protein maturation (k_M). mRNA (*R*), unfolded proteins (*U*) and GFP (*G*) are degraded with rates δ_R, δ_U and δ_G, respectively. Solving the corresponding rate equations, the steady state distribution of fluorescent proteins, *P*(*G*), can be related to that of active plasmids, *P*(*X*).

experiment and model is remarkable considering that there is only one free parameter in the fit. This is due to the fact that two additional experimental constraints have to be met. These are the measured fraction of transfected cells, TR, defined as the percentage of cells expressing one or more plasmids, and the average number of GFP molecules per cell, $\langle G \rangle$, determined by calibration. The parameters μ and mq are fixed by these constraints. The remaining unknown is the expression factor, $k_{exp,}$ which is determined by the fit shown in Figure 5. We obtain $k_{exp} \approx 1 \times 10^6$, $m_{eff} \approx 3$, and $\mu \approx 0.3$–0.5.

The red curve in Figure 5 represents the distribution of successfully expressed plasmids. This distribution, which is directly related to the number of expressed GFP protein though Equation (6), has a well defined mean value given by

$$\langle [\text{plasmids}] \rangle = \mu m q \tag{8}$$

The transfection ratio, TR, is related to this mean plasmid number. It depends on the average number of complexes delivered μ and the effective probability $\tilde{q}$ that from any given complex at least one plasmid is transcribed (we present a complete derivation of these quantities in the Supplementary Data).

$$\text{TR}(\mu, m, q) = 1 - \exp\{-\mu\tilde{q}\} \tag{9}$$

For the data shown in Figure 6, TR is of order 20%.

Cotransfection and Correlated Delivery

One important ingredient of our model is the delivery of DNA in units or complexes and the subsequent correlated coexpression of multiple plasmid copies. This assumption is closely related to the question of whether DNA-complexes fully dissociate before nuclear entry or complexes enter the nucleus as a whole. To elucidate this issue, we studied cotransfection of two distinguishable plasmids (CFP and YFP) and analyzed the outcome of transfection using pre-mixed and post-mixed complexes. Pre-mixed complexes

Table I. Literature values for the kinetic rates of the linear gene expression model

Rate	Best estimate in h^{-1}	Literature values in h^{-1}	References
s_A (transcription)	180	145–240	Hume (2000)
s_P (translation)	100	60–180	Molecular Biology of the Cell 3rd Edition (1995)
k_M (fluorophore maturation)	1.0	0.96–1.28	Sniegowski et al. (2005)
δ_M (mRNA degradation)	0.10	<0.14	Sacchetti et al. (2001)
δ_P (protein degradation)	0.035	0.03–0.04	Sacchetti et al. (2001)
δ_{desG} (destabilized protein degradation)	0.5	0.5	BD Biosciences Clontech

contain both CFP- and YFP-plasmids in a single complex, whereas post-mixed complexes contain either CFP- or YFP-plasmids (for clarity see Fig. 6a and b). The steady-state CFP/YFP expression was analyzed at 24 h post-transfection for approximately 15,000 cells. We define the cotransfection ratio, r, as the number of cells expressing both CFP and YFP divided by the number of cells expressing either CFP or YFP. We find that the cotransfection ratio increases from 12.9% for post-mixed complexes to 21.9% for premixed complexes. The significant difference could not be explained, if complexes were completely dissolved in the cytosol and delivery of plasmids was independent from the complexes. The two-step delivery model, however, naturally explains the discrepancy between pre-mixed and post-mixed complexes. Based on our model, an analytical expression for the cotransfection ratio can be derived (see Supplementary Data) which predicts correctly the measured cotransfection ratios, if the same parameters are used as determined from the EGFP distribution function.

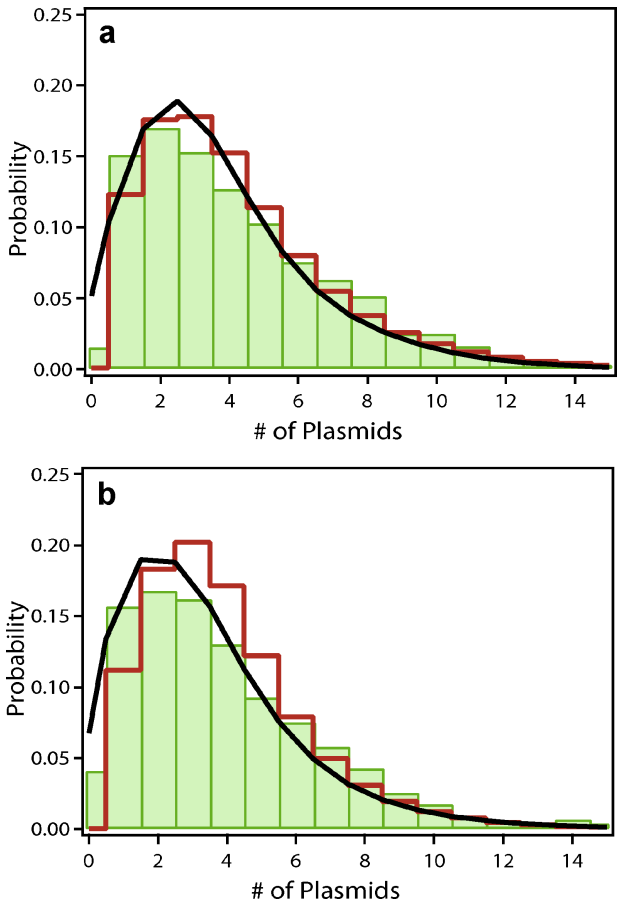


Figure 5. Comparison of single-cell data with the theoretical model. The theoretical EGFP distribution (black) is intimately connected with the underlying distribution of expressing plasmids (red). To facilitate comparison, the protein distribution has been scaled down by the average number of proteins per active plasmid in steady state, k_{exp}. **a** and **b**: For synchronized cultures the measured protein distribution (green) is fitted very well by our theoretical model (black). The fit for PEI transfection (a) yields an average number of delivered complexes, $\mu = 0.53$, and an average number of activated plasmids per complex, $m_{eff} = 3.2$. In the case of Lipofectamine (b), we find $\mu = 0.37$ and $m_{eff} = 3.2$.

Discussion

We have measured the distribution of expression onset times and steady-state expression levels derived from single cell fluorescence time courses. Distributions of onset times of PEI and Lipofectamine collapse on a single curve for synchronized cell cultures, suggesting a universal cell cycle-dependent gene delivery mechanism. Synchronized cells exhibit a broad Poissonian distribution in expression levels and cotransfection experiments reveal correlations in the delivery probability for plasmids contained in one complex.

Invoking Occam's razor, we analyzed the findings in terms of an idealized minimalist model of gene transfection, which describes gene delivery as a two-step stochastic process. Yet our model proves to have considerable predictive power by relating measurable quantities such as the overall transfection efficiency, the cotransfection probability and the shape of the gene expression distribution with each other. Thus, the model allows the derivation of the expression factor, the number of activated plasmids per complex and the average number of delivered complexes from the measured single cell transfection statistics. The model also elucidates the origin of expression variance, separating the noise due to small number fluctuations of complexes, which is inherent to the delivery process and extrinsic sources of noise due to cell–cell variability.

In our gene expression model, we refer to complexes as units of coherently delivered plasmids. Those indirectly inferred complexes are consistent with but not necessarily identical to the complexes described in many physico-chemical studies of PEI and lipofectamine mediated transfection. Cationic-lipid complexes are known to form multi-lamellar aggregates that contain a large number of plasmids (Lasic et al., 1997; Rädler et al., 1997; Zabner et al., 1995). Following endocytotic uptake and release, the complexes slowly dissociate in a stepwise, unwrapping mechanism (Kamiya et al., 2002; Lin et al., 2003). PEI complexes are torroids or rods with a typical hydrodynamic radius of 100 nm (Boussif et al., 1995; DeRouchey et al., 2005), which have been seen to be actively transported inside cells (de Bruin et al., 2007) and to accumulate in the periphery of the nucleus (Suh et al., 2003). Both scenarios describe a situation where numerous small complexes have equal chances of entering the nucleus during the course of mitosis, which is consistent with our model assumptions. Microscopy studies have argued favorably for complexes being at least not fully dissolved at the final delivery stage (Lin et al., 2003; Tseng et al., 1999). However, single nuclear entry events have not been documented explicitly. The probability of transgene expression in the nucleus again depends on the nature of the transfection agent. Pollard et al. (1998) reported that cationic lipids, but not PEI prevent gene expression when complexes are directly injected in the

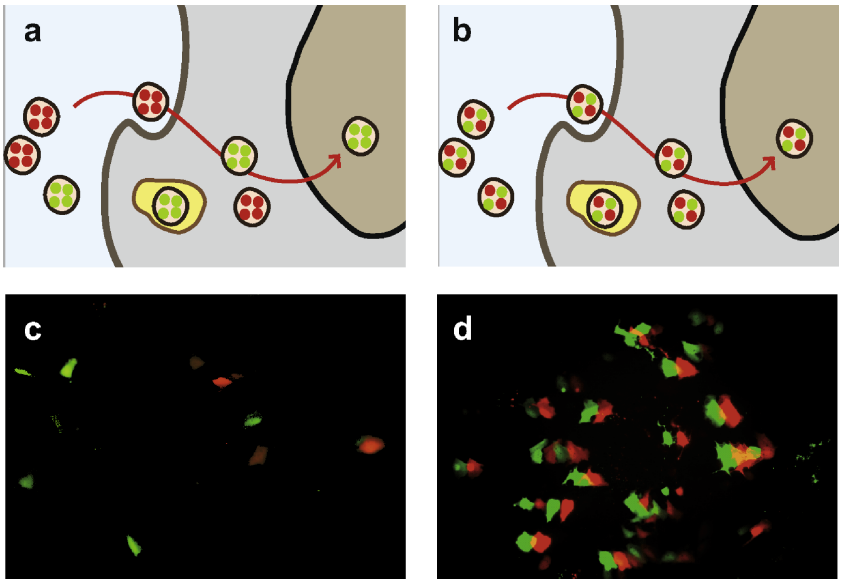


Figure 6. Correlated delivery in CFP/YFP cotransfection with post-mixed (uni-colored) complexes (left) and pre-mixed (dual-colored) complexes (right). **a** and **b**: Post-mixed (uni-colored) and pre-mixed (dual-colored) complexes carry different plasmid content, but take the same pathway to the nucleus. **c** and **d**: Superposition of CFP and YFP fluorescence after transfection reveals a qualitatively different expression pattern for the two distinct experimental protocols. Cyan fluorescence is slightly displaced to permit identification of cotransfected cells. All micrographs are artificially colored.

nucleus. Such findings can only be consolidated with our model, if the delivered complexes transform during the course of the delivery, rather than being the same physical complexes as originally prepared under in vitro conditions. Within the context of our model we restrict ourselves to a narrowed meaning of "complexes" as units of plasmids that are co-delivered. In this framework, we determine the average number of successfully delivered complexes and the effective number of activated plasmids per complex from the analysis of single cell statistics. It will be interesting to corroborate the physical fate and the expression outcome of single complexes by high resolution studies in single cells (de Bruin et al., 2007).

The method to use transfection assays based on automated high-throughput microscopy combined with image processing might evolve into a routine tool for the assessment of transfection efficiency. In contrast to ensemble averaged fluorescence or luminescence data, single cell assays yield precise distribution functions and single cell expression dynamics, which allow a more detailed comparison to theoretical models. As shown here, the analysis of steady state expression levels provides access to the probability of successful plasmid delivery ($P(X)$) and yields an absolute number for the expression factor (k_{exp}). In forthcoming work we will discuss in more detail the distribution of expression onset times and the expression dynamics. We expect that our particular mathematical model can be adapted to a wider class of transfection agents and different types of cells. Their distinct transfection ratios, rate constants and numbers of effective complexes will become even more meaningful in the context of comparative theoretical modeling. A combined experimental and modeling approach will hence help to identify rate-limiting barriers to gene transfer and will result in improved data comparability, making it a versatile tool in the continuous evaluation and improvement of existing synthetic vectors.

We thank Josef Rosenecker, Carsten Rudolph and Ernst Wagner for fruitful discussions and Susanne Kempter for continuous support. The work was supported by the Deutsche Forschungsgemeinschaft through grant SFB486-B10 and SFB TR12, through the Excellence Cluster "Nanosystems Initiative Munich (NIM)," and through the LMU innovative program "Analysis and Modelling of Complex Systems." S.Y., J-T.K. and M.P.D. gratefully acknowledge scholarships from Microsoft Research, the IDK NanoBioTechnology and the DAAD, respectively.

References

Blake WJ, Mads KA, Cantor CR, Collins JJ. 2003. Noise in eukaryotic gene expression. Nature 422(6932):633–637.

Boussif O, Lezoualch F, Zanta MA, Mergny MD, Scherman D, Demeneix B, Behr JP. 1995. A versatile vector for gene and oligonucleotide transfer into cells in culture and in vivo polyethylenimine. Proc Natl Acad Sci USA 92(16):7297–7301.

Brunner S, Sauer T, Carotta S, Cotten M, Saltik M, Wagner E. 2000. Cell cycle dependence of gene transfer by lipoplex polyplex and recombinant adenovirus. Gene Ther 7(5):401–407.

de Bruin K, Ruthardt N, von Gersdorff K, Bausinger R, Wagner E, Ogris M, Braeuchle C. 2007. Cellular dynamics of EGF receptor-targeted synthetic viruses. Mol Ther 15(7):1297–1305.

DeRouchey J, Netz RR, Radler JO. 2005. Structural investigations of DNA-polycation complexes. Eur Phys J E Soft Matter 16(1):17–28.

Dinh A, Pangarkar C, Theofanous T, Mitragotri S. 2007. Understanding intracellular transport processes pertinent to synthetic gene delivery via stochastic simulations and sensitivity analyses. Biophys J 92(3):831–846.
Douglas KL. 2008. Toward development of artificial viruses for gene therapy: A comparative evaluation of viral and non-viral transfection. Biotechnol Prog 24(4):871–883.
Elowitz MB, Levine AJ, Siggia ED, Swain PS. 2002. Stochastic gene expression in a single cell. Science 297(5584):1183–1186.
Ferber D. 2001. Gene therapy: Safer and virus-free? Science 294(5547): 1638–1642.
Hume DA. 2000. Probability in transcriptional regulation and its implications for leukocyte differentiation and inducible gene expression. Blood 96(7):2323–2328.
Kain S. 1999. Green fluorescent protein (GFP): Applications in cell-based assays for drug discovery. Drug Discov Today 4(7):304–312.
Kamiya H, Fujimura Y, Matsuoka I, Harashima H. 2002. Visualization of intracellular trafficking of exogenous DNA delivered by cationic liposomes. Biochem Biophys Res Commun 298(4):591–597.
Kircheis R, Wightman L, Wagner E. 2001. Design and gene delivery activity of modified polyethylenimines. Adv Drug Deliv Rev 53(3): 341–358.
Lasic D, Strey H, Stuart M, Podgornik R, Frederik P. 1997. The structure of DNA-liposome complexes. J Am Chem Soc 119(4):832–833.
Lechardeur D, Verkman AS, Lukacs GL. 2005. Intracellular routing of plasmid DNA during non-viral gene transfer. Adv Drug Deliv Rev 57(5):755–767.
Leveau JHJ, Lindow SE. 2001. Predictive and interpretive simulation of green fluorescent protein expression in reporter bacteria. J Bacteriol 183(23):6752–6762.
Lin A, Slack N, Ahmad A, George C, Samuel C, Safinya C. 2003. Three-dimensional imaging of lipid gene-carriers: Membrane charge density controls universal transfection behavior in lamellar cationic liposome-DNA complexes. Biophys J 84(5):3307–3316.
Longo D, Hasty J. 2006. Dynamics of single-cell gene expression. Mol Syst Biol 2(64):1–10.
McAdams HH, Arkin A. 1997. Stochastic mechanisms in gene expression. Proc Natl Acad Sci USA 94(3):814–819.
McAdams HH, Arkin A. 1999. It's a noisy business! Genetic regulation at the nanomolar scale. Trends Genet 15(2):65–69.
Merrill G. 1998. Cell Synchronization. Methods Cell Biol 57:229–249.
Mortimer I, Tam P, MacLachlan I, Graham RW, Saravolac EG, Joshi PB. 1999. Cationic lipid-mediated transfection of cells in culture requires mitotic activity. Gene Ther 6(3):403–411.
Patil SD, Rhodes DG, Burgess DJ. 2005. DNA-based therapeutics and DNA delivery systems: A comprehensive review. AAPS J 7(1):E61–E77.
Pollard H, Remy J, Loussouarn G, Demolombe S, Behr J, Escande D. 1998. Polyethylenimine but not cationic lipids promotes transgene delivery to the nucleus in mammalian cells. J Biol Chem 273(13):7507–7511.
Rädler JO, Koltover I, Salditt T, Safinya CR. 1997. Structure of DNA-cationic liposome complexes: DNA intercalation in multilamellar membranes in distinct interhelical packing regimes. Science 275 (5301):810–814.
Rao CV, Wolf DM, Arkin AP. 2002. Control, exploitation and tolerance of intracellular noise. Nature 420(6912):231–237.
Raser JM, O'Shea EK. 2004. Control of stochasticity in eukaryotic gene expression. Science 304(5678):1811–1814.
Rosenfeld N, Young JW, Alon U, Swain PS, Elowitz MB. 2005. Gene regulation at the single-cell level. Science 307(5717):1962–1965.
Roth CM, Sundaram S. 2004. Engineering synthetic vectors for improved DNA delivery: Insights from intracellular pathways. Annu Rev Biomed Eng 6:397–426.
Sacchetti A, El Sewedy T, Nasr AF, Alberti S. 2001. Efficient GFP mutations profoundly affect mRNA transcription and translation rates. FEBS Lett 492(1-2):151–155.
Safinya CR. 2001. Structures of lipid-DNA complexes: Supramolecular assembly and gene delivery. Curr Opi Struct Biol 11(4):440–448.
Sniegowski JA, Lappe JW, Patel HN, Huffman HA, Wachter RM. 2005. Base catalysis of chromophore formation in Arg96 and Glu222 variants of green fluorescent protein. J Biol Chem 280(28):26248–26255.
Subramanian S, Srienc F. 1996. Quantitative analysis of transient gene expression in mammalian cells using the green fluorescent protein. J Biotechnol 49(1–3):137–151.
Suh J, Wirtz D, Hanes J. 2003. Efficient active transport of gene nanocarriers to the cell nucleus. Proc Natl Acad Sci USA 100(7):3878–3882.
Tseng W, Haselton F, Giorgio T. 1999. Mitosis enhances transgene expression of plasmid delivered by cationic liposomes. Biochim Biophys Acta 1445(1):53–64.
Varga C, Wickham T, Lauffenburger D. 2000. Receptor-mediated targeting of gene delivery vectors: Insights from molecular mechanisms for improved vehicle design. Biotechnol Bioeng 70(6):593–605.
Varga C, Hong K, Lauffenburger D. 2001. Quantitative analysis of synthetic gene delivery vector design properties. Mol Ther 4(5):438–446.
Varga CM, Tedford NC, Thomas M, Klibanov AM, Griffith LG, Lauffenburger DA. 2005. Quantitative comparison of polyethylenimine formulations and adenoviral vectors in terms of intracellular gene delivery processes. Gene Ther 12(13):1023–1032.
Volfson D, Marciniak J, Blake W, Ostroff N, Tsimring L, Hasty J. 2006. Origins of extrinsic variability in eukaryotic gene expression. Nature 439(7078):861–864.
Zabner J, Fasbender AJ, Moninger T, Poellinger KA, Welsh MJ. 1995. Cellular and molecular barriers to gene-transfer by a cationic lipid. J Biol Chem 270(32):18997–19007.
Zhou J, Yockman JW, Kim SW, Kern SE. 2007. Intracellular kinetics of non-viral gene delivery using polyethylenimine carriers. Pharm Res 24(6): 1079–1087.

Predictive Modeling of non-viral Gene Transfer

Gerlinde Schwake, Simon Youssef, Jan-Timm Kuhr, Sebastian Gude, Maria Pamela David, Eduardo Mendoza, Erwin Frey, Joachim O. Rädler

Supplementary Data

S1 Supplementary Online Information Probabilistic Modeling

Probability distribution of active plasmids per cell

The probability *P(X)* of finding *X* plasmids expressed in a given cell can be computed from a convolution of all underlying stochastic processes that occur prior to transcription initiation. Supposing *X* plasmids have been activated, then $n \geq X$ plasmids first had to be delivered to the nucleus, with a probability *q* for each plasmid to be expressed. This results in a binomial distribution with sample size *n* and parameter *q*:

$$P(X \mid n) = \binom{n}{X} q^X (1-q)^{n-X} \quad . \qquad \textbf{[S1]}$$

Two relevant stochastic processes determine the number of delivered plasmids *n*, namely, the number of complexes *C* that arrive in the nucleus, and the number of plasmids in a given complex. We assume Poisson distributions for both, with means μ and *m*, respectively. Summing over all possibilities, we get the distribution

$$P(n) = \sum_{C=0}^{\infty} \frac{\mu^C}{C!} e^{-\mu} \sum_{n=0}^{\infty} \frac{(Cm)^n}{n!} e^{-Cm} \qquad \textbf{[S2]}$$

for *n*. Here we have used that the convolution of *C* Poisson distributions, each with mean *m*, is again a Poissonian with mean $C \cdot m$.

Considering the previous two equations the overall probability of having *X* active plasmids is

$$P(X) = \sum_{n=0}^{\infty} P(X \mid n) P(n) = \sum_{C=0}^{\infty} \frac{\mu^C}{C!} e^{-\mu} \sum_{n=0}^{\infty} \frac{(Cm)^n}{n!} e^{-Cm} \binom{n}{X} q^X (1-q)^{n-X} \quad . \qquad \textbf{[S3]}$$

By interchanging the order of summation, shifting summation indices and using the normalization condition of the Poisson distribution, this can be rewritten as

$$P(X) = \frac{(mq)^X}{X!} e^{-\mu} \sum_{C=0}^{\infty} \frac{(\mu e^{-mq})^C}{C!} C^X \ . \qquad \textbf{[S4]}$$

Summing from $X = 1$ to infinity yields the transfection probability

$$TE := \text{Prop}(X > 0) = 1 - \exp\left\{\mu(e^{-mq} - 1)\right\} \qquad \textbf{[S5]}$$

which corresponds to Eq. **9** of the main text.

Cotransfection Probabilities

We are interested in the number of cells that are either monochromatic, dichromatic or not fluorescent at all. To compute the probabilities for each, a sum over all possible plasmid numbers X has to be evaluated, with each term in the sum weighted with the probability of activation of zero, one, or two species, depending on the case being considered. If there are i plasmids of one color in the nucleus, the probability that none are activated is $(1-q)^i$, while the probability that at least one is activated is $1-(1-q)^i$

The two cotransfection experiment setups were explained in the Material and Methods section. For uni-colored complexes (post-mixing), the total number of complexes can be subdivided into complexes of either color, yielding a binomial term in the complex number. Thus, for example, the probability of having non-fluorescent cells (not (CFP OR YFP)) is given by

$$\text{Prob}_{\text{post}}(\neg(C \vee Y)) = \sum_{C=0}^{\infty} \frac{\mu^C}{C!} e^{-\mu} \sum_{k=0}^{C} \binom{C}{k} \left(\frac{1}{2}\right)^C \left(\sum_{i=0}^{\infty} \frac{(km)^i}{i!} e^{-km} (1-q)^i \right) \left(\sum_{i=0}^{\infty} \frac{((C-k)m)^i}{i!} e^{-(C-k)m} (1-q)^i \right)$$

. **[S6]**

In the case of dual-colored complexes (pre-mixing), the total number of plasmids is binomial distributed between YFP and GFP, such that the probability of finding, for example, dichromatic cells (CFP AND YFP) is given by:

$$\mathrm{Prob}_{\mathrm{pre}}(C \wedge Y) = \sum_{C=0}^{\infty} \frac{\mu^C}{C!} e^{-\mu} \sum_{n=0}^{\infty} \frac{(Cm)^n}{n!} e^{-Cm} \sum_{i=0}^{n} \binom{n}{i} (1-(1-q)^i)(1-(1-q)^{n-i}) . \quad \textbf{[S7]}$$

Similar expressions can be set up for all other cases. These can be algebraically simplified with the results given in Table SI.

From these expressions, it is easy to compute the cotransfection ratio,

$$r(\mu, m, q) = \frac{\mathrm{Prob}(c \wedge y)}{\mathrm{Prob}(c \vee y)} = \frac{\mathrm{Prob}(c \wedge y)}{2 \cdot \mathrm{Prob}(c \wedge \neg y) + \mathrm{Prob}(c \wedge y)} . \quad \textbf{[S8]}$$

Figure S1 is a representative result for the cotransfection ratio, *r* as a function of the transfection ratio, *TR,* for pre- and post-mixed complexes. Our model predicts that cotransfection is *enhanced* in pre-mixed complexes, and that the probability of cotransfection approaches 1 as *TR* approaches 100%. This is consistent with experimental results. The result shown in Figure S1 is particularly relevant in experiments, since one relies on cotransfection for the simultaneous delivery of two different plasmids.

Distribution of Proteins

The protein number distribution *P(G)* inherently carries the signature of the associated plasmid distribution *P(X)*. Ignoring intrinsic and extrinsic noise in gene expression the mean number of proteins can simply be computed from the distribution of plasmids Eq. **S4** and the expression factor:

$$\langle G \rangle = k_{\mathrm{exp}} \langle X \rangle = k_{\mathrm{exp}} \mu m q \quad \textbf{[S9]}$$

The mean protein number $\langle G \rangle$ can be obtained from single cell statistics. Additional relations are found between the parameters in Eq. **S9** by evaluating how the percentage of non-fluorescent cells, p_0 depends on them. p_0 is identical to the percentage of cells with no activated plasmids in Eq. **S4** or 1- *TR*, where *TR* is the transfection ratio.

$$p_0 := \mathrm{Prob}(X = 0) = \exp\left\{\mu(e^{-mq} - 1)\right\}. \quad \textbf{[S10]}$$

Eliminating μ from Eqs. **S9** and **S10,** and with rearrangements, one finds

$$ae^{a} = we^{w}. \qquad \textbf{[S11]}$$

where $a := \dfrac{\langle G \rangle}{k_{\exp} \ln p_0}$ and $w := mq + a$. Solving Eq. **S11** for w gives the Lambert W-function.

Hence,

$$m_{eff} =: mq = \text{LambertW}\left(ae^{a}\right) - a \qquad \textbf{[S12]}$$

which only depends on measurable quantities and k_{exp}. Fitting the expression factor as the only free parameter, μ and m_{eff} can be determined from single cell data. Consequently, the distribution of active plasmids is set by Eq. **S4**. The distribution of proteins then follows by stretching the distribution of plasmids according to Eq. **6**. As argued in the main text, theory predicts discrete protein distributions, with peaks spaced by k_{exp}. Of course, there are additional noise sources like all post-transfectional fluctuations and limited measurement accuracy. To compare theory with experiment, we replaced the peaks of the discrete protein distribution by Gaussians with the same area and a standard deviation of 0.3 of each peak's position to approximate extrinsic noise. Figure 5 shows the complex, plasmid and protein distributions for a set of single cell data obtained from this theory. A full list of parameters for all four data sets is given in Table SII.

Figure S1 (Schwake et al.)

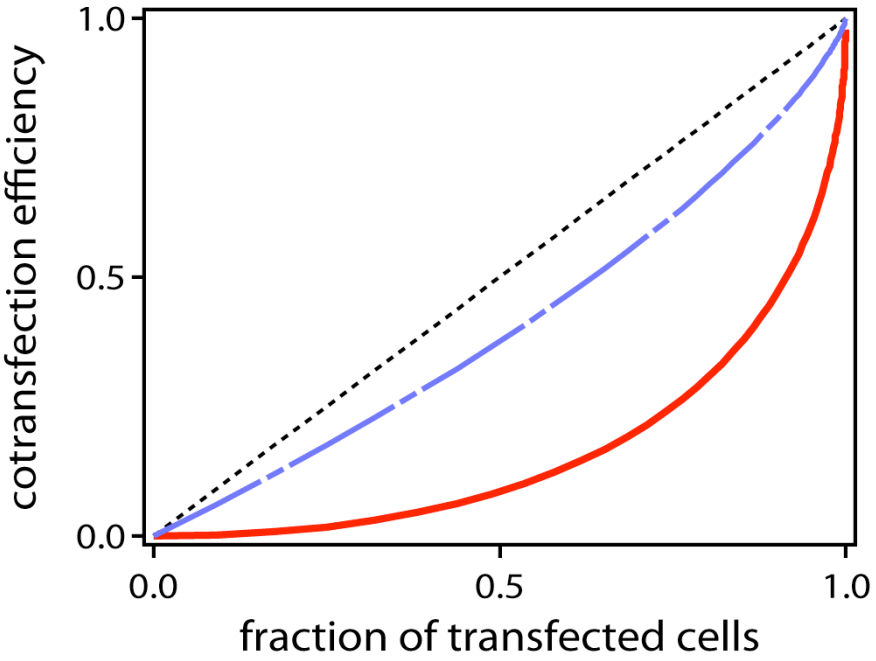


Figure S1. Theoretical prediction for the probability of cotransfection, *r,* as a function of the transfection ratio, *t*, in terms of percent cells transfected. The stochastic model of complex delivery predicts a strong discrepancy between pre-mixed and post-mixed (uni-colored) complexes. The analytic solution is given in the supporting information. Shown are parametric plots with μ varying between 0 and infinity, while $m_{eff} = 3.2$ is held constant.

Table SI (Schwake et al.)

$\mathrm{Prob}_{\mathrm{post}}(C \wedge \neg Y)$	$\exp\left\{\frac{\mu}{2}(e^{-mq}-1)\right\} - \exp\left\{\mu(e^{-mq}-1)\right\}$
$\mathrm{Prob}_{\mathrm{post}}(C \wedge Y)$	$\left(1-\exp\left\{\frac{\mu}{2}(e^{-mq}-1)\right\}\right)^2$
$\mathrm{Prob}_{\mathrm{post}}(\neg(C \vee Y))$	$\exp\left\{\mu(e^{-mq}-1)\right\}$
$\mathrm{Prob}_{\mathrm{pre}}(C \wedge \neg Y)$	$\exp\left\{\mu\left(e^{-\frac{m}{2}q}-1\right)\right\} - \exp\left\{\mu(e^{-mq}-1)\right\}$
$\mathrm{Prob}_{\mathrm{pre}}(C \wedge Y)$	$1-2\exp\left\{\mu\left(e^{-\frac{m}{2}q}-1\right)\right\} + \exp\left\{\mu(e^{-mq}-1)\right\}$
$\mathrm{Prob}_{\mathrm{pre}}(\neg(C \vee Y))$	$\exp\left\{\mu(e^{-mq}-1)\right\}$

Table SI. Probabilities of finding non-fluorescent, monochromatic, and dichromatic cells for pre-mixing (hetero-complex) and post-mixing (homo-complex) cotransfection.

Table SII (Schwake et al.)

data	p_0	$\langle G \rangle$	k_{exp}	m_{eff}	μ	$\langle X \rangle^*$
Synchronized cells, Lipofectamine	0.70	$1.55\cdot10^6$	$1.3\cdot10^6$	3.21	0.37	4.0
Asynchronous cells, Lipofectamine	0.70	$8.72\cdot10^5$	$0.8\cdot10^6$	2.89	0.38	3.6
Synchronized cells, PEI	0.60	$1.71\cdot10^6$	$1.0\cdot10^6$	3.22	0.53	4.3
Asynchronous cells, PEI	0.77	$6.06\cdot10^5$	$0.7\cdot10^6$	3.18	0.27	3.8

Table SII. Parameters of the theoretical protein distribution as inferred from single cell data. The number of non-fluorescent cells, p_0, and the mean number of proteins, $\langle G \rangle$, have been determined experimentally. Fitting the expression factor k_{exp}, to give optimal agreement between theory and experiment, the mean number of delivered plasmids, μ, and the mean number of activated plasmids per complex, $m_{eff} = m \cdot q$, are set by theory. From these parameters, the plasmid distribution, $P(X)$, is determined by Eq. **S4**. $\langle X \rangle^*$ is the mean number of expressing plasmids per fluorescent cell.

S2 Supplementary online information to Materials and Methods

Materials. Earle's MEM (Gibco, Catalog no. 41090-028), FBS (Invitrogen, Catalog no. 10106-185), Leibovitz's L-15 Medium (Gibco, Catalog no. 21083-027), Lipofectamine^{TM}2000 (Invitrogen, Catalog no. 11668-027) and OptiMEM (Gibco, Catalog no. 51985-026) were purchased from Invitrogen. 6-well culture plates (Falcon, Catalog no.353046) and thymidine (Catalog no. 6060-5) were purchased from VWR International GmbH. Linear PEI (25 kDa, Catalog no. 23966) was purchased from Polysciences, Europe GmbH, Eppelheim.

All plasmids were obtained from BD Biosciences: the pEGFP-N1 plasmid (Catalog no. 6085-1) consists of the transcriptional regulatory domain of cytomegalo virus (CMV) preceding the EGFP sequence; the pd2EGFP-N1 plasmid (Catalog no. 6009-1), derived from EGFP, contains a PEST amino acid sequence that targets the protein for degradation and results in rapid protein turnover; the pECFP-N1 plasmid (Catalog no. 6900-1) and the pEYFP-C1 plasmid (Catalog no. 6005-1), both derived from EGFP, encodes cyan and yellow fluorescent variants, respectively.

BD Living Colors EGFP Calibration Beads (Catalog no. 632394) were purchased from BD Biosciences. Sterile PBS and HEPES buffered saline (HBS) were prepared in-house. Trypsin-EDTA (Catalog no. Z-26-M) was purchased from c.c.pro GmbH.

Cell culture. A human bronchial epithelial cell line (BEAS-2B, ATCC) was grown in Earle's MEM supplemented with 10% FBS at 37°C in a humidified atmosphere, 5% CO_2 level. Cells were maintained at 85% confluence. Transfection was performed on both non-synchronized and synchronized cultures. A thymidine kinase double-block was performed to synchronize cells (Merrill 1998). Briefly, cells were cultured in synchronization medium (growth medium with 2mM thymidine) for 18 hours. Initial release was facilitated by an eight-hour incubation in regular growth medium. A second block was performed by incubating cells in

synchronization medium for another 17 hours, which was followed by a three-hour incubation in regular growth medium to allow mid-S-Phase transfection.

Transfection optimization. Culture and transfection conditions were optimized at different levels. Different media on which the cells could grow were initially screened for autofluorescence and CO_2 independence. Leibovitz's L-15 Medium was found to be ideal for growing cells with minimal autofluorescence levels. Transfection was optimized by varying the Lipofectamine^{TM}2000:DNA and PEI:DNA ratios. Relative efficiencies were determined by Fluorolog Fluorescence Spectrofluorometer (Data not shown). Highest transfection efficiencies were obtained when 2µl of Lipofectamine^{TM}2000 or a Nitrogen:Phosphate ratio (N:P) = 8 of PEI:DNA was used with 1µg of DNA. EGFP expression stability was determined using time-course experiments; expression without photobleaching was found to be stable for at least four hours.

Transfection. BEAS-2B cells were grown to 80% confluence from an initial seeding density of 1 x 10^5 cells/well in six-well plates at 37°C, 5% CO_2 level 24 hours before transfection. Cells were washed and the medium is replaced with 1 ml OptiMEM/well immediately before transfection. Optimized transfection procedures were performed using either Lipofectamine^{TM}2000 or PEI; 1 µg of pEGFP-N1 or pd2EGFP-N1, corresponding to $4 \cdot 10^{11}$ plasmids, is used for transfecting each batch of cells. For Lipofectamine-mediated transfection, separate 100 µl preparations of 1% w/v pEGFP-N1/OptiMEM or pd2EGFP-N1/OptiMEM and 2% v/v Lipofectamine^{TM}2000/OptiMEM were made at ambient temperature and allowed to stand for five minutes. The Lipofectamine transfection medium was prepared by adding the Lipofectamine solution to the plasmid solution. For PEI-mediated transfection, separate 100µl preparations of 1% w/v pEGFP-N1/HBS or pd2EGFP-N1/HBS and PEI (N:P = 8)/HBS were made and the solutions were combined by adding the PEI

solution to the DNA solution. Transfection media were allowed to stand for an additional 20 minutes at ambient temperature to facilitate complex formation. Cells were incubated with 200 µl/well Lipofectamine or PEI transfection medium for 3 hours at 37°C, 5% CO_2 level. Negative controls were transfected with non-EGFP containing plasmids using the same delivery systems. After 3 hours of incubation the medium was removed, and cells were washed with PBS. Cells were reincubated with Leibovitz's L-15 Medium with 10% FBS prior to EGFP expression monitoring.

Cotransfection. For the cotransfection experiments, we choose ECFP and EYFP plasmids on account of their similarities in terms of size (4731bp vs. 4733bp), promoter (early promoter of CMV), DNA replication origin (SV40) and translational efficiency (Kozak consensus translation initiation site). These plasmids only differ in the position of the MCS-site (C-terminal in ECFP, N-terminal in EYFP), apart from the fluorescence gene that they express, which are necessarily different to distinguish ECFP- from EYFP-expressing cells. Cotransfection was performed with two kinds of preparations containing the same molar amount per plasmid. For one preparation, ECFP/Lipofectamine, EYFP/Lipofectamine, ECFP/PEI and EYFP/PEI were complexed separately. For the other, a mixture of ECFP and EYFP hetero-complexeswere complexed with Lipofectamine or PEI. Transfection using either the hetero-complexes (pre-mixed) or a mixture of homo-complexes (post-mixed) were performed as previously described. Cells were reincubated in growth medium and CFP and YFP expression was monitored by fluorescence microscopy after 24 hours.

Instrumentation. EGFP expression was monitored using a motorized inverted microscope (Axiovert100M, Zeiss) equipped with a temparature-controlled mounting frame for the microscope stage. SimplePCI (Compix) was used for all microscope controls. A mercury light source (HB 100) was used for illumination, and a reflector slider with three different filter

blocks was used for detection. EGFP and d2EGFP were detected with filter set 41024 (Chroma Technology Corp., BP450-490, FT510, LP510-565), EYFP with F41-028 (AHF analysentechnik AG, BP500/20, FT515, BP535/30) and ECFP with F31-044 (AHF analysentechnik AG, BP436/20, FT455, BP480/40). An illumination shutter control was used to prevent bleaching.

Data acquisition and quantitative image analysis. Images were taken at 10× magnification, with a constant exposure time of 1s, at 10 minute-intervals for at least 30 hours post-transfection. Automated scanning and image capture for 25 view fields within a predefined 20mm diameter was programmed using the AIC module of SimplePCI. Briefly, the procedure captures individual brightfield images of the 25 view fields, then proceeds with a 30 hour loop where fluorescence images of each view field are taken at 10 minute intervals. At the end of the loop, brightfield images of the 25 view fields are taken again. Fluorescence images are consolidated into single image sequence files per view field. Negative control images were taken to assess lamp threshold values and autofluorescence, and were subtracted from corresponding image sequence files in SimplePCI to eliminate autofluorescence effects. To capture cell fluorescence over the entire sequence, regions of interest (ROIs) were manually defined around each cell in SimplePCI (Fig. 2). Changes in total gray measurements in individual ROIs were determined for each time point considered using a built-in function of SimplePCI.

EGFP quantification. EGFP is a Phe64Leu/Ser65Thr mutant of the *Auquorea victoria* protein, GFP (Heim et al. 1995). It exhibits an intense emission spectrum, higher photostability, and a half life of > 24 hours in mammalian cells (Bi et al. 2002); (Cotlet et al. 2001); (Tsien 1998). The accurate determination of the total number of EGFP molecules expressed in each cell is impeded by experimental limitations, notably bleaching and

autofluorescence. We optimized culture, microscope and image acquisition settings to maximize the dynamic range while minimizing the effects of bleaching and autofluorescence. Our calibration procedure involved the use of an EGFP standard originally designed for flow cytometry. We used the microbead population coated with $1.16 \cdot 10^5$ EGFP molecules. For each transfection experiment, 50µl Bead4 of the EGFP-Calibration Kit was suspended in 1ml HBS in one well of a six-well plate. Fluorescence images of the beads were taken under the same conditions as the transfected cells. For each experiment, total gray values were obtained for at least 80 beads and an equal number of background regions adjacent to each bead. The average background corrected total gray value is obtained and used for converting intensity values to molecules of equivalent soluble fluorochrome (MESF) units. In our experiments, average calibration factors typically do not exceed a deviation of 11.2%. Calibrations between different assays have higher deviations due to illumination variations. Consequently, the calibration factor and the number of expressed EGFP molecules that we report are not absolute as a result of inherent and non-quantifiable differences in the quantum efficiencies of the EGFP chromophores inside the cells and the beads. However, relative intensity levels and the intensity distribution functions that are discussed in the paper are unaffected by the absolute calibration.

References (Supplementary Online Information)

Bi XJ, Wirth M, Beer C, Kim EJ, Gu MB, Zeng AP. 2002. Dynamic characterization of recombinant Chinese Hamster Ovary cells containing an inducible c-fos promoter GFP expression system as a biomarker. J Biotechnol 93(3):231-242.

Cotlet M, Hofkens J, Maus M, Gensch T, Van der Auweraer M, Michiels J, Dirix G, Van Guyse M, Vanderleyden J, Visser A and others. 2001. Excited-state dynamics in the enhanced green fluorescent protein mutant probed by picosecond time-resolved single photon counting spectroscopy. J Phys Chem B 105(21):4999-5006.

Heim R, Cubitt AB, Tsien RY. 1995. Improved Green Fluorescence. Nature 373(6516):663-664.

Merrill G. 1998. Cell Synchronization. Methods Cell Biol 57:229-249.

Tsien RY. 1998. The green fluorescent protein. Annu Rev Biochem 67:509-544.

4 Phenotypic Switching

Not all individuals of a species are the same. For example there may be variation in size, shape or other aspects of appearance or in behavioral features. Often, the reason for this are differences in the genetic information, that is, the *genotype*. With different genetic information comes a different protein expression pattern, as detailed in Section 2.1.2, which results in different characteristics, all of which together are called the *phenotype*. However, even two bacteria which have been created by (asexual) cell division from a single ancestor, and therefore share an identical genome, may have different physical properties and hence different phenotypes.

Of course, the inherent fluctuations in the molecular numbers cause small and transient heterogeneity across a clonal cell population. However, phenotypic traits differ qualitatively from these microscopic fluctuations, as they are substantial and stable for longer periods of time. This is a consequence of the underlying genetic regulations, as shortly addressed in Section 2.2.3. At the end of this chapter we will consider a specific example: Competence regulation in *Bacillus subtilis* and in particular how individual cells switch from a non-competent to a competent phenotype.

4.1 The Competent Phenotype of Bacillus subtilis

The soil-dwelling bacterium *Bacillus subtilis* is widespread in nature and a model organism for laboratory experiments. It was first described by Ehrenberg in 1835 [115] and has since then been extensively studied, especially to elucidate gene regulation, cell cycle events like chromosomal replication, and phenotypic differentiation like sporulation and competence development [116].

Differentiation of clonal populations of *B. subtilis* is most apparent upon stressing the cells by deprivation of nutrients. Under these conditions cells first enter a highly motile state, as necessary for the search of new food sources. If this fails, a fraction of the cells becomes *competent* for the transformation of exogenous DNA, which constitutes an example of horizontal gene transfer, cf. Section 3.1. Competence requires the costly production of a DNA-uptake machinery and further proteins, which are responsible for incorporating newly acquired DNA into the bacterial genetic material. Competence has probably evolved to acquire useful abilities, like the digestion of a wider array of nutrient sources, and thereby survive in the stressful environment. It is obviously a costly investment, not only due to the metabolic costs, but also due to the possibilities of incorporating harmful genes or misplacement of incorporated gene sequences. Therefore, it immediately makes sense, that not all cells become competent, but that competence development is instead strongly regulated, with respect to both when

cells may become competent and how probable this is. As a last resort cells which do not become competent can form spores and thereby ride out the lean times in a dormant state.

4.2 The Advantage of Multiple Phenotypes

During the lifetime of an organism it might move between different environments, or the environment might change due to some external factor. To ensure survival in the everlasting struggle for resources, it appears necessary to adapt to a given environment as fast as possible by changing to some optimal (with respect to the present condition) phenotype. Astonishingly, even for constant external conditions, clonal populations of microorganisms tend to show phenotypic heterogeneity. From an evolutionary standpoint this seems strange: Under favorable conditions microbes grow exponentially and a population where all individuals have the optimally reproducing phenotype would quickly outgrow a population exhibiting phenotypic heterogeneity. Why put part of the population in a state in which reproduction is not optimal? Possible solutions to this seeming paradox are the costs (in terms of resources and time) of switching between phenotypes and possible risks of being in a phenotypic state which is not adapted to the environment at all. The idea is that evolution has optimized not every single organisms, but the species as a whole, with respect to a changing environment [117]. While under one condition one phenotype may thrive, it may be severely hampered upon an environmental change and even may perish. Therefore, phenotypic diversity can be seen as a strategy to minimize risk, also known as bet-hedging.

4.3 Stochastic Switching

4.3.1 Auto-feedback: Phenotype-encoding through Non-linear Reaction Networks

Many bacterial systems show phenotypic variability [118]. Often these systems can be modeled by rate equations that have more than one steady state solution. An early example is a genetic control system where a gene's protein induces stronger expression of the gene itself [119]. This is known as an auto-feedback loop, a very simple and widespread mechanism which relies on the non-linearity of the activating feedback. The dynamics can be described by coupled differential equations:

$$\dot{M} = \frac{\beta_M}{1+(Q/P)^n} - \delta_M M \,, \tag{4.1a}$$

$$\dot{P} = \beta_P M - \delta_P P \tag{4.1b}$$

Here M is the concentration of the mRNA, while P is the concentration of the corresponding protein. Auto-feedback is ensured through steplike activation of transcription with cooperatively n, half-maximal activation at $P = Q$ and maximal transcription rate β_M. Translation is linear in mRNA concentration at rate β_P, while δ_M and δ_K are degradation rates of mRNA and proteins, respectively. Usually, equations (4.1) have three fixed points, one unstable and two stable ones (with low and high protein concentration, respectively). The stable fixed

points can be associated with different phenotypes, which are characterized by the concentration of protein P.

4.3.2 Switching Mediated by Stochasticity in Gene Expression

While in reality one rarely has such a simple system, one may still take this as a paradigm. It explains how different phenotypes can be encoded. Still, another ingredient is necessary to explain how a homogeneous population may split up into different subpopulations: intrinsic fluctuations. When taking small number fluctuations (see Chapter 2.3) into account, it is possible that protein numbers in a given individual fluctuate from one stable fixed point across the unstable fixed point into the basin of attraction of the other stable fixed point. This is easily shown by stochastic simulations (cf. Section 2.4.3), but may also be argued for analytically in the case of the auto-feedback loop and other simple network motifs [4].

Obviously switching from one expression pattern to another requires sufficiently strong fluctuations. This is can be achieved by feed-forward cascades. Here, one molecular species or compound triggers the enhanced production of another molecular species. If the former is subject to large relative fluctuations in the number of molecules, these are passed on to the latter, which may then display large fluctuations in its absolute number of molecules, known as bursts. Examples include eukaryotic genes under strong transcriptional control, which exhibit slow transitions between active and inactive states [120]. For bacteria the mRNA number is often very small and thus protein production is burst-like [5], which makes the explicit modeling of mRNA necessary to understand phenotypic switching [121].

Often, phenotypic switching needs to be fast to adapt to rapid changes of a stressful environment. At the same time switching might be risky, as the environment might change back, leaving the cell badly adapted. Under these conditions, the fraction of the population which actually switches, needs to be tightly controlled. A good bet-hedging strategy is to control the number of switching individuals by transiently enhancing noise through the mechanisms mentioned above. This mode of control of stochastic switching between phenotypes is termed a switching window [122].

In the manuscript entitled “Kinetics of Genetic Switching into the State of Bacterial Competence” by M. Leisner, J.-T. Kuhr, J. O. Rädler, E. Frey, and B. Maier, single cell time-lapse data of *B. subtilis* bacteria, switching to the competent phenotype, is explained by a non-linear, stochastic model. Quantitative agreement between theory and experiment is found for both the switching probability and the switching time. The consequences of a knock-out mutation are linked to a bifurcation in the fixed points of differential equations describing the underlying non-linear reaction network.

Kinetics of Genetic Switching into the State of Bacterial Competence

Madeleine Leisner,[†‡] Jan-Timm Kuhr,[§] Joachim O. Rädler,[‡] Erwin Frey,[§] and Berenike Maier[†*]
[†]Institut für Allgemeine Zoologie und Genetik, Westfälische Wilhelms Universität, Münster, Germany; [‡]Department für Physik, Ludwig-Maximilians-Universität München, München, Germany; and [§]Arnold Sommerfeld Center for Theoretical Physics and Center for NanoScience, Department für Physik, Ludwig-Maximilians-Universität München, München, Germany

ABSTRACT Nonlinear amplification of gene expression of master regulators is essential for cellular differentiation. Here we investigated determinants that control the kinetics of the genetic switching process from the vegetative state (B-state) to the competent state (K-state) of *Bacillus subtilis*, explicitly including the switching window which controls the probability for competence initiation in a cell population. For individual cells, we found that after initiation of switching, the levels of the master regulator [ComK](t) increased with sigmoid shape and saturation occurred at two distinct levels of [ComK]. We analyzed the switching kinetics into the state with highest [ComK] and found saturation after a switching period of length 1.4 $\pm$ 0.3 h. The duration of the switching period was robust against variations in the gene regulatory network of the master regulator, whereas the saturation levels showed large variations between individual isogenic cells. We developed a nonlinear dynamics model, taking into account low-number stochastic effects. The model quantitatively describes the probability and timescale of switching at the single cell level and explains why the ComK level in the K-state is highly sensitive to extrinsic parameter variations. Furthermore, the model predicts a transition from stochastic to deterministic switching at increased production rates of ComK in agreement with experimental data.

INTRODUCTION

Populations of genetically identical cells often maintain a diversity of phenotypes, characterized by different patterns of gene expression. This is usually triggered by stochastic fluctuations, which are amplified by the underlying gene regulatory networks (1–4). The benefits of such non-genotype-derived heterogeneity lie in the enhanced adaptability to environmental changes of the population as a whole (5–8). To analyze phenotypic heterogeneity it is necessary to monitor gene expression in individual cells (9). Real-time kinetics of gene expression have been extensively measured in individual cells to characterize noise in gene expression (10–14), multistability (1,5,15,16), oscillations (16–19), and timing of gene activities (20), but the determinants for genetic switching kinetics are not well characterized so far.

Competence development in *Bacillus subtilis* is one example in which a genetic switch determines cell fate. At low cell density, a homogeneous cell population undergoes exponential growth, but at high cell density (stationary growth phase), the cell population becomes heterogeneous in its phenotype, with a well-defined fraction of 15% expressing genes that code for a strong DNA import machine and recombination proteins (21). These cells are called competent for DNA transformation and they express the master regulator *comK* at high level (22–24). In this study, the state in which *comK* expression levels are high is denoted the K-state. The entry into the K-state is switchlike (25,26). The positive feedback loop in the genetic control circuit is important for the establishment of the competent phenotype in *Bacillus subtilis* (27,28). In noncompetent cells, the positive autoregulatory loop is not activated and *comK* expression is low (B-state). During exponential growth, ComK is kept at a basal level through degradation by the MecA/ClpC/ClpP protease complex, and by transcriptional repressors including Rok, AbrB, and CodY (Fig. 1). Due to a quorum-sensing mechanism, the concentration of ComS (an inhibitor of MecA/ClpC/ClpP) rises with increasing cell density. Work by Maamar et al. (29) revealed that at T_0 (i.e., at the entry into stationary phase, $T_x=x$ h after transition point) the average number of mRNA coding for ComK per cell is of order 1. In this regime, small number fluctuations are, relative to the mean, of paramount importance. This is especially noteworthy since the reaction kinetics of ComK is highly nonlinear. Experiments and simulation have shown that the fraction of cells that switch into the K-state is determined by the magnitude of intrinsic fluctuations in *comK* expression (16,29). The second important determinant of the fraction of cells in the K-state is the length of a switching window in which basal *comK* expression rate is enhanced, which facilitates switching (30). Under conditions in which nutrient concentrations are constantly low, cycles of competence initiation and decay have been observed in real-time experiments in individual cells and a mathematical model described the system as an excitable regulatory circuit (15,16). Escape from the K-state has been attributed to negative feedback between ComK and the inhibitor of ComK proteolysis, ComS. Theoretical models of the competence decision system can be divided into two different categories, by the description of the system as excitable (15,16,31) as opposed to bistable (5,27–29,32).

Submitted May 9, 2008, and accepted for publication October 15, 2008.

*Correspondence: maierb@uni-muenster.de

Editor: Herbert Levine.

0006-3495/09/02/1178/11 $2.00

doi: 10.1016/j.bpj.2008.10.034

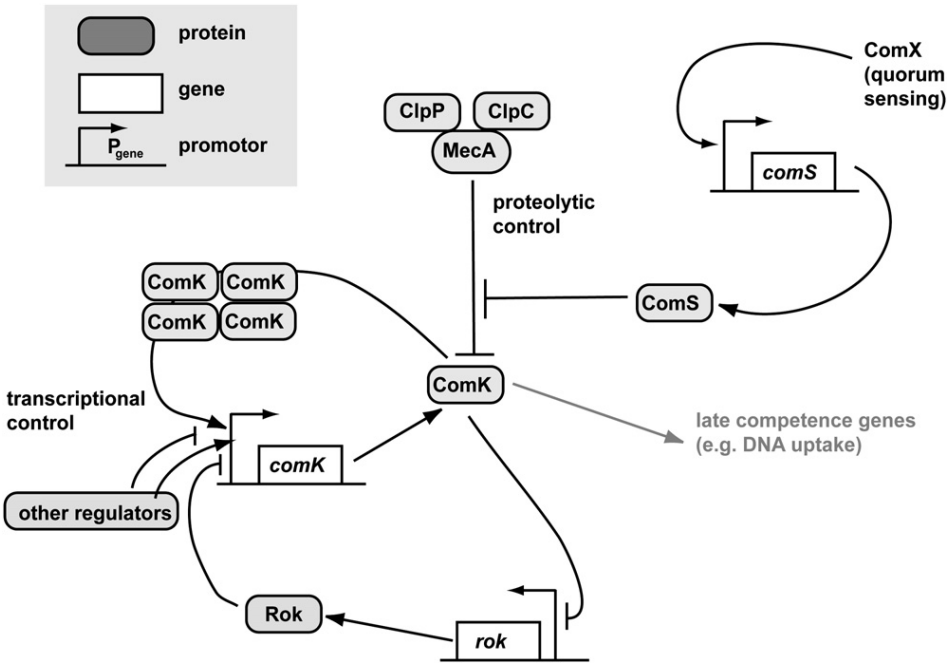


FIGURE 1 Core of the competence circuit. Arrows denote upregulation, blunt ends denote downregulation.

Here we investigated the switching kinetics of individual cells of *Bacillus subtilis* both experimentally and theoretically. We quantified the expression rate of the promoter of *comK* in real-time and mimicked increasing cell density with individual cells by continuously adjusting the growth medium. Using this method, we investigated the variability in the length of switching period, switching kinetics, and accumulated *comK* expression against varying external conditions due to growth phase and against genetic modification of the competence circuit. We developed a mathematical model predicting bistability during cell growth that retrieved the length of the switching period and the fraction of competent cells. The model explains why the saturation level of the K-state is highly sensitive to external fluctuations as observed in the experiments. Transition to deterministic switching, i.e., independence of fluctuations, as found for a mutant strain with increased transcription rates, is neatly explained by loss of the B-state in the stationary phase.

MATERIALS AND METHODS

Materials and strains

All *B. subtilis* strains used were derived from strain BD630 and are described in Table S1 in Supporting Material. *Bacillus* strains were grown in liquid competence medium (33) supplemented with glucose (0.5%), L-histidine, L-leucine, L-methionine (50 μg/ml), and chloramphenicol, kanamycin (5 μg/ml), or spectinomycin (100 μg/ml) at 37°C. Competent cells were prepared as described previously (33). BM101 was created by transformation of genomic DNA of BD2711 into BD2955 ((28,41) (Table S1)).

Microscopy

Cells were taken at $T_0 \pm 1$ h (where T_0 is the transition from the exponential to the stationary growth phase, $T_x = x$ h after the transition point), permitted to attach to a polystyrene-coated cover slide and mounted onto a flow chamber. An image was acquired every 15 min. To confirm that the conditions under the microscope were equal to the conditions in an Erlenmeyer flask, the medium was exchanged directly after every exposure for the supernatant of a parallel culture grown in an Erlenmeyer flask (shaken at 300 rpm, at 37 °C).

Microscopy was performed with a Axiovert 200 M microscope (Carl Zeiss, Jena, Germany) equipped with a digital camera (Andor, Belfast, UK), and an EC Plan-Neofluar 100×/1.30 oil immersion objective (Zeiss). Andor software was used for image acquisition. Stability of absolute fluorescence values was verified using a microscope image intensity calibration kit (Focal-Check fluorescence microscope test slide #3, F 36914; Molecular Probes, Eugene, OR). Homogeneity of illumination was tested using fluorescent slides and the maximum deviation was <5%.

Analysis

To analyze the ComK protein levels in single cells, we used strains containing a *gfp-comK* reporter construct standing under the control of the promoter of *comK* (P_{comK}) in addition to the original copy of the *comK* gene. As the green fluorescent protein (GFP) molecule is not likely to be a substrate for MecA/ClpC/ClpP proteolysis, the concentration of GFP controlled by P_{comK} is not a direct measure of ComK. However, the expression rates, i.e., the production rate of ComK and GFP, are expected to be similar, as the complete native promoter of *comK* including all native signals for *comK* expression is present. We analyzed ComK expression in single cells as the fluorescence intensity (FI) represented by the mean shaded value (measured in arbitrary fluorescence units (FU)).

Images were processed using ImageJ software (National Institutes of Health, Bethesda, MD). The image background was corrected using a rolling ball algorithm with a radius of 50. An intensity threshold tool was used to delimit the boundaries of the cells in the bright-field image (Fig. 2 *a*). The areas encircled by the boundaries were then selected as a region of interest using the ImageJ wand tool. These regions of interest were then applied to the fluorescence images and the mean shaded value of each cell was obtained using the ImageJ measurement tool.

The autofluorescence level of noncompetent BD630 was $FI_{auto\ T-2} = 23 \pm 1$ FU, and the width of the Gaussian fit was $FI_{auto\ T-2} = 9.2 \pm 1$ FU. We verified recently that the autofluorescence level did not shift until the basal fluorescence level reached its maximum (30). Since the time delay between the expression of the GFP reporter and the onset of fluorescence is 5 min (30), but switching into the K-state occurred on a significantly longer

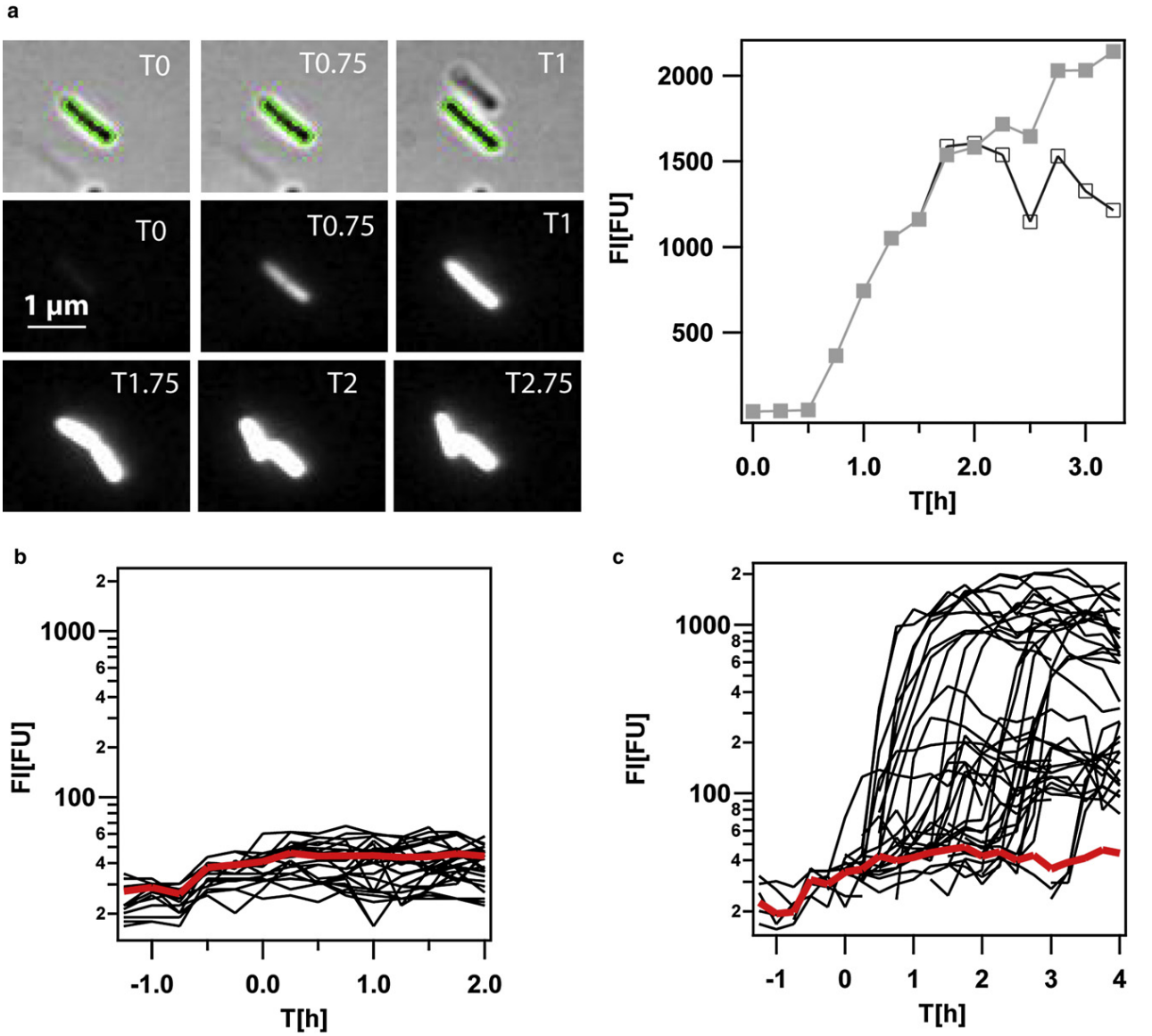


FIGURE 2 Time course of P_{comK}-gfp expression in individual cells. (*a*) Time-lapse of a switching cell and corresponding plot of fluorescence intensity FI as a function of time. (*Green contour*) Outline of the cell. Scale bar 1 μm. (*b*) Time course of fluorescence intensity of cells remaining in the B-state. Red line indicates the average fluorescence intensity of cells remaining in the B-state. (*c*) Time course of fluorescence intensity of cells switching into the K-state. Red line indicates the average fluorescence intensity of cells still in the B-state before switching.

timescale, we did not correct our data for GFP-folding. In addition, no significant GFP degradation could be found on the relevant timescale of 3 h (30).

We defined the switching threshold differently as in previous work with fixed cells (30). Instead of defining the threshold from the histogram of fluorescence intensity, we defined the switching time as the point where the intensity increased sharply in the $FI(t)$ plot (Fig. 2 *a*). This led to slightly lower values for the threshold but the fraction of competent cells was not affected (Fig. S4).

RESULTS

The basal *comK* expression rate increases within the whole cell population

We quantified the expression of the master regulator *comK* in individual cells by measuring the fluorescence intensity of GFP fused to the promoter of *comK* (Fig. 2) (see Materials and Methods). To stimulate growth-phase-dependent development of cells on the microscope and to supply the cells with oxygen, we exchanged the medium in the flow chamber with medium from a cell culture growing under standard competence conditions every 15 min. We quantified the expression from the *comK* promoter by measuring the average fluorescence intensity *FI* of individual cells (Fig. 2 *a*). In agreement with previous experiments, we found that in our real-time experiments 15% of the cells switched into the K-state (Fig. S4). Initiation of switching was defined as a sharp increase in fluorescence intensity, which was followed by a sigmoidal increase in fluorescence intensity as a

function of time. Before entry into the stationary phase at T_0, basal expression of *comK* increased significantly in all cells between $T_{-0.75}$ and $T_{-0.5}$ (Fig. 2, *b* and *c*). The basal level of *comK* expression was indistinguishable for cells that later triggered the switch and cells that remained in the B-state, indicating that no preselection had taken place as yet.

A fraction of cells switches into a state with intermediate *comK* expression

The saturation levels showed a large variation between individual cells with two distinct peaks of saturation fluorescence intensity (Figs. 2 *c* and 3 *a*). An intermediate peak was distributed at ~200 FU, i.e., at a level between the basal fluorescence level at 50 FU (B-state) and high fluorescence level (K-state) at 1300 FU. We did not observe switching between the intermediate fluorescence mode and the K-state. Furthermore, we found no correlation between saturation fluorescence intensity and basal fluorescence, daughter cells after cell division, or cell length (data not shown). The intermediate fluorescence peak disappeared in the *rok*-strain (Fig. S5) in which the major repressor of *comK* transcription, Rok, was knocked-out. Since the characterization of functionality and molecular determinants for switching into the intermediate fluorescence mode is subject to future experiments, we decided to analyze the kinetics of switching into the K-state in the following, i.e., we concentrated on cells that obtained saturation fluorescence values exceeding 300 FU.

The length of the switching period of individual cells is well defined and independent of growth phase

To quantify the switching kinetics, we normalized the individual switching curves to their saturation level and shifted the *T*-axis to overlay all curves at half-maximum expression level at $\tau_{1/2}$ (Fig. 3 *b*). To investigate whether the switching kinetics were dependent on growth phase, we sampled switching curves between $T_{-0.25}$ and $T_{2.5}$. The overlay showed very little variation, indicating that the duration of the switching period is well defined and independent of growth phase (data not shown). The expression rate *exr* was defined as the first derivative with respect to time of the fluorescence intensity, FI (Fig. 3 *c*). Please note that the cells did not grow during ≈ 1 h after the onset of switching (M. Leisner, K. Stingl, and B. Maier, unpublished data), and therefore we did not correct the expression rate for cell length. However, we attribute the negative expression rate at $T - \tau_{1/2} > 1$ h to cell growth, since GFP fluorescence was stable within several hours (Materials and Methods). We defined the length of the switching period ρ as the time during which *exr* exceeded $0.1 exr_{\max}$. The average period for switching into the K-state was $\rho = 1.4 \pm 0.3$ h, with a standard deviation of 0.3 h (Fig. S7 *a*). To address the question whether the length of the switching period was well defined in individual cells, we plotted the switching period ρ as a function of growth phase T and found no significant variation (Fig. S7 *b*).

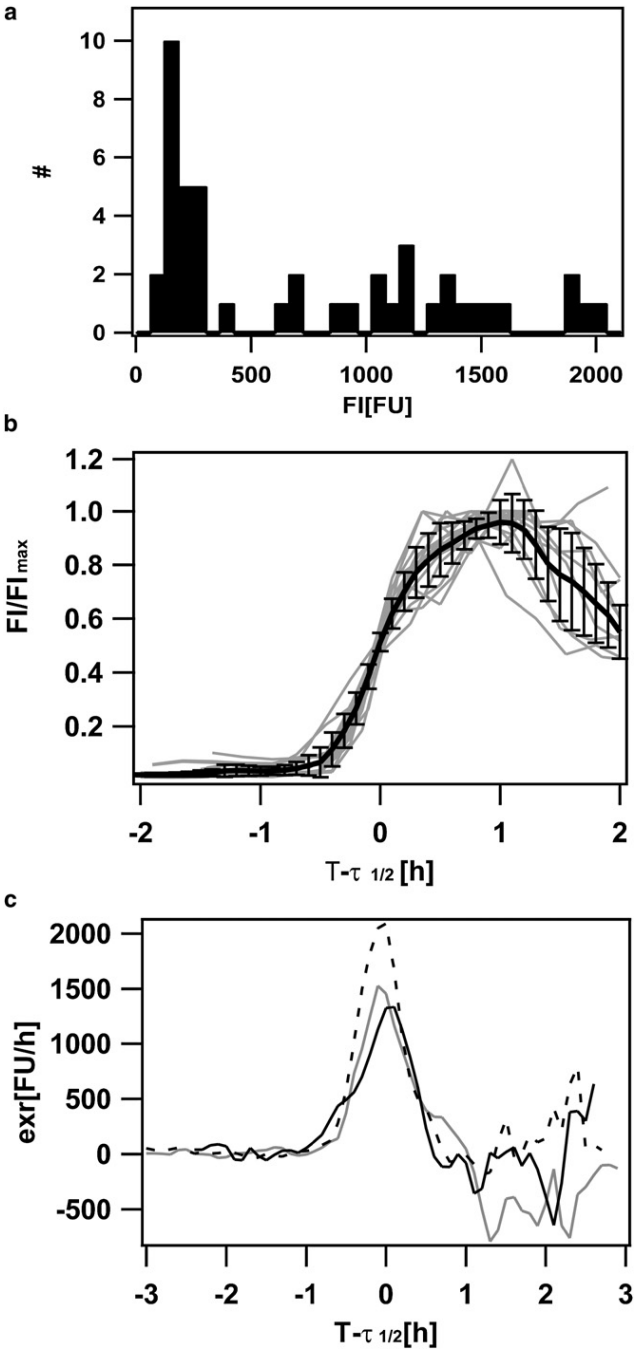


FIGURE 3 Switching characteristics of wild-type (wt) cells. (*a*) Distribution of maximum fluorescence intensity in the wild-type (BD2711). (*b*) The fluorescence intensity in 28 individual cells was normalized to the cumulative expression (maximum fluorescence intensity). The time axis was shifted to $\tau_{1/2}$, where cells had half-maximum fluorescence intensity. (*Solid line*) Averaged fluorescence intensity; (*shaded lines*) individual cells. (*c*) Expression rates *exr* of wt (*shaded*), *rok*-strain (*dotted*), and ComS (*solid*) overproducing strain. Expression rates were obtained by multiplication of the normalized values with the average maximum value of 28–50 cells.

With increasing ComS concentration, the probability for proteolysis of ComK by the MecA/ClpC/ClpP complex decreases, leading to increased ComK levels. We investigated whether the levels of ComS influenced the kinetics of switching using a ComS overproducing strain *comS++*. Although the fraction of switching cells increased from 15% to 85%, the switching kinetics of the individual bacteria did not alter significantly as compared to the wild-type (Fig. S6). In particular, the expression rate *exr* was comparable to the wild-type (Fig. 3 *c*).

A two-variable model elucidates the switching process

Nonlinear dynamics together with fluctuations can lead to bimodal protein distributions (34). Here we analyze a fully stochastic model for competence development. It incorporates production and degradation of both ComK proteins (K) and corresponding mRNA molecules (M). Mathematically the model is captured by a Master equation; a formal description of the model in terms of transition rates is given in the Supporting Material. Since analytical solutions are not feasible, we analyze the model by employing stochastic simulations (see below).

To give a frame of reference for the dynamics of competence development, and to explain the emergence of distinct phenotypes, it is instructive to analyze the equations:

$$\frac{\partial \mathrm{K}}{\partial t} = \beta_{\mathrm{K}}\mathrm{M} - \frac{\delta_{\mathrm{K}}\mathrm{K}}{q_{\mathrm{K}} + \mathrm{S} + \mathrm{K}} - \delta_0\mathrm{K}, \qquad (1)$$

$$\frac{\partial \mathrm{M}}{\partial t} = \alpha_{\mathrm{M}} + \frac{\beta_{\mathrm{M}}}{1 + (p_{\mathrm{K}}/K_{\mathrm{f}})^{\gamma_{\mathrm{K}}}} - \delta_{\mathrm{M}}\mathrm{M}. \qquad (2)$$

While these equations disregard stochastic effects (see further below), they are easily discussed by a nonlinear dynamics analysis.

The first term in Eq. 1 represents translation of K, which happens with rate β_{K} for each M. The second term describes degradation of K according to Michaelis-Menten kinetics. For very high K this happens at a maximal rate δ_{K}, and half-maximal degradation of K is found where $\mathrm{K} = \mathrm{S} + q_{\mathrm{K}}$. S corresponds to the number of ComS (S) proteins which compete with ComK for degradation by the MecA/ClpC/ClpP protease complex (see Supporting Material). Representing the effect of quorum sensing, $S(t)$ communicates the cell density which rises sigmoidal with time to the individual cell. In our model, $S(t)$ takes the role of an externally determined control function (Fig. S1). The third term covers linear degradation of ComK with rate $\delta_0 << \delta_{\mathrm{K}}$, which must always be present owing to nonspecific degradation/denaturation of ComK. This term is important only for large K and ensures that K cannot rise arbitrarily, as is biologically necessary. In Eq. 2, α_{M} is the basal transcription rate in absence of K. The next term describes autocatalytic feedback of ComK: Two ComK dimers bind cooperatively to the *comK*-promoter (35,36), thereby strongly activating transcription. Assuming dimer formation and dissociation as well as promoter binding and unbinding are fast enough to equilibrate, autocatalytic transcription can be modeled by a Hill function, with maximal (for large K) transcription rate β_{M}, half-maximal concentration p_{K}, and cooperativity γ_{K}. The value K_{f}, appearing in the denominator, is the number of free ComK, i.e., not bound to the MecA/ClpC/ClpP protease complex (see Supporting Material). Finally, degradation of mRNA is proportional to M and the rate δ_{M}. For further details on the model and its parameters, see Table S2.

Using nonlinear dynamics analysis (37) and literature data to set parameters, we investigated whether our two variable model (Eqs. 1 and 2) could reproduce the switching behavior. For low numbers of ComS we find only one stable fixed point corresponding to the vegetative state ($\mathrm{K} \approx 100$, $\mathrm{M} \approx 1$) (Fig. 4 *a*). As S rises, a saddle-node bifurcation appears, creating a further stable fixed point with higher K and M values and an intermediate unstable fixed point. This, in principle, allows cells to stochastically switch to the state with higher K and M. However, due to the high activation threshold, this is highly unlikely. At $S \sim 1500$, the lower stable and the unstable fixed point almost meet in another bifurcation. The lower stable fixed point (now shifted to $\mathrm{K} \approx 200$, $\mathrm{M} \approx 1$) corresponds to the B-state. The intermediate unstable fixed point at $\mathrm{K} \approx 600$, $\mathrm{M} \approx 3$ is very close to the B-state (Fig. 4 *b*). Under these conditions, intrinsic noise is strong enough to carry some cells across the switching threshold. A sample path in phase space is given in Fig. S2, illustrating escape from B-state and passing of threshold. Cells that cross the threshold evolve toward the upper stable fixed point (K-state, $\mathrm{K} \approx 7000$, $\mathrm{M} \approx 10$), i.e., bacteria switch into the K-state. Back-transitions from the K-state into the B-state are highly improbable, because in the K-state, fluctuations are too small compared to the distance between the K-state and the threshold to revert switching.

Noise-driven switching requires vicinity to region with quantitatively different dynamics

The above scheme describes the evolution for a standard wild-type cell as it evolves from the low cell density exponential growth phase to the high cell density stationary phase. In that sense, it can be considered as averaged description, while for single cells reaction rates may vary, which is usually termed extrinsic noise (13). This can have major influence on the single cell dynamics to the point of loss of stochastic switching. To explore the consequences of variation in the rates we performed a stability analysis by examining the topology of nullclines and fixed points of the model. If the topology changes, one expects to find qualitatively different dynamics to that described above. Fixed points of the dynamics may emerge or disappear in bifurcations, changing the switching behavior significantly. Since

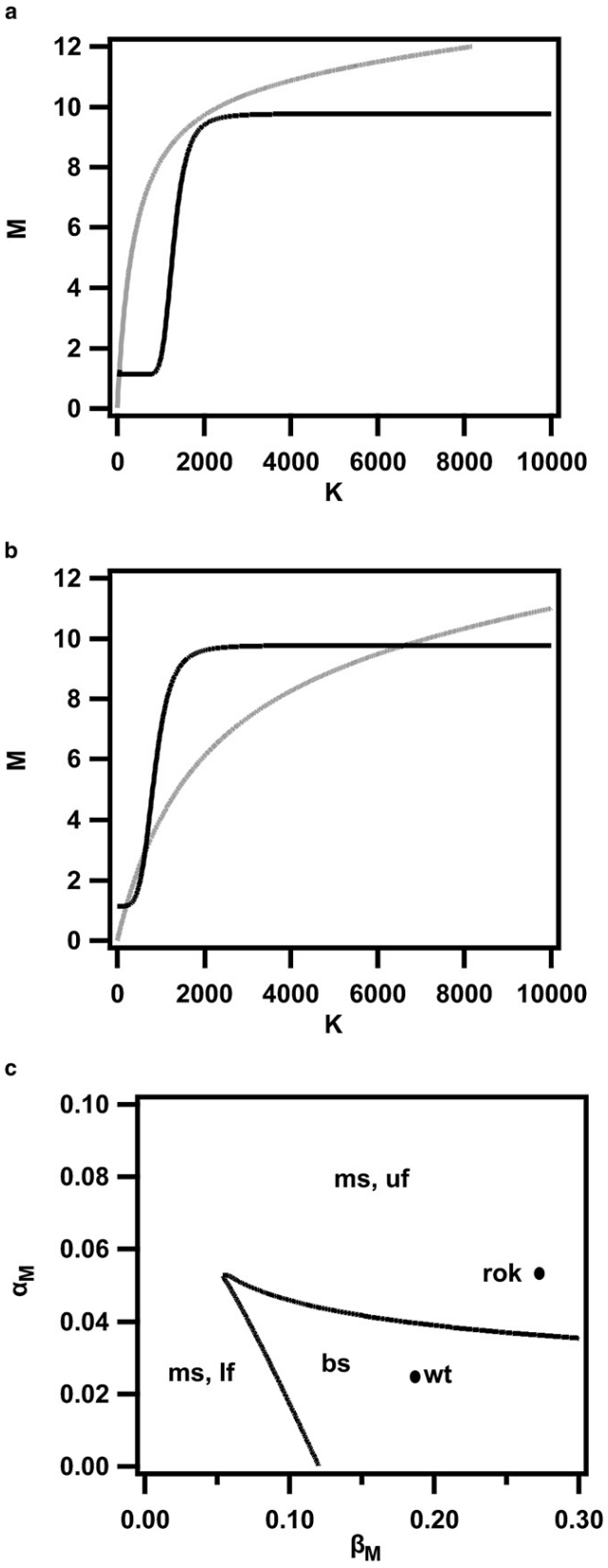


FIGURE 4 Nonlinear dynamics model. (*a* and *b*) Bifurcation analysis. Nullclines of Eqs. 1 and 2. (*a*) S=0. (*b*) S=1500 are plotted (*solid*, dM/dt=0; *shaded*, dK/dt=0). (*c*) A stability diagram, α_M versus β_M is plotted, monostable (*ms*) and bistable (*bs*) regions with only upper (*uf*) or only lower (*lf*) are indicated. Parameter combinations of the model for the wild-type (*wt*) and *rok*-strain (*rok*) are indicated.

the expression rates are experimentally accessible, we show the stability diagram (a cut through the parameter space) for variation in the basal transcription rate α_M and the transcription rate with feedback β_M (Fig. 4 *c*). Similar stability diagrams can be drawn for any pair of parameters.

A general result for bistable systems exhibiting stochastically induced switching is that they are located close to lines that separate regions with different topology. This is necessary, since only then are random fluctuations frequent and wide enough to enable a trajectory (an individual cell in our case) to cross the threshold with a sufficiently high probability. It follows that the switching probability per unit time is sensitive to the distance to the line of separation. This feature is likely to be present in systems similar to ours (38).

Effects of perturbing the promoter activity of *comK* on initiation probability, switching kinetics, and saturation levels

The stability diagram (Fig. 4 *c*) predicts the loss of bistability at increased transcription rates of *comK*. Previous experiments using a *rok*-strain showed that the *rok*-strain has a twofold increased basal expression (29,30,39). We further measured that the maximum expression rate during switching increased ~1.5-fold (Fig. 3 *c*). We therefore analyzed how the dynamics of the model qualitatively changes when α_M and β_M vary strongly. For the *rok*-strain, one would expect $\alpha_M \rightarrow 2\ \alpha_M$ and $(\alpha_M + \beta_M) \rightarrow 1.5\ (\alpha_M + \beta_M)$. In this regime, as the stationary phase is reached ($S \sim 1500$), the system has lost its lower and intermediate fixed point, corresponding to B-state and threshold. This means achievement of competence is independent of fluctuations and is initiated in all cells. In this altered reaction network a transition from a bistable system with stochastic switching to a monostable, deterministic system occurs.

Experimental data showed indeed that 99 ± 3% of all cells reach high FI for the *rok*-strain (30). We further investigated experimentally whether deletion of the major repressor protein of *comK* transcription, Rok, affected switching characteristics (Fig. S5). The sigmoidal shape of [ComK](t) and the switching period of $\rho = 1.4$ h were not altered, while the maximum expression rate and the cumulative expression increased by a factor of ~1.5 as compared to the wild-type strain (Fig. 3 *c* and Fig. S5 *a*).

Extrinsic noise determines the saturation value of individual cells in the K-state

The observed spread in saturation levels (Figs. 2 *c* and 3 *a*) is explained by extrinsic noise, which is present in any cell population. This does not alter the behavior of the population as a whole but affects every individual bacterium differently.

Besides altering dynamics of a few cells qualitatively (as discussed above), the value of K at the upper fixed point is highly sensitive to these fluctuations. This results from the fact that the nullclines in Fig. 4 *b* are close to parallel at the upper fixed point. Variations in the model's parameters thus shift the K-state to higher or lower values of K.

We quantified the effect of variation of the parameters on the saturation value of ComK. Computing the relative change of K at the upper fixed point due to a relative change in a given parameter yields the sensitivity (see Supporting Material). We find that variation in the overall degradation rates δ_M and $(\delta_K + \delta_0)$ alter the fixed point's position by the same magnitudes, but opposite sign as the overall production rates $(\alpha_M + \beta_M)$ and β_K. By comparison, the half-maximal degradation of ComK, q_K, has a minor effect on the fixed point's position, while variations in the other parameters can be neglected (see Table S2).

Stochastic simulations with well-motivated parameters fortify predictive character of the model

Quantitative data, obtained by stochastic simulations (40), corroborated the predictive value of the model (Fig. 5 and Fig. S2). As found in earlier experiments of the whole bacterial population (30), single cell time-lapse microscopy revealed a transient increase in *comK* transcription α_M before T_0 (Fig. 2, *b* and *c*), which we introduced in our simulations (Fig. S1).

Extrinsic noise was introduced in the stochastic simulations by choosing model parameters from a Gaussian with a standard deviation of 5% about their mean for each realization (see Supporting Material for details). Theses variations in the model parameters can change number and positions of fixed points. The former can be addressed analytically by stability analysis (see above, Fig. 4 *c*), the latter by computation of sensitivities, as was done for the position of the upper stable fixed point (Table S2).

Before entry into stationary phase, i.e., at values of $S \ll 1500$, cells showed fluctuations about the vegetative state, but no escape was possible. Approximately 14.7% of all cells entered the K-state, showing saturation in ComK level within $\rho = 1.2$ h, in accordance with experimental findings of ~15% competent cells and a length of the switching period of ρ ~1.4 h. Further, we find a wide spread in saturation levels and expression curves very much like those recorded in the experiment (Fig. 2 *c*), indicating that extrinsic noise is an important determinant of ComK levels.

These findings show that an effective two-component model with well-motivated parameters is not only sufficient to explain switching dynamics qualitatively but also returns quantitative data in excellent agreement with the experiment.

DISCUSSION

Competence development in *B. subtilis* provides an interesting regulatory network to study the kinetics of genetic switching between different phenotypes. We developed an experimental approach and a stochastic nonlinear dynamics model to quantify the switching behavior of individual cells of *B. subtilis* into the K-state in response to quorum sensing. Both experiment and theory reveal that the switching period is a robust property of the genetic switch whereas the saturation levels show strong variability most likely due to external fluctuations.

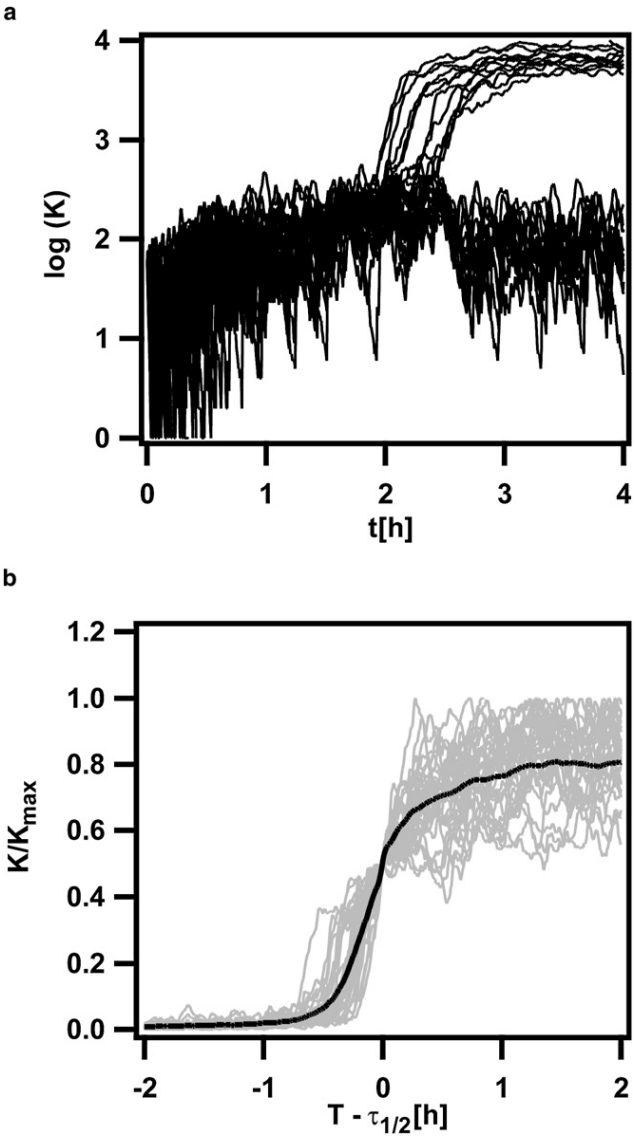


FIGURE 5 Stochastic simulations. (*a*) 100 realizations of the stochastic simulation of the model are shown. (*b*) Time courses of cells that entered the K-state were normalized to the maximal value. The time axis was shifted to $\tau_{1/2}$, where cells had half-maximal *K* values; compare to Fig. 3 *b*.

Rise in basal expression from the *comK* promoter does not select cells for switching

In agreement with previous studies that concentrated on cell populations, we found that the basal expression rate of

ComK increased before entry into stationary phase (30). Increased expression rate of *comK* may be explained by decreasing expression of the repressor AbrB and by deactivation of the GTP-dependent repressor function of CodY. Using real-time experiments with individual cells, we were able to distinguish the rise in basal expression rate between cells that switched eventually after entry into the stationary phase and cells that did not switch before the switching window closed. The magnitude of increase was similar in cells that remained in the B-state and cells that switched into the K-state. This result indicates that the sharp rise in the expression rate of the *comK* promoter upon switching is not a result of predetermination in the B-state. Interestingly, in the simulations, switching of individual cells could not be predicted from the corresponding timecourses before switching. While different cells must have different switching probabilities (due to extrinsic noise), the ones which actually switch are picked randomly by fluctuations in the number of molecules. This observation is in agreement with the finding that noise in transcription from the *comK* promoter was mainly intrinsic, i.e., that the transcription probability from two copies of P_{comK} is uncorrelated (29).

Is the competence system multimodal?

We found that the maximum levels of ComK in the wild-type strain were highly variable with two distinct levels. This observation is reminiscent of a previous result reported by Maamar and Dubnau (28) showing the existence of two subpopulations with an intermediate and a high ComK-expressing class of cells. Since their setup only allowed for snapshots in time for a whole-cell population, they could not resolve whether the cells expressing *comK* at intermediate levels had already reached their maximum. Using single cell time-lapse microscopy, we showed that indeed the intermediate fraction has already reached their maximum ComK concentration.

Multimodality in the distribution of ComK of cells in the K-state hints toward further distinguishable phenotypes. There are two possible sources for multimodal expression: stochastic bursting or a deterministic mechanism due to additional regulation circuits. Transcriptional bursts mediated by slow promoter dynamics could lead to different protein levels. Then, however, typical times for binding/unbinding of transcription factors would need to be on the order of several hours, as no transitions between the two levels are evident. This is unlikely, as for prokaryotes it is known that promoter dynamics are usually very fast. Translational bursts might also lead to multimodality if mRNA dynamics are much slower than protein dynamics. In that case, one would expect to find multiple peaks in the protein distribution which are equidistant to each other. Again, the timescale of mRNA dynamics would need to be on the order of hours to explain the absence of transitions between the distribution's modes. Since these two stochastic explanations seem unlikely we suppose that a deterministic mechanism due to additional regulatory circuits is responsible for the observed multimodality. The circuit could couple into the dynamics of the *comK* promoter, thereby altering the maximal transcription rate ($\alpha_{\mathrm{M}} + \beta_{\mathrm{M}}$) by a significant amount. This assumption would be consistent with the observation that the intermediate peak is not present in the *rok*-strain, where transcription rates are elevated. A more thorough understanding of the observed multimodality has to await further experimental investigations.

A minimal model can describe the time course of switching in individual cells

Here we presented a plausible two-component model incorporating well-known processes, intrinsic and extrinsic noise. All parameters are taken directly from literature or within biological meaningful ranges, and quantitative results remarkably match those of the experimental system. Our model covers the development of competence on the single cell level which we also recorded experimentally. To include escape from competence, extended models with further interactions and more molecular species are necessary. Since the mechanisms of escape from competence are as yet sketchy and extended models tend to be less descriptive, we decided to focus on competence onset and clarify it in detail. Of special significance is the accordance in the timescale for reaching the K-state, as well as in the fraction of cells attaining competence. A discussion of the system gives straight explanations of probabilistic and deterministic switching in different *B. subtilis* strains.

For the wild-type, we identify the low-number mRNA fluctuations about the mean value 1, together with a large burst factor, as the outstanding source of fluctuations in the vegetative state. While in exponential phase our model predicts monostability (thereby forbidding switching), in stationary phase, switching initiation is facilitated due to the vicinity of B-state and threshold. Values of ComK saturate in K-state, whose position is highly sensitive to extrinsic noise, explaining broad distributions in saturation levels.

We see the minimal number of species and reaction parameters as an advantage of our model as compared to others (16,29). It grants an intuitive understanding of the key properties of competence decision: the interplay of low-number fluctuations and nonlinear dynamics.

Bistable versus excitable models

We focused on the time interval of transition from exponential to stationary phase. Similar to independent research (5,27–29,32), we find that a model featuring bistability is well fit to explain experiments performed in that temporal regime. Studies characterizing the competence system as excitable (15,16,31) focus on longer time intervals later in stationary phase. They find that an indirect inhibition of *comS* expression by ComK acts as a further feedback loop, rendering the system excitable in a region of parameter

space. These two different approaches appear to be contradictory at first sight only: The models address experiments using different strains and growth conditions. In particular, these differences led to different length scales of the switching period, to different fractions of competent cells and length of cell cycle. Another highly important difference is that Suel et al. (16) observed initiation and decay of competence over a timescale of days without actively changing the medium. The effect of quorum sensing and nutrient limitation (that is known to initiate competence development during entry into the stationary growth phase) is not explicitly involved in these studies, although change in cell concentration under microscope conditions may have an effect on the probability of switching initiation. In our experiments, we explicitly included the effect of cell density by continuously adjusting the medium thus simulating quorum sensing.

At the transition from exponential to stationary phase, the environment of a single cell, which couples into its dynamics by quorum sensing, is not in equilibrium yet. This constantly pushes the system away from steady state dynamics so that slow processes (like indirect feedback) are of less importance. Furthermore, model parameters are influenced by the overall state of the cell, so that the relative importance of network components might change over time. These variations in parameters can also carry excitable systems into regions of mono- or bistability (15,31). If indirect feedback is bypassed cells are no longer excitable (15). From an evolutionary point of view a system that initially reacts fast and in a bistable manner to stressful conditions and then transits to an excitable mode is meaningful: A fast response enhances fitness, but only cells that can exit competence and reproduce can make use of that advantage. If a first wave of transient competent cells does not succeed in producing a beneficial phenotype, there are always other individuals entering and leaving the competent state.

All models identify noise as a necessary ingredient of the competence decision network of *B. subtilis*, supporting the idea of functional noise. Noise is needed to leave the vicinity of a stable fixed and cross a threshold, which either leads to an excursion in phase space or to the approach of another stable fixed point. The origin of these fluctuations can be manifold but were analyzed by Schultz et al. (31) in a systematic way. There it was found that fluctuations in binding to the protease complex can be neglected which we accounted for by adiabatic elimination of this fast reaction. In previous theoretical studies (31,32), fluctuations in the mRNA were ignored. However, since transcript numbers per cell are extremely low, and promoter dynamics are usually assumed to be fast in prokaryotes (4) we regard the finite number effect of mRNA and the consequential translational bursting as the main source of noise (1,4). Quantitative switching times cannot be determined reliably if noise in mRNA levels is neglected (38). If promoter dynamics are slow, the corresponding fluctuations will leave their mark in the downstream mRNA, which amounts to large fluctuations in mRNA. Here we showed that extremely low mRNA numbers coupled to transcriptional bursting suffice to trigger crossing of threshold and onset of competence in a stochastic manner.

The duration of the switching period is intrinsically defined and robust toward variations of the regulation circuit

We studied the switching kinetics of individual cells, dependent on growth phase. Switching was accomplished after a period of $\rho = 1.4$ h in individual cells. One explanation could be that dependent on growth phase an external signal accumulated in the surrounding medium. However we found, that overlaying switching curves between T_0 and T_4 did not reveal large variation, indicating the expression rate and switching period did not depend on growth phase. We therefore conclude that in the K-state expression of *comK* and degradation of the protein equilibrate to a new steady state, without further extracellular regulation. Furthermore, we found that neither overproduction of the proteolysis inhibitor ComS nor knockout of the repressor for *comK* transcription had an effect on the length of the switching period, indicating that the period is a robust property of the system.

Transition from stochastic to deterministic switching at increased promoter activity

In the *rok*-strain, B-state and threshold are nonexistent in stationary phase and render the system monostable, forcing individuals into the K-state in a deterministic manner. Smits et al. (27,39) and Maamar et al. (28) showed bistable behavior in their *rok*-strain. Since we found deterministic switching with different *rok*-strains both in real-time and in bulk experiments, we attribute the difference to our more rigorous quantification of fluorescence levels to determine the threshold between cells in the B-state and cells in the K-state. Furthermore, the close proximity to the stability transition (Fig. 4 *c*) could explain that slightly different growth conditions may have a strong effect on the stability of the system.

CONCLUSION

Our combined experimental and theoretical approach showed that the length of the switching period from the B-state to the K-state of *Bacillus subtilis* is robust against variations of regulation circuit of the master regulator for competence and against the growth phase. Switching characteristics of *B. subtilis* can be understood by a well-defined two-component mathematical model that couples nonlinear dynamics to probabilistic effects. The model quantitatively retrieves the fraction of competent cells and the switching period and predicts a transition from stochastic to deterministic genetic switching at increased basal transcription rate.

Experiments verified this prediction. Furthermore, the model explains the large fluctuations between individual cells in *comK* expression in the K-state. We hypothesize that the expression of late competence proteins may be induced once a threshold concentration of ComK is reached and that therefore the cell does not require tight control over the saturation level of ComK.

SUPPORTING MATERIAL

Three tables and seven figures are available at http://www.biophysj.org/biophysj/supplemental/S0006-3495(08)00107-0.

We thank D. Dubnau for the gift of BD 2711, several plasmids, and useful discussions. We thank Hanna Engelke and Judith Megerle for support with data acquisition and processing and Kerstin Stingl and Gerda Scheidgen-Kleyboldt for valuable discussions.

This work has been supported by the Deutsche Forschungsgemeinschaft, grant No. MA3898. M.L. and J-T.K. acknowledge a stipend by the Elite-Netzwerk Bayern. Financial support of the German Excellence Initiative via the program "Nanosystems Initiative Munich (NIM)" is gratefully acknowledged. E.F. and J.O.R. acknowledge financial support by LMUinnovativ "Analysis and Modelling of Complex Systems".

REFERENCES

1. Ozbudak, E. M., M. Thattai, H. N. Lim, B. I. Shraiman, and A. van Oudenaarden. 2004. Multistability in the lactose utilization network of *Escherichia coli*. *Nature*. 427:737–740.
2. Smits, W. K., O. P. Kuipers, and J. W. Veening. 2006. Phenotypic variation in bacteria: the role of feedback regulation. *Nat. Rev. Microbiol.* 4:259–271.
3. Raser, J. M., and E. K. O'Shea. 2004. Control of stochasticity in eukaryotic gene expression. *Science*. 304:1811–1814.
4. Kaern, M., T. C. Elston, W. J. Blake, and J. J. Collins. 2005. Stochasticity in gene expression: from theories to phenotypes. *Nat. Rev. Genet.* 6:451–464.
5. Avery, S. V. 2005. Cell individuality: the bistability of competence development. *Trends Microbiol.* 13:459–462.
6. Sumner, E. R., and S. V. Avery. 2002. Phenotypic heterogeneity: differential stress resistance among individual cells of the yeast *Saccharomyces cerevisiae*. *Microbiology*. 148:345–351.
7. Thattai, M., and A. van Oudenaarden. 2004. Stochastic gene expression in fluctuating environments. *Genetics*. 167:523–530.
8. Kussell, E., and S. Leibler. 2005. Phenotypic diversity, population growth, and information in fluctuating environments. *Science*. 309:2075–2078.
9. Longo, D., and J. Hasty. 2006. Imaging gene expression: tiny signals make a big noise. *Nat. Chem. Biol.* 2:181–182.
10. Maheshri, N., and E. K. O'Shea. 2007. Living with noisy genes: how cells function reliably with inherent variability in gene expression. *Annu. Rev. Biophys. Biomol. Struct.* 36:413–434.
11. Kaufmann, B. B., Q. Yang, J. T. Mettetal, and A. van Oudenaarden. 2007. Heritable stochastic switching revealed by single-cell genealogy. *PLoS Biol.* 5:1973–1980.
12. Austin, D. W., M. S. Allen, J. M. McCollum, R. D. Dar, J. R. Wilgus, et al. 2006. Gene network shaping of inherent noise spectra. *Nature*. 439:608–611.
13. Elowitz, M. B., A. J. Levine, E. D. Siggia, and P. S. Swain. 2002. Stochastic gene expression in a single cell. *Science*. 297:1183–1186.
14. Ozbudak, E. M., M. Thattai, I. Kurtser, A. D. Grossman, and A. van Oudenaarden. 2002. Regulation of noise in the expression of a single gene. *Nat. Genet.* 31:69–73.
15. Suel, G. M., J. Garcia-Ojalvo, L. M. Liberman, and M. B. Elowitz. 2006. An excitable gene regulatory circuit induces transient cellular differentiation. *Nature*. 440:545–550.
16. Suel, G. M., R. P. Kulkarni, J. Dworkin, J. Garcia-Ojalvo, and M. B. Elowitz. 2007. Tunability and noise dependence in differentiation dynamics. *Science*. 315:1716–1719.
17. Elowitz, M. B., and S. Leibler. 2000. A synthetic oscillatory network of transcriptional regulators. *Nature*. 403:335–338.
18. Lahav, G., N. Rosenfeld, A. Sigal, N. Geva-Zatorsky, A. J. Levine, et al. 2004. Dynamics of the p53-Mdm2 feedback loop in individual cells. *Nat. Genet.* 36:147–150.
19. Mihalcescu, I., W. H. Hsing, and S. Leibler. 2004. Resilient circadian oscillator revealed in individual cyanobacteria. *Nature*. 430:81–85.
20. Amir, A., O. Kobiler, A. Rokney, A. B. Oppenheim, and J. Stavans. 2007. Noise in timing and precision of gene activities in a genetic cascade. *Mol. Sys. Biol.* 3:71.
21. Nester, E. W., and B. A. D. Stocker. 1963. Biosynthetic latency in early stages of deoxyribonucleic acid transformation in *Bacillus subtilis*. *J. Bacteriol.* 86:785–796.
22. Berka, R. M., J. Hahn, M. Albano, I. Draskovic, M. Persuh, et al. 2002. Microarray analysis of the *Bacillus subtilis* K-state: genome-wide expression changes dependent on ComK. *Mol. Microbiol.* 43:1331–1345.
23. Vansinderen, D., A. Luttinger, L. Y. Kong, D. Dubnau, G. Venema, et al. 1995. ComK encodes the competence transcription factor, the key regulatory protein for competence development in *Bacillus subtilis*. *Mol. Microbiol.* 15:455–462.
24. Hamoen, L. W., A. F. Van Werkhoven, J. J. E. Bijlsma, D. Dubnau, and G. Venema. 1998. The competence transcription factor of *Bacillus subtilis* recognizes short A/T-rich sequences arranged in a unique, flexible pattern along the DNA helix. *Genes Dev.* 12:1539–1550.
25. Vansinderen, D., and G. Venema. 1994. ComK acts as an autoregulatory control switch in the signal-transduction route to competence in *Bacillus subtilis*. *J. Bacteriol.* 176:5762–5770.
26. Turgay, K., L. W. Hamoen, G. Venema, and D. Dubnau. 1997. Biochemical characterization of a molecular switch involving the heat shock protein ClpC, which controls the activity of ComK, the competence transcription factor of *Bacillus subtilis*. *Genes Dev.* 11:119–128.
27. Smits, W. K., C. C. Eschevins, K. A. Susanna, S. Bron, O. P. Kuipers, et al. 2005. Stripping *Bacillus*: ComK auto-stimulation is responsible for the bistable response in competence development. *Mol. Microbiol.* 56:604–614.
28. Maamar, H., and D. Dubnau. 2005. Bistability in the *Bacillus subtilis* K-state (competence) system requires a positive feedback loop. *Mol. Microbiol.* 56:615–624.
29. Maamar, H., A. Raj, and D. Dubnau. 2007. Noise in gene expression determines cell fate in *Bacillus subtilis*. *Science*. 317:526–529.
30. Leisner, M., K. Stingl, J. O. Radler, and B. Maier. 2007. Basal expression rate of comK sets a "switching-window" into the K-state of *Bacillus subtilis*. *Mol. Microbiol.* 63:1806–1816.
31. Schultz, D., E. Ben Jacob, J. N. Onuchic, and P. G. Wolynes. 2007. Molecular level stochastic model for competence cycles in *Bacillus subtilis*. *Proc. Natl. Acad. Sci. USA*. 104:17582–17587.
32. Karmakar, R., and I. Bose. 2007. Positive feedback, stochasticity and genetic competence. *Phys. Biol.* 4:29–37.
33. Albano, M., J. Hahn, and D. Dubnau. 1987. Expression of competence genes in *Bacillus subtilis*. *J. Bacteriol.* 169:3110–3117.
34. Thattai, M., and A. van Oudenaarden. 2001. Intrinsic noise in gene regulatory networks. *Proc. Natl. Acad. Sci. USA*. 98:8614–8619.
35. Hamoen, L. W., A. F. Van Werkhoven, G. Venema, and D. Dubnau. 2000. The pleiotropic response regulator DegU functions as a priming protein in competence development in *Bacillus subtilis*. *Proc. Natl. Acad. Sci. USA*. 97:9246–9251.
36. Hamoen, L. W., D. Kausche, M. A. Marahiel, D. van Sinderen, G. Venema, et al. 2003. The *Bacillus subtilis* transition state regulator AbrB binds to the 35-promoter region of ComK. *FEMS Microbiol. Lett.* 218:299–304.

37. Strogatz, S. H. 2000. Nonlinear Dynamics and Chaos. Westview Press, Cambridge, MA.

38. Mehta, P., R. Mukhopadhyay, and N. S. Wingreen. 2008. Exponential sensitivity of noise-driven switching in genetic networks. *Phys. Biol.* 16:5.

39. Smits, W. K., C. Bongiorni, J. W. Veening, L. W. Hamoen, O. P. Kuipers, et al. 2007. Temporal separation of distinct differentiation pathways by a dual specificity Rap-Phr system in *Bacillus subtilis*. *Mol. Microbiol.* 65:103–120.

40. Gillespie, D. T. 1977. Exact stochastic simulation of coupled chemical reactions. *J. Phys. Chem.* 81:2340–2361.

41. Maier, B., I. Chen, D. Dubnau, and M. P. Sheetz. 2004. DNA transport into *Bacillus subtilis* requires proton motive force to generate large molecular forces. *Nat. Struct. Mol. Biol.* 11:643–649.

Biophysical Journal, Volume 96

Supporting Material

Kinetics of genetic switching into the state of bacterial competence

Madeleine Leisner, Jan-Timm Kuhr, Joachim O. Rädler, Erwin Frey, and Berenike Maier

Supplementary information for

Kinetics of genetic switching into the state of bacterial competence

Madeleine Leisner, Jan-Timm Kuhr, Joachim O. Rädler, Erwin Frey, Berenike Maier

Table of contents

Supplemental Text

Stochastic model and simulations

If molecular fluctuations are neglected, initial conditions determine the dynamics of the model and each cell develops towards one of the stable fixed points. Since numbers of M and K are low initially, a deterministic numerical simulation would end up with 0% cells in the K-state. To explain the heterogeneity of a cell culture molecular fluctuations are thus of prime importance.

To evaluate our model including stochastic fluctuations we implemented Gillespie's stochastic algorithm (14). All possible reactions and the corresponding transition rates are given in Supplemental Table 3.

To render our model more realistic, we further allowed parameter values to vary to incorporate cell-to-cell variability (size, variable number of cellular machinery, etc.), i.e. we introduced extrinsic noise. For each realization of the simulation every parameter was chosen out of a Gaussian distribution about its mean with a standard deviation of 5% of its mean, and then held constant for that run (γ_K was not varied, since its value is set by the topology of the model). The magnitude of variation in the parameters was chosen to lie in a realistic range, since no quantitative data on this subject could be obtained. The qualitative and quantitative influence of extrinsic noise can be investigated by stability and sensitivity analysis, respectively (see below). In each run the evolution of an individual cell was simulated for 6 hours.

The influence of the cell population as a whole is incorporated via S(t). Information on cell density and nutrient supply is integrated in this function by quorum sensing. For simplicity we assumed that there is a linear relation between cell density (which rises sigmoidal with time) and S(t). Thus a sigmoidal function as given by a Hill-function is appropriate. We empirically chose

$$S(t) = \frac{1500}{1 + \left(\frac{0.5h}{t}\right)^3} \qquad \textbf{(1)}$$

as shown in Supplemental Fig. 1. Realizations of our simulations are shown as ComK vs. time in Fig. 5a,b and in M-K phase space in Supplemental Fig. 2. In the former, one can see that saturation is achieved for most cells and that a band of saturation values is found, similarly to that of Fig. 2c.

We evaluated 1000 simulations and found that in 14.7% of all runs the K-state was reached, reproducing the known literature value. The K-state was defined as K > 900 at the end of the simulation period.

We also extracted the switching periods for our simulations. To this purpose we normalized the

timecourses of all cells that entered K-state to their maximal value and then shifted the time axis to overlay them at half maximal time (Fig. 5c). The temporal derivative of the mean of these curves was, in analogy with the experimental data, taken to obtain the switching rate. The switching period was defined as the time interval, where this rate was at least 0.1 of its maximal value. We found a mean switching period of 1.2h, in good accordance with experiment.

Motivation of reaction rates

Rates of the distinct chemical processes, as given in the terms of equations (1) and (2) of the main text are either well motivated from literature or generic representations of the molecular processes. Michaelis-Menten degradation of ComK is the primary mechanism which keeps protein numbers low in the vegetative state. Linear degradation, which is not catalyzed by protease complexes, is much slower (5), but establishes an equilibrium between degradation and production of ComK at high levels. Recent studies have postulated an inhibition of the *comS* promoter by high ComK concentrations on long timescales (6-8). This would lead to S becoming a function of not only t but also of K. Taylor expansion of this function S(K,t) in K then recovers to first order linear degradation.

Autocatalytic feedback in *comK* transcription is modeled by a Hill-function which is the generic function if formation of oligomers and promoter binding/unbinding are fast. It is known that four ComK proteins can bind to the *comK* promoter cooperatively (9), which gives $\gamma_K = 4$.

The small basal transcription rate is attributed to *comK* promoter leaking (10). The two linear terms for ComK production and mRNA degradation are the generic choices for these elementary processes. For parameter values of the model see Supplemental Table 2.

Rise of ComS facilitates activation of auto-feedback in two distinct manners.

Degradation of ComK follows Michaelis-Menten kinetics, catalyzed by the MecA/ClpC/ClpP protease complex (D). The reaction scheme is

$$K_f + D_f \underset{\lambda_{-1}}{\overset{\lambda_1}{\leftrightarrow}} [KD] \overset{\lambda_2}{\rightarrow} D_f \qquad \textbf{(2)}$$

where K_f, D_f, and [KD] refer to free ComK, free protease complex, and ComK bound to the protease complex, respectively. Reactions λ_1 and λ_{-1} are fast in comparison with the actual degradation λ_2.

Cells respond to high density by increasing ComS concentration which couples into the competence decision network ((9), (Fig. 1)). ComS is a small peptide that is degraded by the MecA/ClpC/ClpP protease complex (11). Protease complexes are thus partially occupied by ComS and degradation of ComK is slowed down. Furthermore K_f is increased in the presence of S and thus more K is available

for dimer formation and binding to the *comK* promoter, which has a prominent effect near the lower fixed point and threshold (Supplemental Fig. 3). Since degradation of ComS and ComK happens through the same processes (9) we assume identical rates,

$$S_f + D_f \underset{\lambda_{-1}}{\overset{\lambda_1}{\leftrightarrow}} [SD] \overset{\lambda_2}{\rightarrow} D_f \qquad \textbf{(3)}$$

Neither absolute affinities of ComK nor ComS towards the protease complex nor total ComS numbers are known. But since we are interested in the dynamics of K, only the number of protease complexes currently occupied by S is of importance. In this sense S is understood as an effective ComS number which gives, when assuming identical affinities towards the protease complex as for ComK, the right number of occupied MecA/ClpC/ClpP binding sites.

By the law of mass action one finds

$$K_f(K,S) = \frac{1}{2}\frac{K}{K+S}\left(K+S-D-\frac{\lambda_1}{\lambda_{-1}} + \sqrt{\left(K+S-D-\frac{\lambda_1}{\lambda_{-1}}\right)^2 + 4\frac{\lambda_1}{\lambda_{-1}}(K+S)}\right) \qquad \textbf{(4)}$$

which appears in eqn. (2) of the main text. Here particle conservation, i.e. $K = K_f + [KD]$, $S = S_f + [SD]$, and $D = D_f + [KD] + [SD]$, was used.

S as a function of time was set empirically as an external control parameter (see above).

Parameters of the model

Our model includes ten parameters (Supplemental Table 2) and the external control functions S(t) and $\alpha_M(t)$ (Supplemental Figure 1). All parameters were chosen consistently with results of earlier studies. δ_K, D and β_K/δ_M, could be obtained directly from literature. Further constraints were $M \approx 1$ in the vegetative state in stationary phase (12), a switching threshold of a few hundred ComK proteins and a saturation value of 10^4-10^5 proteins in the K-state (5).

In earlier studies (3) we found that basal expression increases about an hour before entrance to stationary phase and that after a 'switching-window' the percentage of competent cells stay constant. We incorporated that into the model by reducing α_M by a factor of r outside of that window in all simulations. $r = 0.5$ was chosen small enough so that switching was very rare for the reduced basal transcription rate $r \cdot \alpha_M$ (Supplemental Figure 1).

The rates of the degradation reaction λ_1, λ_{-1}, and λ_2 have not been addressed experimentally. However, this is not a drawback to the model, since λ_2 is incorporated in δ_K (Michaelis-Menten theory). λ_1 and λ_{-1} appeared explicitly in $K_f(K,S)$ and since both rates are fast, only their ratio is of interest. Varying λ_1/λ_{-1} has only minor influence on the model dynamics.

Sensitivity and Stability

The broad distribution of saturation levels was attributed to variability in the parameters of the model, known as extrinsic noise (13). To quantify its influence on the saturation value of ComK one can define the sensitivity $\Sigma(\mu)$ with respect to a given parameter μ as the relative change $\Delta K/K^*$ of the upper fixed point $K^*(\mu)$ due to a relative change $\Delta\mu/\mu$ in that parameter (Supplemental Table 2). An analytical expression is given by

$$\Sigma(\mu) = \frac{\Delta K / K^*}{\Delta \mu / \mu_0} = \frac{\mu_0}{K^*} \frac{\partial K^*}{\partial \mu}\bigg|_{\mu=\mu_0} \qquad \textbf{(5)}$$

as follows by Taylor expansion up to linear order of the fixed point's position.

If parameters of the model are varied strongly the qualitative features of stationary states can change (bifurcation). This may for example be the case for non-wild type strains, for which the competence circuit (Fig. 1) is altered. Numerically we can compute the number and kind of fixed points of our model and thereby analyze its possible dynamics. An example is given in Fig. 4b, which also explains deterministic switching for the *rok*- strain. Note that the wild type parameters have to be set close to the bifurcation in which the lower stable and the instable fixed point annihilate to admit stochastically induced crossing of threshold.

Supplementary Tables

Strain	Genotype	Source
BD 630	*his leu met*	-
BD 2955	*his leu met rok- (spc[a]),*	(1)
BD 2711	*his leu met comK-gfp (CBL[b], cat[a])*	(2)
BM 101	*his leu met rok- (spc[a]), comK-gfp (CBL[b], cat[a])*	this study
BM 77	*his leu met comK-gfp (CBL[b], cat[a]), multicopy comS (kan[a])*	(3)
BM 2528	*his leu met, multicopy comS (kan[a])*	(4)

Supplemental Tab. 1. *B. subtilis* strains used in this study.
[a] kan, cat and spc stand for resistance to kanamycin, chloramphenicol and spectinomycin respectively.
[b] Inserted by Campell like integration.

Parameter	Significance	Mean value	Motivation/ Explanation	Sensitivity of upper fixed point Σ
β_K	Translation rate per mRNA	1 1/s	burst factor: $\beta_K/\delta_M \approx 50$; (12)	3.31
δ_K	Maximal Michaelis-Menten degradation rate of K	11.3 1/s	≈ 11.5 1/s; (5) Fig. 1A therein	-2.98
δ_0	Linear degradation rate of K	$1.5 \cdot 10^{-4}$ 1/s	Ration of saturation of *rok*- and wild type	-0.34
q_K	(K+S) at half-maximal enzymatic degradation	400	<9000; (5) Fig. 1A therein; sets position of fixed points	0.14
α_M	Minimal transcription rate	0.025 1/s	M ≈ 1 before switching; (12)	0.39
β_M	Maximal additional transcription rate by feedback	0.19 1/s	position of upper fixed points/ saturation	2.93
p_K	Half-maximal feedback	600	position of switching threshold	negligible
γ_K	Cooperativity/Hill Coefficient	4	four K proteins bind to the promoter cooperatively (9)	negligible
δ_M	Degradation rate per mRNA	0.022 1/s	burst factor: $\beta_K/\delta_M \approx 50$ and M ≈ 1 before switching; (12)	-3.31
D	Number of protease complexes	700	(5)	negligible
λ_1/λ_{-1}	Equilibrium constant of degradation reaction	1	variation of this parameter are of minor influence only	negligible

Supplemental Tab. 2. Parameters of the model. Besides the constraints given in the column motivation/explanation, parameters were chosen to be in line with our experimental findings.

Description	Reaction	Transition rate
protein production	$K \rightarrow K+1$	$\beta_K M$
protein degradation	$K \rightarrow K-1$	$\frac{\delta_K K}{q_K + S + K} - \delta_0 K$
mRNA production	$M \rightarrow M+1$	$\alpha_M + \frac{\beta_M}{1+(p_K / K_f)^{\gamma_K}}$
mRNA degradation	$M \rightarrow M-1$	$\delta_M M$

Supplemental Tab. 3. Reactions of the stochastic model. Processes and their corresponding transition rates are given. Note that the rates depend on the number of K and M. These transition rates are used in the stochastic simulations.

Supplementary Figure Captions

Supplemental Fig. 1. Empirically set development of the number of ComS molecules S(t) used for the simulations and of the basal transcription rate $\alpha_M(t)$. The latter gives rise to a window of opportunity.

Supplemental Fig. 2. Fluctuations in vegetative state and crossing of switching threshold. The lower left part of Figures 4a,b of the main text are shown. For a typical realization (black line) switching occurs after a long period of fluctuations about the lower fixed point. Note that only discrete molecule numbers are taken, which can be seen very well for the number of mRNAs. Nullclines (Blue: dM/dt = 0; red: dK/dt = 0) are plotted for S = 0 (dotted lines) and S = 1500 (straight lines).

Supplemental Fig. 3. Influence of [ComS] on the fraction of free ComK. ComS binds competitively to the MecA/ClpC/ClpP protease complex, thereby increasing the amount of free ComK (K_F), see eqn. (4) of Supplemental text. Dashed line: S = 0, straight line: S = 1500. Near the lower stable and unstable fixed points (K ≈ 200 and K ≈ 600, respectively) this has an important effect on free ComK concentration.

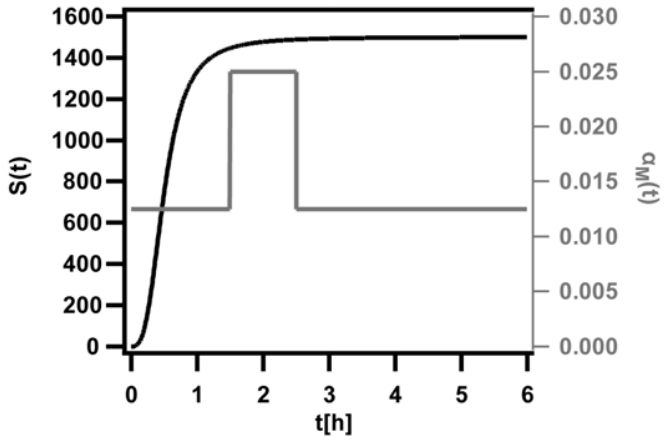
Supplemental Fig. 4. Comparison of switching kinetics between bulk and single cell measurements to confirm equal development under microscopic conditions and in liquid culture. Fraction of cells in the K-state. Black: cells grown in Erlenmeyer flask, grey: cells grown on a microscope slide. a) wt, b) *rok*-. Black and grey lines: best fit to a sigmoidal function.

Supplemental Fig. 5. Switching kinetics of individual *rok*- cells (BM101). a) Time course of fluorescence intensity of cells switching into the K-state. b) The fluorescence of 50 individual cells was normalized to the cumulative expression (maximum fluorescence intensity). The time axis was shifted by $\tau_{1/2}$, where cells had half maximum fluorescence intensity.

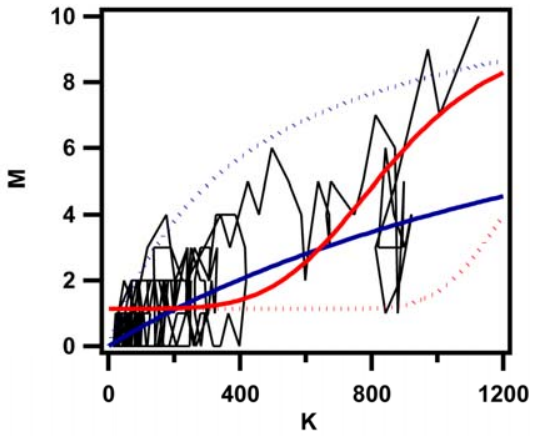
Supplemental Fig. 6. Switching kinetics of individual ComS overproducing cells (BM77). a) Time course of fluorescence intensity of cells switching into the K-state. b) The fluorescence of 50 individual cells was normalized to the cumulative expression (maximum fluorescence intensity). The time axis was shifted by $\tau_{1/2}$, where cells had half maximum fluorescence intensity.

Supplemental Fig. 7. Switching period is independent of growth phase. a) Histogram of switching

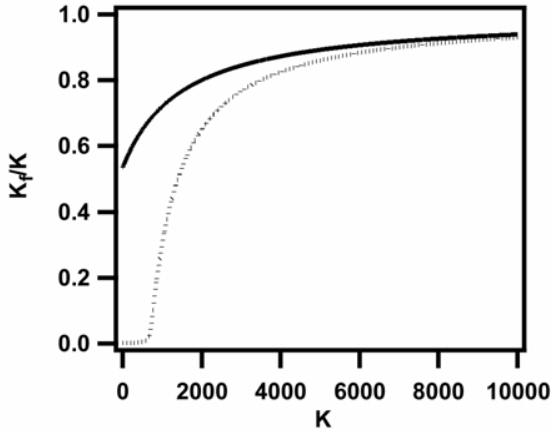
period ρ. Red line: best fit to a Gaussian function with $\langle\rho\rangle = 1.44 \pm 0.02$ and a width of 0.33 ± 0.02. b) Switching period ρ as a function of growth phase *T*.



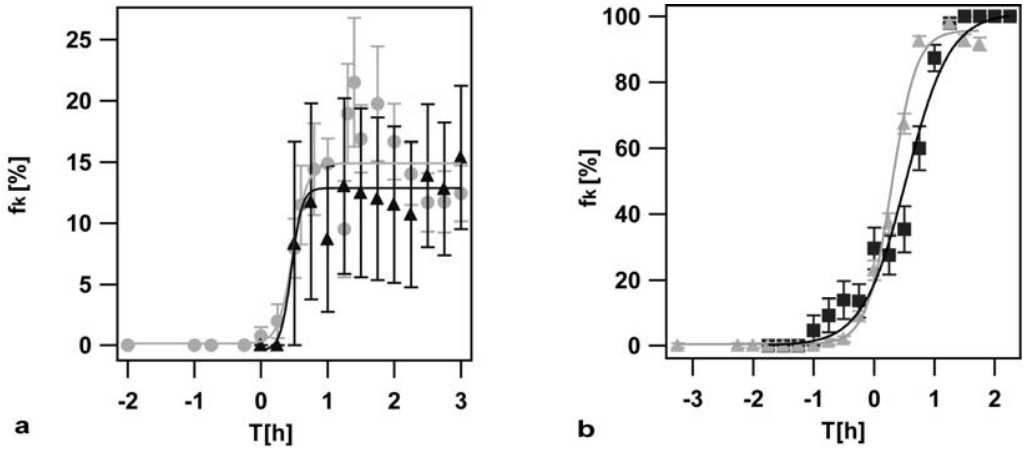
Supplemental Fig. 1



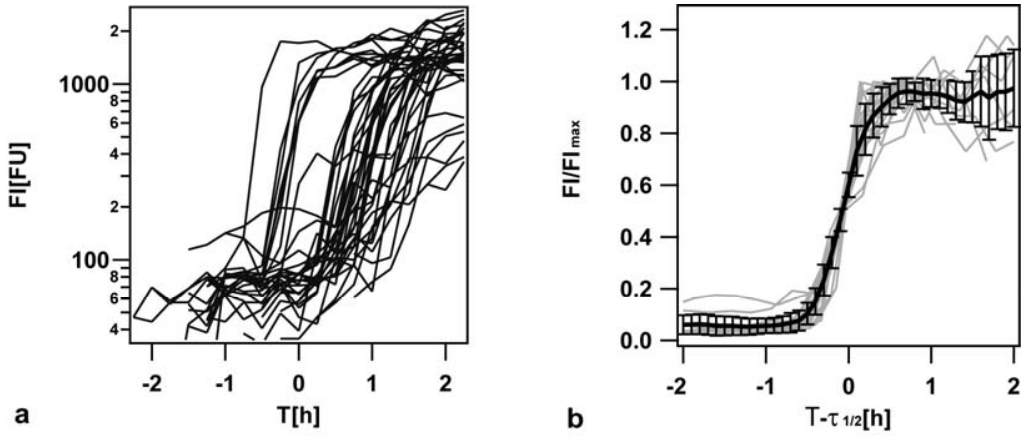
Supplemental Fig. 2



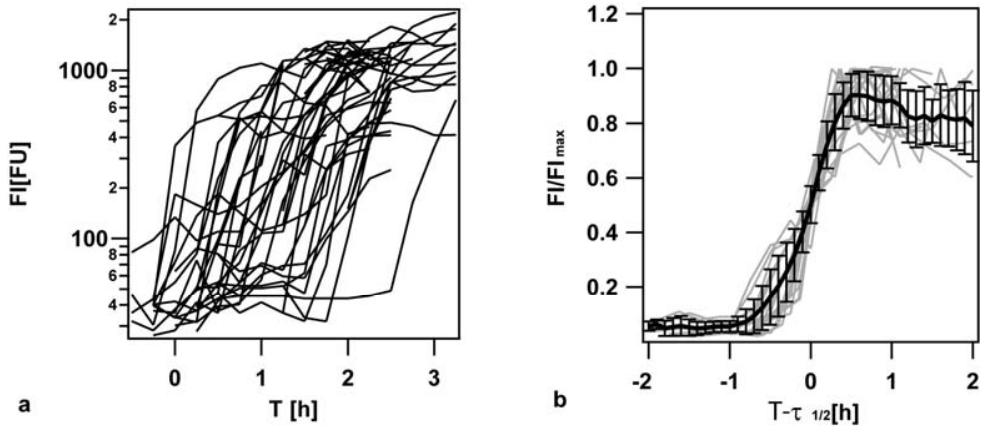
Supplemental Fig. 3



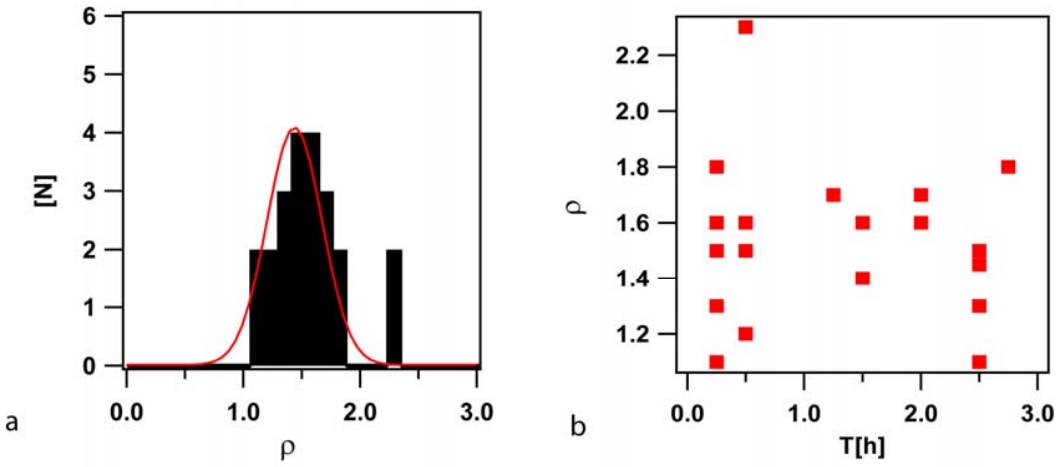
Supplemental Fig.4



Supplemental Fig. 5



Supplemental Fig. 6



Supplemental Fig. 7

Supplemental References

1. Hoa, T. T., P. Tortosa, M. Albano, and D. Dubnau. 2002. Rok (YkuW) regulates genetic competence in Bacillus subtilis by directly repressing comK. Molecular Microbiology 43:15-26.
2. Haijema, B. J., J. Hahn, J. Haynes, and D. Dubnau. 2001. A ComGA-dependent checkpoint limits growth during the escape from competence. Molecular Microbiology 40:52-64.
3. Leisner, M., K. Stingl, J. O. Radler, and B. Maier. 2007. Basal expression rate of comK sets a 'switching-window' into the K-state of Bacillus subtilis. Molecular Microbiology 63:1806-1816.
4. Liu, L., M. M. Nakano, O. H. Lee, and P. Zuber. 1996. Plasmid-amplified comS enhances genetic competence and suppresses sinR in Bacillus subtilis. Journal of Bacteriology 178:5144-5152.
5. Turgay, K., J. Hahn, J. Burghoorn, and D. Dubnau. 1998. Competence in Bacillus subtilis is controlled by regulated proteolysis of a transcription factor. Embo Journal 17:6730-6738.
6. Suel, G. M., J. Garcia-Ojalvo, L. M. Liberman, and M. B. Elowitz. 2006. An excitable gene regulatory circuit induces transient cellular differentiation. Nature 440:545-550.
7. Suel, G. M., R. P. Kulkarni, J. Dworkin, J. Garcia-Ojalvo, and M. B. Elowitz. 2007. Tunability and noise dependence in differentiation dynamics. Science 315:1716-1719.
8. Smits, W. K., and V. Bongiorni C., J.W., Hamoen, L.W., Kuipers, O.P., and M. Perego. 2007. Temporal separation of distinct differentiation pathways by a dual specificity Rap-Phr system in Bacillus subtilis. Molecular Microbiology 65:103-120.
9. Hamoen, L. W., G. Venema, and O. P. Kuipers. 2003. Controlling competence in Bacillus subtilis: shared use of regulators. Microbiology-Sgm 149:9-17.
10. Avery, S. V. 2005. Cell individuality: the bistability of competence development. Trends in Microbiology 13:459-462.
11. Prepiak, P., and D. Dubnau. 2007. A peptide signal for adapter protein-mediated degradation by the AAA(+) protease ClpCP. Molecular Cell 26:639-647.
12. Maamar, H., A. Raj, and D. Dubnau. 2007. Noise in gene expression determines cell fate in Bacillus subtilis. Science 317:526-529.
13. Elowitz, M. B., A. J. Levine, E. D. Siggia, and P. S. Swain. 2002. Stochastic gene expression in a single cell. Science 297:1183-1186.
14. Gillespie, D. T. 1977. Exact Stochastic Simulation of Coupled Chemical-Reactions. Journal of Physical Chemistry 81:2340-2361.

Part II

Growth of Microbial Colonies as a Stochastic Non-Equilibrium Process

5 Range Expansion and Microbial Colony Growth

Often in biology, species are restricted to some kind of territory. This might be an island for example or a region with given climatic or other environmental conditions, necessary for the species' prosperity. However, the spatial restriction may be overcome if the species enters an, as yet unoccupied, territory (like another island), if environmental conditions change (say, as a result of climate change), or if the species acquires a new property (e.g. through mutation) which lets it withstand a wider range of environmental conditions. In this case, the species under consideration can extend its territory, i.e., undergo a *range expansion*. Of course, this is a complicated process, which depends on the properties of the species as well as on that of the environment. We will therefore have a short look at some macroscopic examples of range expansion, before we consider more controllable microbial experiments, which aim to take into account the central aspects of this fascinating process.

5.1 Macroscopic Range Expansions

5.1.1 Homo sapiens

A common theory for the early dispersal of humans purports that modern humans, originating from Africa, where our species evolved to its present form, have spread across all continents of the Earth [123, 124, 125]. About 120000 to 60000 years ago members of *Homo sapiens* crossed the Red Sea and started to settle on the Arabic Peninsula and in the Near East. From there waves of expansions spread to (i) India and South East and South Asia and further to Australia roughly 25–70 thousand years ago[1] (ii) Europe about 40 thousand years ago. 10–20 thousand years ago, at the end of the Last Glacial Maximum the rest of Eurasia was settled and humans crossed the still existing, ice-free Bering land bridge to the Americas. From there the whole American continent was successively populated. This is an example of a *range expansions* in our own species' history.

5.1.2 Invasive Species

Scenarios like the human colonization of the continents must also have happened for other species over and over again and on different scales since life first appeared. These range

[1] A recent study finds evidence that the ancestors of the Aboriginals have indeed left from Africa and entered Australia in an independent early migration wave [126].

expansions still happen today and actually are very frequent as a result of human travel and globalization effects. *Introduced species*, which have been brought to a new habitat by humans (by accident or intentionally) often have an adverse impact on the ecosystem in the newly occupied habitat and may threaten biodiversity, since natural enemies (predators or herbivores) are rare, prey is unused to the new species, and native, *endemic* species may be outcompeted in some way. In this case one speaks of an *invasive species* [127]. Humans intentionally brought a huge number of non-native species to new habitats, for various reasons, including agriculture, forestry, livestock breeding, hunting, or even for aesthetic and nostalgic reasons or as pets. Prominent examples of purposely introduced and now invasive species include [128]:

- The Eastern grey squirrel was first brought from America to the parks in the United Kingdom in the late 19^{th} century. They spread very fast and have by now far outnumbered the native red squirrel.
- Cane toads are originally form South and Central America, and were introduced to control pests of local crops on various islands. While they did do much harm there, they were in the 1930s introduced to Australia, to reduce a native beetle. This did not work particularly well, but instead the toads reproduced wildly, outcompeting local competitors, and simultaneously colonizing the continent from the northeast.
- Sailors and colonists brought European rabbits to many places as a cheap, fast reproducing source of food and fur. In Australia this had devastating consequences as many plant species served as food sources to the rabbits and were decimated to the point of extinction. This brought along erosion and other ecological problems.
- About one hundred European Starlings where set free in New York City in 1890, as a first step in the attempt to bring all birds mentioned in Shakespeare's works to North America. The starling expanded rapidly and reached California by 1955, and heavily competes for nest sites with endemic species. The European Starling was also introduced to (between others) Australia to control pests, where it had mayor impact on the environment and promotes spreading of introduced plant's seeds.

While one could continue this list to include many pet animals and more plants, other major invasion are by species which have by chance travelled along with humans, e.g., the zebra mussel in the ballast water of ships from Europe to North America. Like in Europe the zebra mussel colonizes lakes, rivers and channels [129].

There are also examples where the introduction of new species led to a benefit, at least to humans [130]. Further, humans try, with some success, to reintroduce extirpated, i.e. locally extinct, species (e.g. wolves) to their former territory. Sadly, these are rare cases compared to introductions with a detrimental impact.

5.1.3 Climate Change

One essential part of the environment a species or population senses are the climatic conditions. Species often react to climatic changes by migration, where the best known example is seasonal migration, e.g., by birds. On much longer, global timescales, the climate changes

with ice ages and warm periods, and this has undoubtedly led to range shifts for all kinds of animals and plants.

However, as an effect of anthropogenic global warming, the marked effect of climate change on the migration of many species can be observed one shorter timescales. While the human impact on world climate might have started millennia ago [131], and might have influenced animal migration since then, present data [132] is virtually indisputable: Species move to colder regions, be it towards the poles or to higher altitudes in mountainous regions [133, 134].

5.1.4 The Founder Effect

A very effective way to follow range expansions, as those introduced above, is by genetic analysis. By taking genetic samples from a population at different times and places, one can deduce the degree of relationship between individuals.[2] In particular, from the number of different genes, it is possible to assess the time when the *last common ancestor* of a specific gene lived. A descriptive example is the historic human range expansion discussed in Section 5.1.1: One finds that genetic diversity in the native human populations decreases with the distance from Africa (e.g. along the coastlines of the Indian and Pacific Oceans and going from north to south in the Americas) [135, 136].

This is the result of sampling from a small population, also known as *genetic drift* [137], which happens during range expansions: Usually only a small group of individuals is at the very front of the population, thus constituting a *population bottleneck*. These individuals typically are the ancestors of following generations contributing to the expansion. Sampling from them proceeds in the fashion of a Fisher-Wright process [137, 138], which in this spatial setting is known as the *founder effect* [139]. As the number or individuals is small, fluctuations are relevant during sampling, and genetic diversity can be lost by chance. Some of the descendants, which only carry part of the genetic diversity, will constitute the next group that continues the range expansion, which in turn leads to what is called the *serial founder effect* [135, 136].

5.2 Growth of Microbial Colonies in the Microlab

In macroscopic range expansions many different aspects influence the dynamics: geography, vegetation, climatic conditions, competition and cooperation between individuals, interactions between species, and mobility. Moreover typical timescales are decades to millennia, and typical distances can reach up to several hundreds of kilometers. To find generic properties, it is therefore necessary to reduce the complexity and scale, by finding small and (comparatively) simple model systems. During the last years growing microbial colonies have been used to mimic range expansions, though similar systems have been studied in their own right for decades [140, 141, 142, 143, 144].

[2] Complications which arise due to the possibility of horizontal gene transfer are briefly discussed in Chapter 3.

5.2.1 Single Species Colonies

Most commonly, microbial colony growth is studied on an agar substrate in a Petri dish. The morphology of the created colonies is subject to many different influences, with mobility being one of prime importance. For mobile microorganisms the resulting patterns are often of stunning beauty and show a wide range of complicated patterns [140, 141, 142, 143, 144] which, however, do not resemble that of typical macroscopic range expansions. We will therefore focus on the growth of non-motile microbial cells, which mainly spread through reproduction and a resulting pushing force at the front.

Colony Morphology for Non-motile Microorganisms

In this case, depending on the species, the softness of the substrate (controlled by agar concentration), the concentration of nutrients, and other properties, the morphology of a colony still varies substantially. Research is usually confined to a few model organisms like *E. coli* or *B. subtilis*. Even under constant conditions, colony growth is an organized non-trivial $3d$ process, with spatiotemporal change in phenotype composition [145] and cell density [146], mediated through self-organization and hierarchical patterning [144]. However, there are two main and distinct morphologies depending on nutrient and agar concentration: *compact* and *branched* [147, 148].

The branched morphology is encountered if nutrients are limited, since then only those parts of the colony keep growing which are both advancing into new territory and are not too close to other parts of the colony [149, 142, 143, 147]. The most successful model for these conditions is the *diffusion-limited-aggregation* model [150, 151], which explicitly takes the scarce nutrients into account.

Rough Fronts of Compact Colonies

While the branched morphology displays enormous complexity [142, 143], compact colonies, which evolve in nutrient-rich settings resemble coarse-grained range expansions much closer. In this setting, colonies seem to expand in a much smoother manner as compared to branched morphologies, at least at a macroscopic level. While there is growth in the vertical direction it is usually limited to $\sim$10–100 cell layers [140], and the colony mostly grows horizontally at the interface between occupied and unoccupied territory, known as the *growth front*. Microscopically, growth cannot be completely smooth, since on the length scale of the individual there must exist protrusions and indentations, as a result of the inherent stochasticity in reproduction. Growth is perpendicular to the local orientation of the expanding front and therefore protrusions also grow laterally with time. This leads to the creation of ever longer ranged fluctuations in the position of the colony's front, an effect known as *kinetic surface roughening* [152].

Theoretical aspects and modeling of surface roughness will be discussed in Section 6.3, where we will also have a look at the Eden model [153, 152] a surface growth model for compact bacterial colonies. Experimental studies involving bacteria [154] found unusually rough fronts,

while recent experiments with eukaryotes [155] displayed characteristic Eden roughening. Though roughening in itself is a rather theoretical field of research, it can heavily influence the outcome of microbial range expansions where there is competition between different strains, since there is a strong coupling between population dynamics at the front and the front's roughness properties.

5.2.2 Heterogeneous Populations

One of the central questions with respect to range expansions is the influence of the spatial settings and the founder effect on the composition of a population composed of different strains, which cannot be captured by continuous well-mixed approaches, see Section 5.3. In this context the relative importance of selection between different phenotypes and genetic drift at the expanding front is central: Are fluctuations strong enough that a selective advantage no longer ensures survival? For well-mixed systems it is known that for small population sizes selective advantage is of less importance as compared to genetic drift [156, 157].

Single-Species Sectors

In spatial microbial experiments, phenotypes which have much larger growth rates than their competitors form lateral expanding sectors [158, 140, 142, 159, 160], since growth of protrusion also proceeds perpendicular to the primary direction of the range expansion. These sectors take over linear fronts deterministically, while for circular colonies slower growing phenotypes may prevail. Interestingly, even in the neutral case sectors establish and coarsen at the expanding front, as phenotypes locally fixate at the front, which is the result of the founder effect [161, 162, 160]. An important finding of these works is that boundaries between distinct sectors fluctuate in a super-diffusive manner, which results from the front's roughness, which is on all length and times scales imprinted in the meandering of the sector boundaries. This can be explained by the scaling between lateral and longitudinal extensions of surface undulations and the typical timescale on which such undulations decay [163, 161]. The general idea is sketched in a descriptive way in Figure 5.1.

Mutations and Selection during Colony Growth

On large scales selection will always dominate over fluctuations, but what if mutations are present, which constantly create sectors of mutants on the microscopic length scale? If mutations and selection alter the composition of the front in the same direction (i.e. towards a higher portion of a given strain) it is clear that the result is an accelerated dominance of the faster growing strain. This can, for example, be seen from Figure 6.6 which shows experiments which have been carried out, by the author of this work, in the laboratory of the soft condensed matter group at the chair of Prof. J. O. Rädler at the Ludwig-Maximilians-Universität in Munich.

More interesting than the case where mutations and selection act in concert, is the situation where they are antagonists. For the well-mixed case, Eigen argued that there must be an

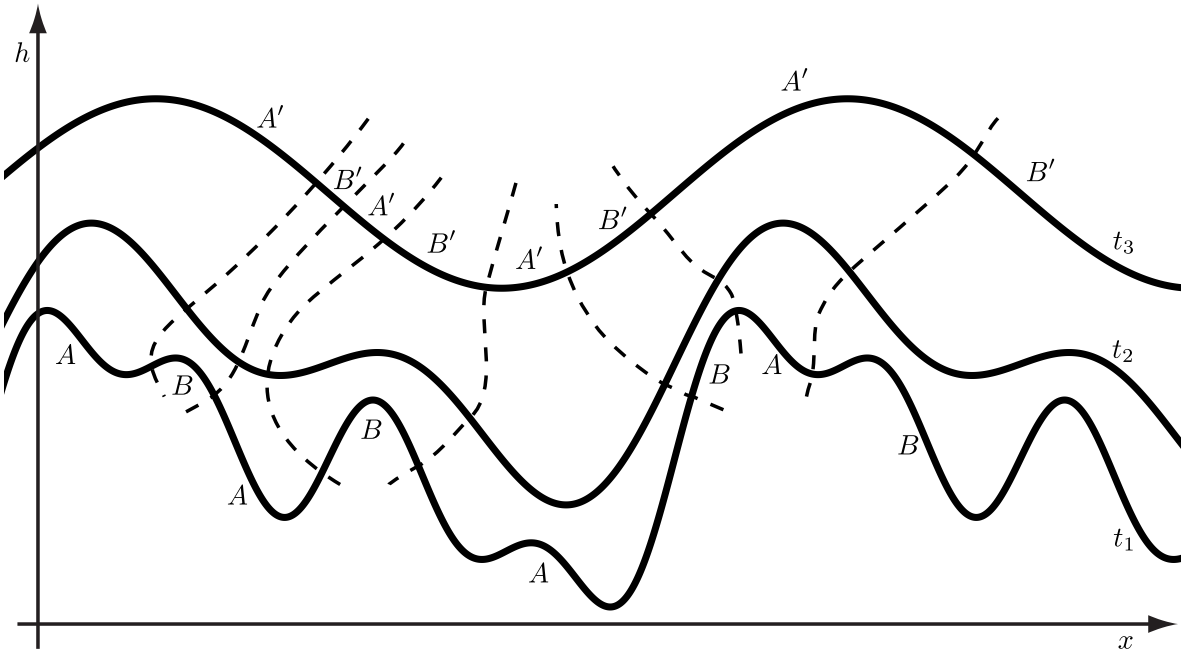


Figure 5.1: Influence of surface roughness on the sector boundaries between domains of different bacterial phenotypes during colony growth. Let the lowest bold line be the front of a growing bacterial colony at time t_1, which is composed of alternating sectors of neutral (same growth rate) strains A and B. The other two bold black lines show the most probable front position (or the average front position) at later times t_2 and t_3, where short length fluctuations of the front have decayed. The dashed black lines are the most probable (average) trajectories of the boundaries between sectors. As one can see the shape of sectors is strongly influenced by the roughness of the front, since sector boundaries always move perpendicular to the local orientation of the front.

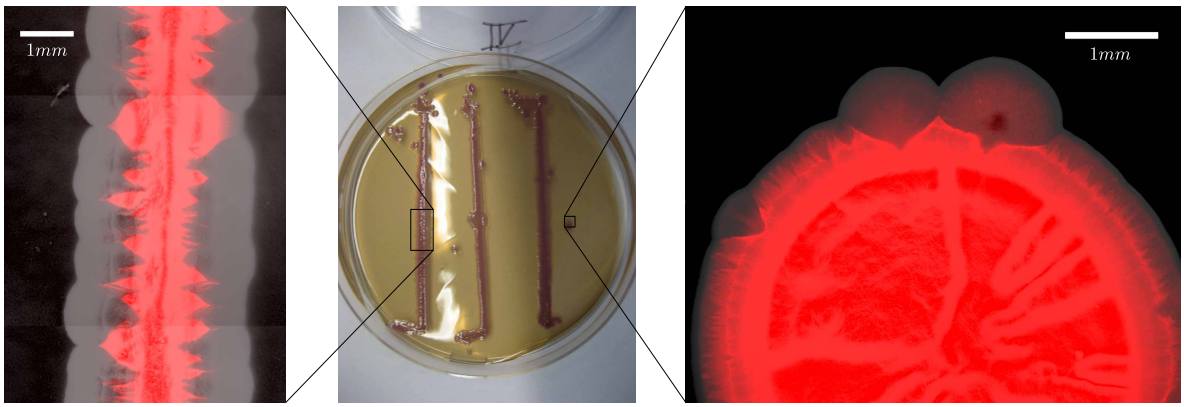


Figure 5.2: Experiments on microbial colony growth under the influence of mutation and selection. **Center:** Fluorescently labeled *E. coli* bacteria have been inoculated in a strip (left) or circular (right) geometry and then let grow for 48 hours. **Left and right:** Overlays of fluorescent and bright field microscopic images reveal details of colony growth. Mutation-like events result in loss of fluorescence, enhanced growth rate and formation of laterally expanding sectors of non-fluorescent bacteria.

error threshold, a critical mutation rate, where selection is not sufficiently strong any more to withstand the large number of newly arising mutations [164, 160]. If this is the case, then the mutants will overtake the population very fast and the wild-type will be wiped out even though it has a higher reproduction rate. This collective phenomenon of many independently arising mutations acting in concert can also be found in spatial settings. One example where this happens at an expanding rough front is the model introduced in the manuscript at the end of the next chapter. There, the error threshold, which from a physicist's point of view is a phase transition, is examined in detail and it is found that the front's roughness has a severe impact on the resulting patterns of mutant and wild-type sectors. To understand the approach of said manuscript, we introduce the key concepts and properties of phase transitions to absorbing states and of dynamic surface roughening in Chapter 6.

5.3 Modeling Approaches Different from Surface Growth

With respect to both macroscopic and microscopic range expansions a number of interesting questions may be asked (and hopefully answered). What is the speed of the population front, and how does it depend on birth and migration rates and other properties of the individuals from which the population is composed. How important are fluctuations and how do they influence the evolution of the population? How is the diversity of a population shaped under the influence of genetic drift, selection and mutations and how does it change over time? What patterns do evolve in the case of a non-homogeneous population? To answer these questions one needs models that describe the spatial growth of the population and make the measurable quantities accessible and comparable.

Besides surface growth models (see Chapter 6) there exist other approaches to model range expansion and bacterial colony growth, which we will now shortly discuss.

5.3.1 The Fisher-Kolmogorov Equation

In the simplest case, where local populations are large and one considers a coarse enough scale, the change in population density $\rho(x,t)$ during range expansion is often modeled by the *Fisher-Kolmogorov equation*,

$$\dot{\rho}(x,t) = r\rho(1-\rho/K) + D\nabla^2\rho\,. \tag{5.1}$$

This equation describes locally *logistic growth* with a growth rate at low concentrations r and a *carrying capacity* K through the first term, and spatial diffusion with diffusion constant D, through the second term, which leads to spatial expansion [74]. The Fisher-Kolmogorov equation is a non-linear reaction-diffusion equation which was set up to model the spread of a gene through a population [165]. One can show that in $1d$ it has traveling waves, $\rho(x,t) = f(x \pm ct)$, as asymptotic solutions, with wave speed $c \sim \sqrt{rD}$, when initialized with a concentration which is not zero everywhere [166]. Due to this intuitive $1d$ result, equation (5.1) has mainly been used to model the spread of species along rivers, coastlines, boundaries of distinct ecosystems, and other linear habitats in the context of range expansions [167]. In higher dimensions one has the additional problem of curvature of the front, which prohibits the

derivation of many analytic results, while comparison with data is still possible by numerical solution. Further, there exist a variety of extensions of the Fisher-Kolmogorov equation, which widen its range of applications by taking into account more details of the process under consideration [74, 167].

The Fisher-Kolmogorov equation implies that the population is well described by a continuous function and fluctuations can be completely neglected. This implies that neither the founder effect nor surface roughness nor possible consequences of the two are accessible by analyzing equation (5.1). To make things worse, equation (5.1) predicts regions of arbitrarily low density in the wake of the population wave, and at some point the discrete nature of individuals must be observable. These leading individuals are of prime importance as they are the ones which "pull" the population wave and determine the future of the expanding population, since they are the very few founders which have occupied new territory. To incorporate fluctuations one can add a noise term proportional to $\sqrt{\rho(1-\rho)}\eta(x,t)$, where η is Gaussian white noise, to equation (5.1). Calculations show strong deviations [168, 169], especially with respect to the speed of wave propagation, from the non-stochastic, continuous solutions. The results of stochastic calculations agree with simulations of the *stepping stone model* [170], which is a discrete (in space, time and density) and stochastic model, which in the continuum limit is analogous to the Fisher-Kolmogorov equation.

5.3.2 Simulations of Microbial Colony Growth

While analytical approaches are important to understand mechanisms, when it comes to competition between several species or mutations in the course of range expansions (see e.g. [162, 171]), simulations, often generalizations of the stepping stone model [170], are an essential tool.

One of the most important result is that due to genetic drift, and the resulting founder effect, selectively disadvantaged competitors may stay much longer on the front than expected from well-mixed scenarios. If the reduced fitness is the result of a mutation, then the descendants of the original mutant (i.e. the mutant strain) can "surf" the expanding population wave [172, 173, 174, 175]. Simulations like this can, of course, be conducted for any geometry, and have even been applied to the European landmass to model human range expansion [176].

Simulations have also been employed to investigate in the growth and organization of biofilms [177, 178, 179], where many different species interact to build up growing aggregates which exhibit complex morphologies and intricate interspecies interactions. In this context especially the problem of cooperation has attracted much attention [180, 181, 182]. A common finding for biofilm evolution is that local interactions lead to the emergence of global changes in composition and structure.

6 Absorbing State Phase Transitions and Surface Growth Roughening

Statistical mechanics aims to explain the macroscopic properties of a system from the specifications and interactions of its microscopic constituents. It is especially successful in doing so for systems in equilibrium. Due to this success, many non-equilibrium approaches follow the track laid out for equilibrium problems as far as possible and it turns out that there are many similarities and analogies to be found. For this reason, we here shortly discuss the basic phenomenology for second order equilibrium phase transitions, using the Ising model as an example. After addressing critical exponents, the idea of universality classes, and scaling (and also finite size scaling) we try to generalize to one particular class of non-equilibrium systems, which undergo transitions to absorbing states. As an example we introduce directed percolation (DP), as it is very descriptive and makes it easy to see both similarities and differences to equilibrium models. A more extended discussion of the equilibrium theory is found in any textbook on statistical mechanics, e.g., [183, 184, 185], while articulate references in the context of non-equilibrium phase transitions are, e.g., [186, 187].

We then turn to another class of non-equilibrium processes, known as surface growth models, which do not undergo phase transitions, but nonetheless exhibit critical behavior, self-affine properties and scaling and can therefore be sorted into universality classes. An excellent introduction to surface growth is the book by Barábasi and Stanley [152].

6.1 Equilibrium Phase Transitions

6.1.1 The Framework of Statistical Mechanics

For systems in equilibrium there exists a formalism to derive macroscopic variables of interest (e.g. response functions) which is, at least in principle, straightforward. The formalism rests on the assertion that in equilibrium one can assign each possible micro state s of the system under consideration a probability $p(s)$ to be the one which is presently realized. In the case of the canonical ensemble it is the Boltzmann factor $\exp(-\beta\mathcal{H}(s))$, which assigns each state its relative statistical weight according to the ration of its internal energy, given by the Hamiltonian $\mathcal{H}(s)$, to the typical thermal energy $\beta^{-1} = k_B T$ of the system, where k_B is the Boltzmann constant and T is the temperature. The sum of the statistical weights of all states is the partition sum

$$\mathcal{Z} = \sum_{\{s\}} \exp(-\beta\mathcal{H}(s)) \tag{6.1}$$

which contains all information about a (canonical) ensemble of systems and from which all quantities of interest can, in principle, be derived. The tough part of the formalism is to carry out the sum, especially if the system's constituents are interacting and their number N goes to infinity, which is the thermodynamic limit. Typically, the number of possible states grows like a power law in N, and thus the computation of the partition sum becomes increasingly cumbersome.

6.1.2 The Ising Model

As an example of the approach discussed above we consider the *Ising model*, a paradigm of equilibrium statistical mechanics. It is a toy model for a ferromagnet[1], composed of magnetic spins σ on a lattice in a magnetic field h. The spins can take two discrete states and are, in the simplest case, all equal and interact only with nearest neighbors, with a parameter J which gives the strength of the interaction. In this case the Hamiltonian reads

$$\mathcal{H}(s) = -\frac{J}{2}\sum_{\langle i,j\rangle}\sigma_i\sigma_j - h\sum_i \sigma_i \,, \tag{6.2}$$

where the state s is the sequence of states the different spins take $s = \{\ldots, \sigma_{i-1}, \sigma_i, \sigma_{i+1}, \ldots\}$ and the square brackets denote nearest neighbors.

Phase Transition and Critical Exponents

While the Ising model is, in the thermodynamic limit $N \to \infty$, analytically solvable in one and two dimensions, the three dimensional case is only accessible by approximative approaches like the renormalization group (see, e.g. [188]). Importantly, at zero external field h, in two and more dimensions one finds a second-order phase transition at a critical temperature T_c, below which the system has a non-zero spontaneous magnetization $m = \langle\sigma\rangle$, in contrast to $T > T_c$ where the magnetization is strictly zero. One calls T the *control parameter* and m the *order parameter*. Approaching T_c from below, the magnetization goes to zero as

$$m \sim \Delta^\beta \tag{6.3}$$

where $\Delta := T_c - T$ is the distance to the phase transition. *Repose functions*, like the heat capacity C and the magnetic susceptibility χ, describe the change of macroscopic observables upon infinitesimal alterations of the control parameter. Response functions usually diverge like power laws near continuous phase transitions, like that of the Ising model:

$$C \sim |\Delta|^{-\alpha} \,, \tag{6.4}$$

$$\chi \sim |\Delta|^{-\gamma} \,. \tag{6.5}$$

This is the result of the overwhelming importance of fluctuations at the transition itself, vividly reflected in the likewise diverging correlation length

$$\xi \sim |\Delta|^{-\nu} \,. \tag{6.6}$$

[1] A ferromagnet is a material which, for no external magnetic field, has a phase transition at the Curie temperature T_C. Below T_C the ferromagnet has a finite magnetization m, above T_C one recovers paramagnetic behavior with $m = 0$. The Ising model was devised to study this phase behavior.

The correlation length can for example be observed as connected clusters of spins in the same state. At the transition these clusters reach arbitrary large extensions.

The phase transition is characterized by its *critical exponents* β, α, γ, and ν. There exist further observables with different critical exponents, which we will not discuss here. Interestingly, all critical exponents are related to each other by a number of *scaling laws*, so that any set of two of them determines all others.

Though the Ising model appears to be but a caricature of a real ferromagnet, its solutions nonetheless predict the numerical values of critical exponents found for experimental ferromagnetic systems at the phase transition/Curie temperature. What is more, the very same critical exponents have been found for systems, which are, from a microscopical viewpoint, completely different, e.g., near the critical point of a liquid/vapor mixture.

6.1.3 The Scaling Hypothesis and Universality Classes

Universality Classes

How does this uncanny similarity of seemingly unrelated systems arise? It turns out that near the phase transition all microscopic differences between the systems are qualitatively irrelevant. The reason for this is the importance of long-ranged fluctuations, whose effect by far exceeds that of the microscopic interactions: At the phase transition a system's properties are dominated by the diverging length scale, while the microscopic details are summed up in a numerical factor.

Though microscopic details are irrelevant for the critical exponents, not all systems which can undergo continuous phase transitions exhibit the same critical exponents. Systems with identical exponents have the same "kind" of long-range fluctuations. The kind is determined by the dimensionality of the system, the symmetries of the system and furthermore by the number of components a suitable order parameter of the system has. Systems that share these properties also share the same critical exponents and are said to be part of the same *universality class*.

The Scaling Hypothesis

As a phase transition is approached and the typical length scale of fluctuations diverges, fluctuations of smaller extensions are not suppressed. Instead we have fluctuations of *all* possible extensions nested into each other, down to the microscopic scale of the system's constituents. This can already be observed by just looking at a visualization of an Ising model at criticality, where connected spin clusters of any size are present — with a lower cutoff at the distance between two neighboring spins and an upper cutoff at the extension of the system under consideration.

Zooming in, we find that a magnified part of the system has the same statistical properties as the full system, if we stay away from the cutoffs. This is to say, that even though a part of the the system looks different, it has the same statistical properties and is, in this

sense, indistinguishable from another realization of the full system, a property which is called *statistical self-similarity*.

More generally, for a non-critical, but infinite extended system (i.e. $L \to \infty$, where $L \sim N^{1/d}$ is the linear extension and d is the dimension of the system) at finite distance Δ from the phase transition, there exists a finite correlation length ξ according to (6.6). Upon changing the distance to the phase transition according to

$$\Delta \to c\Delta \,, \tag{6.7}$$

with c positive, the correlation length changes as predicted by (6.6):

$$\xi \to c^{-\nu}\xi \,. \tag{6.8}$$

If, say we move away from the phase transient, $c > 1$, the correlation length is decreased. We can, however, recover the numerical value of our original correlation length ξ by decreasing length units according to

$$x \to c^{-\nu}x \,. \tag{6.9}$$

Summarizing, for an infinite system, when simultaneously changing Δ and the unit length x, as stated in equations (6.7) and (6.9), the new system is indistinguishable from the original one, as long as ξ is not comparable to the microscopic length scale. The magnetization m and response functions also change according to (6.3)–(6.5). To recover the numerical values, one has to rescale the units in which they are measured similarly to (6.9), but with their respective critical exponents. Thus, rescaling all units properly, we can map systems at different distance to criticality onto each other. This is known as the *scaling hypothesis*.

Finite Size Scaling

All real-world systems have a finite extension, which means that the typical length scale ξ cannot really diverge in experimental setups. Nonetheless, critical behavior is found close to the critical point $\Delta = 0$, and it is still possible to map systems at different Δ onto each other, if the finite extension is taken into account.

In finite systems, $L < \infty$, the correlation length has a strict upper bound, $\xi \leq L$, and consequently $\xi(\Delta, L)$ now depends on both Δ and L, and so do all other observables. Luckily, we can still assume that for large Δ the finiteness of the system is not reflected in ξ, as fluctuations of the system's size are overwhelmingly rare, while for $\Delta \to 0$ fluctuations spread through the whole system. We may compare the expected infinite-system correlation length ξ_∞ according to (6.6), which can be considered as the natural length scale of the system, with the system size L, to distinguish between the asymptotic regimes. Taking $y := L/\Delta^{-\nu}$, we have

$$\xi(\Delta, L) \sim \begin{cases} \Delta^{-\nu} & y \gg 1 \,, \\ L & y \ll 1 \,. \end{cases} \tag{6.10}$$

When in the regime $y \gg 1$, upon changing Δ and rescaling length units as in (6.7) and (6.9), we expect to recover the original natural correlation length ξ_∞, but have the problem that in

the new length units the system size L has a different value, and therefore y is altered. The solution of course is to also resize the system,

$$L \to c^{-\nu} L \,, \tag{6.11}$$

and thus keep y unchanged. Summarizing, all systems where the ration $L/\Delta^{-\nu}$ is identical are indistinguishable (again keeping away from the microscopic length scale), and the correlation length behaves as

$$\xi(\Delta, L) = \Delta^{-\nu} f\left(L/\Delta^{-\nu}\right) \,, \tag{6.12}$$

where we recover the asymptotics through

$$f(y) \sim \begin{cases} 1 & y \gg 1 \,, \\ y & y \ll 1 \,. \end{cases} \tag{6.13}$$

f is known as a *scaling function* or *scaling form*, and is identical for all models from the same universality class. The procedure by which one reduces the functional dependency of an observable from two variables, Δ and L, to a single variable $y = L/\Delta^{-\nu}$, as just done, is known as *finite size scaling*, and relies on the fact that near the phase transition one can convert length to distance from criticality and vice versa. For more complicated models, which depend on more than two variables, one can perform this method analogously and thereby reduce the number of variables by one.

6.2 Phase Transitions to Absorbing States

Sometimes there is no way back. In ecology, if all representatives of a species have died, the species has become extinct and will usually not come back into existence. Similarly, a forest fire, if extinguished for good, does not flare up again. If a mouse is trapped in a (perfect) mousetrap, it cannot get out any more. All these are examples of absorbing states, which can be entered, but not left.

Processes which can end up in absorbing states can undergo *phase transitions to absorbing states* [186, 187] and are of interest in statistical physics because they are generic examples of non-equilibrium processes. To elaborate this notion, we will discuss *directed percolation*, a representative of this class of models, which is often considered to be of similar importance for absorbing state phase transitions, as the Ising model is for equilibrium phase transitions. Directed percolation is a dynamical version of *isotropic percolation* [189], a geometric model, where each site or bond of a lattice is occupied with possibility p and empty otherwise.[2] In isotropic percolation, above a critical probability p_c, there exists an infinite cluster of adjacent occupied sites, and near p_c many quantities of interest, like the typical cluster size, exhibit critical behavior as discussed in Section 6.1.

[2] In a sense, isotropic site percolation resembles the statistical mechanics model of a paramagnet, where the temperature and external magnetic field determine the fraction of spins pointing upwards. Identifying the magnetization m with the occupation probability p, the two models map to each other. However, the geometric properties of percolation are much harder to derive than the thermodynamic properties of the paramagnet.

6.2.1 Directed percolation

In directed percolation (DP), one thinks of open or closed bonds/sites on a $d+1$ dimensional lattice. Bonds/sites are open with probability p and closed otherwise. Under the influence of gravity, a liquid may flow from an open bond/site to a lower neighboring open bond/site. Consequently, the clusters of sites/bonds filled with the liquid are different from that of occupied sites in isotropic percolation, since a preferential, time-like direction (in which gravity works) was introduced. Thus, one has d spatial directions and one temporal direction and one speaks of a $d+1$ dimensional process. Figure 6.1 depicts one possible realization starting with a single seed initial condition on a $1{+}1d$ square lattice.

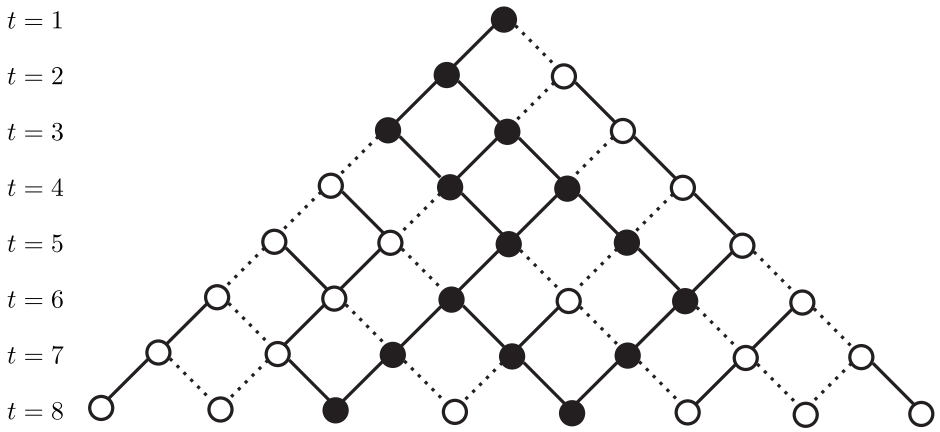


Figure 6.1: An example of a directed percolation process starting on a $1+1$ dimensional square lattice with single seed initial condition. Open and closed bonds are denoted by solid and dotted lines, respectively. Time points downwards, and thus only next-neighbors below occupied sites (full circles) are being occupied in the next time step, while others stay empty (open circles).

In infinite systems ($L \to \infty$), one finds that DP undergoes a phase transition at a critical probability p_c, below which occupied sites always die out after some time and one enters an *absorbing state* with no occupied sites. One takes p as the control parameter of the system. In distinction to isotropic percolation, there is a qualitative difference between the time-like preferential growth direction and space-like directions perpendicular to it: instead of just a single correlation length (i.e. typical isotropic extension of domains) ξ, one finds two distinct correlation lengths, $\xi_{\parallel}$ and $\xi_{\perp}$, parallel and perpendicular to the preferential direction, respectively. They can, between others, be identified with the typical extensions of clusters of non-occupied sites, when above p_c. Typically, one considers two different initial conditions: (i) a fully occupied lattice or (ii) a single occupied site in an otherwise empty lattice. Observables of interest are different depending on the initial condition. For (i) the mean density ρ of occupied sites is of major importance, while for (ii) one usually asks for the probability P_s that there are any occupied sites left and for the mean number of occupied sites N.

In the simplest case, the lattice under consideration is a tilted $(1 + 1)$-dimensional square lattice and update is parallel (as depicted in Figure 6.1). For this case it is easy to see, that

directed percolation can also be considered a reaction-diffusion process on a $1d$ lattice, where particles A may jump (diffuse) to adjacent sites and may undergo different reactions (see also Figure 6.2):

$$A \longrightarrow 2A \quad \text{(reproduction/auto-catalysis)} \tag{6.14}$$
$$2A \longrightarrow A \quad \text{(coalescence)} \tag{6.15}$$
$$A \longrightarrow \emptyset \quad \text{(death/degradation)} \tag{6.16}$$

At a coarse-grained level we may capture this behavior in a stochastic partial differential

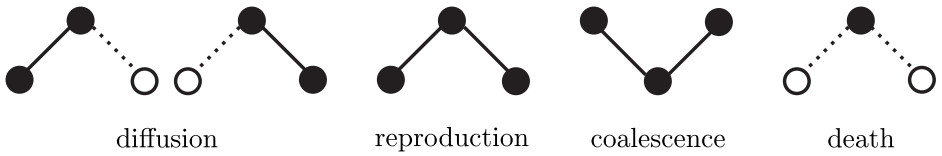


Figure 6.2: Directed percolation can be interpreted as a reaction-diffusion process, made up from microscopic hopping and reaction events. Time points downwards; open and closed bonds are denoted by solid and dotted lines, respectively; occupied and empty sites are depicted as full and open circles, respectively.

equation for the density of particles ρ:

$$\dot{\rho}(x,t) = \tau\rho(x,t) - g\rho(x,t)^2 + D\nabla^2\rho(x,t) + \kappa\sqrt{\rho}\,\eta(x,t) \tag{6.17}$$

Here, the first term, proportional to ρ, corresponds to reproduction and death, and the sign of the parameter τ depends on which reaction is dominant. The second term, proportional to ρ^2, incorporates coalescent events, while the third term results from diffusion of particles. The last term models fluctuations, which are characteristic for the process. They vanish where there are no particles, i.e. for $\rho = 0$. The square root dependence is a result of coarse-graining, since by the central limit theorem, the independent microscopic fluctuations add up to macroscopic Gaussian fluctuations with a standard deviation proportional to the square root of the mean. Consequently, $\eta(r,t)$ is white Gaussian noise,

$$\langle\eta(x,t)\rangle = 0\,, \tag{6.18}$$
$$\langle\eta(x,t)\eta(x',t')\rangle = \delta(x-x')\delta(t-t')\,. \tag{6.19}$$

The parameters τ, g, D, and κ appearing in (6.17) are unknown functions of p, with the restriction that all but τ are strictly positive. It is clear that for $\tau(p_c) = 0$ and $\tau(p < p_c) < 0$, since below $\tau = 0$ there exists no steady-state solution of (6.17) with finite ρ. Equation (6.17) is non-linear, which prohibits an exact solution. Actually, there could be higher order non-linear terms in (6.17), but one can show by renormalization group techniques, that their influence on long-ranged excitations of the system is negligible and, hence, near the phase transitions only the quadratic term is relevant.

6.2.2 Differences to Equilibrium Systems

One might think that DP is not so different from the Ising model, both being two-state lattice gas models. In stark contrast to the Ising model however, it is impossible to assign an energy

to any but the absorbing state of DP.[3] Therefore, it is impossible to compute a partition sum for DP. To understand why the receipt described in Section 6.1 breaks down, it is important to understand the idea of statistical equilibrium: the flow of probability $j_{s\to s'}$ from one micro state s into another one s' is equal to the flow $j_{s'\to s}$ in the opposite direction. The probabilities $p(s)$ and $p(s')$ equilibrate at such levels, that they just balance the transition rates w between the states

$$j_{s\to s'} \equiv w_{s\to s'}p(s) = w_{s'\to s}p(s') \equiv j_{s'\to s} \; . \tag{6.20}$$

This statement is equivalent to equilibrium and known as *detailed balance*. Closely related is the idea of *ergodicity*, which is the notion that for a system in equilibrium, measuring an observable for a sufficiently long time is equivalent to measuring it at a specific moment in a very large number (i.e. a statistical ensemble) of systems.

What happens to these ideas if we look at a process with an absorbing state like DP? The absorbing state can certainly be entered, so there is a flow of probability into it, but as the name suggests, there is no equivalent flow in the opposite direction, which violates detailed balance.[4] When considering ergodicity the problem worsens, since any single system will eventually always enter the absorbing state, as a result of fluctuations (even though they might be very rare). That particular system is not representative for an ensemble of systems anymore, and ergodicity is broken.

From these arguments it is clear, that it is impossible to compute a meaningful partition sum as in (6.1), which carries information about the system. Thus, there is no direct analogue of free energy or response functions. In general it is still possible to evaluate the probability to find a given realization of a process from the microscopic rules, but this becomes unbearable cumbersome for large system and long times, which are necessary conditions to analyze phase transitions. As a result, exact analytical treatment is not feasible and one has to content oneself with approximative approaches, like renormalization group techniques [188], which are challenging in their own right.

6.2.3 Critical Exponents, Universality and Scaling

Even though the standard tools of equilibrium statistical mechanics fail when describing non-equilibrium phase transitions, both classes still bear some striking similarities. Especially one can organize absorbing phase transitions into universality classes according to their critical behavior.

Different models which exhibit phase transitions to absorbing states can be analyzed by using the same, or very similar defined, observables. For sake of clarity we will continue using DP as an example, even though the following statements hold for general absorbing state phase transitions. As prime observables one usually chooses the density ρ (starting from

[3] The energy of the absorbing state would be minus infinity, which is not particularly helpful.

[4] One might argue that there is no problem with detailed balance, and that the absorbing state is the equilibrium state, where all flows are identical zero. This interpretation is equivalent to the tenor of the previous footnote, since any state with an energy of minus infinity is, of course, never left. The argument becomes elusive, however, if there are two or more absorbing states with no flow between them, since, in equilibrium, the relative probability in ending up in either should be related to the difference in their respective energies.

a fully occupied lattice), the survival probability (starting from a single seed) P_s, and the two correlation lengths $\xi_{\parallel}$ and $\xi_{\perp}$ to characterize the system. They are useful as they are easily interpreted graphically and approach steady-state values after initial transients for large enough lattices. At the phase transition they exhibit critical behavior. With the distance to criticality, $\Delta := p - p_c$, they vary as

$$\rho \sim \begin{cases} \Delta^{\beta} & \Delta > 0 \,, \\ 0 & \Delta < 0 \,, \end{cases} \tag{6.21}$$

$$P_s \sim \begin{cases} \Delta^{\beta'} & \Delta > 0 \,, \\ 0 & \Delta < 0 \,, \end{cases} \tag{6.22}$$

$$\xi_{\parallel} \sim |\Delta|^{-\nu_{\parallel}} \,, \tag{6.23}$$

$$\xi_{\perp} \sim |\Delta|^{-\nu_{\perp}} \,. \tag{6.24}$$

One sees that both ρ and P_s fulfill the properties of order parameters. While the corresponding critical exponents are different for general cases of phase transition to absorbing states, they coincide in the special case of DP, $\beta = \beta'$, due to the so-called rapidity-reversal symmetry of DP [190].

As for equilibrium phase transitions, it is possible to organize absorbing state phase transitions into universality classes, according to their set of four, in general independent, exponents defined through equations (6.21)–(6.24). Again, the classes are distinguished by symmetries and dimensionality of the model and on properties of the order parameter, but not by the microscopic details, since again, near the transition, long-range behavior dominates. One important finding is that a wide variety of models falls into the DP universality class, which is therefore considered very robust. Nonetheless, there exist a number of other known classes, see, e.g., [186, 187].

6.3 Surface Growth

Not all non-equilibrium models necessarily undergo phase transitions, but many still have interesting properties. In this section we consider *surface growth* models [152, 191], which inherently display critical behavior, in the sense that typical length scales of surface fluctuations grow without bounds if not put to a stop by a finite system size.

Growth processes are widespread in nature: Water soaking through a paper towel, fire burning down a dry cornfield, a bacterial colony growing and expanding on a surface (as introduced in Section 5.2), atoms being deposited on a substrate surface by molecular beam epitaxy. All these examples have in common that there is a surface which grows into an "empty" region as time progresses. This surface may seem smooth, even flat, from far away, but on smaller length scales it has an irregular, rough shape. This would not seem very remarkable if these surface fluctuations were only present at the level of microscopic constituents of which the "growing medium" is composed. However, the typical extension of fluctuations grows in time and subsequently ever larger segments of the surface display fluctuations of the same kind, meaning they are *self-affine* in a statistical sense. Intuitively self-affinity means that enlarging part of a system, but stretching by different (appropriate) factors in the different directions,

one receives a new system with statistically identical properties.[5] Statistical mechanics aims to study models that are as simple as possible, but still exhibit these characteristic features.

6.3.1 Linear Models

Temporal Increase of the Width: Random Deposition

The simplest model, *random deposition*, considers an $L \times |\mathbb{N}|$ square lattice, where the L direction is usually considered as horizontal (and subject to periodic boundary conditions); thus, there are L half-infinite columns. The lowest site of each column is initially occupied by a particle, and in each time step a random column is chosen (equally distributed), where a new particle is added and falls down to the lowest unoccupied site; for an example see Figure 6.3. The number of deposited particles N is therefore proportional to time and so is the

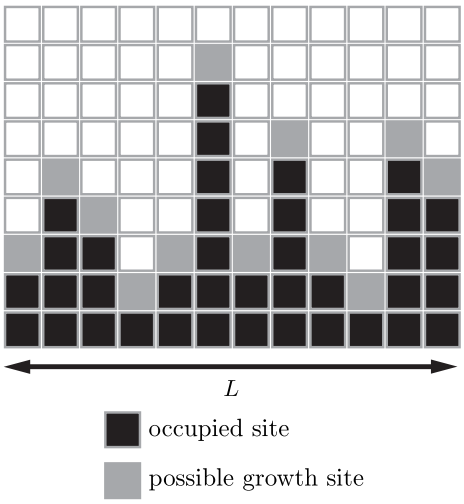


Figure 6.3: A realization of random deposition on a lattice with linear extension $L = 12$. Particles are added to a random column and fall down until they are just above an occupied site, where they stick. A present configuration is shown in black; in the next step a single particle will be added to one of the gray sites with equal probability.

average height of the interface,

$$\bar{h}(t) := \frac{1}{L} \sum_{i=1}^{L} h(i,t) = N(t)/L \,, \tag{6.25}$$

where $h(i,t)$ is the vertical position of the highest occupied lattice site in column i at time t. The *roughness* of the surface is quantified by the *width*

$$w(L,t) := \left(\frac{1}{L} \sum_{i=1}^{L} \left[h(i,t) - \bar{h}(t) \right]^2 \right)^{\frac{1}{2}} \,, \tag{6.26}$$

[5] If the factors are identical for all directions one has a *self-similar* or *fractal* system.

which is easily interpreted as the standard deviation of the height about its mean. For random deposition, all columns are independent of each other, and it is easy to show that the number of particles in each column (i.e. its height), is binomial distributed. This implies that the standard deviation of the height grows like the square root of the mean height. Thus, the ensemble averaged[6] width grows as

$$w \sim t^{\beta} \tag{6.27}$$

where $\beta = 1/2$ is the *growth exponent* of random deposition.[7] It turns out that the interface generated by random deposition is not self-affine, since height fluctuations all have the same order of magnitude, and do not scale with the considered length scale. This results from the lack of correlation between the columns.

Saturation of the Width: Surface Relaxation

The independence of neighboring columns in random deposition seems rather unnatural as it permits steep steps of arbitrary size to build up along the front. To remedy this, one considers random deposition with *surface relaxation*. In contrast to the previous example, particles may move to neighboring columns, if these have a lower height, and continue to do so until they reach a locale minimum. This smoothens the surface immensely, as evident from a reduced roughness exponent, $\beta = 1/4$. More importantly, the width does not grow without bounds, but instead saturates to a finite value $w_{sat}(L) := w(L, t \to \infty)$. For different lattice sizes L one finds

$$w_{sat}(L) \sim L^{\alpha} \,, \tag{6.28}$$

where $\alpha = 1/2$ is the *roughness exponent* for random deposition with surface relaxation. Height fluctuations of all magnitudes lower than w_{sat} are also present on shorter horizontal length scales, and the system is now self-affine. Between the two regimes (given by equations (6.27) and (6.28)) lies a crossover, which appears as the system "feels" its finite size. This should be the case where $t^{\beta} \sim L^{\alpha}$, which gives us the *crossover time*

$$t_{\times} \sim L^{\alpha/\beta} \,, \tag{6.29}$$

where α/β is called the *dynamical exponent* of the growth model, as it interrelates the time-dependent and the size-dependent regime. The set of critical exponents $\{\alpha, \beta\}$ characterizes a given universality class of surface growth models, which have symmetries and dimensionality in common. For random deposition with surface relaxation we have $\alpha/\beta = 2$.

Scaling

The crossover to saturation appears at the time when typical horizontal extensions $\xi(t)$ of the height fluctuations, which grow like $t^{\beta/\alpha}$, reach the system size L, where they necessarily

[6] From here on we will always assume ensemble averaged observables if not stated otherwise.

[7] The growth exponent β should not be confused with the critical exponent of the order parameter in the context of phase transitions (as introduced in Sections 6.1 and 6.2), which is denoted by the same symbol.

saturate. After this time all sites are correlated with each other, which means that no section of the front can lag behind arbitrarily far (or run away, for that matter). As in the context of finite size scaling (see Section 6.1.3), different systems can be related to each other by comparing ξ and L or similarly t and $t_\times$:

$$w = t^\beta \hat{w}(t/t_\times) \ , \tag{6.30}$$

where $\hat{w}(y)$ is a universal scaling form which recovers the asymptotic regimes,

$$\hat{w}(y) \sim \begin{cases} 1 & y \ll 1 \ , \\ y^{-\beta} & y \gg 1 \ . \end{cases} \tag{6.31}$$

This scaling behavior is known as the Family-Vicsek scaling relation [192].

Coarse Graining — Stochastic Continuum Equations

The correlations between different columns render an exact analytical solution of the lattice model infeasible. However, we can coarse-grain the model, using its symmetry properties. We want to obtain a stochastic differential equation of the general form

$$\dot{h}(x,t) = F(h,x,t) + \eta(x,t) \ , \tag{6.32}$$

where $\eta(x,t)$ are stochastic fluctuations modeled as Gaussian white noise,

$$\langle \eta(x,t) \rangle = 0 \ , \tag{6.33}$$

$$\langle \eta(x,t)\eta(x',t') \rangle = 2D\delta(x-x')\delta(t-t') \ , \tag{6.34}$$

and D gives the noise amplitude. The explicit form of F depends on the symmetries, which tell us under which transformations equation (6.32) is invariant. Ballistic deposition with surface relaxation[8] is expected to be invariant under:

- translation in time; $t \to t' = t + t_0,\ x \to x,\ h \to h$: The growth rules do not depend explicitly on time, but only implicitly through $h(x,t)$. All terms which explicitly depend on t are therefore prohibited, while terms depending on $h(x,t)$ or on temporal derivatives of h are allowed.

- translation in space; $t \to t,\ x \to x' = x + x_0,\ h \to h$: All horizontal positions/columns are equal, growth rules do not depend explicitly on x, therefore F does not contain terms which depend explicitly on x, but spatial derivatives are allowed.

- translation in height; $t \to t,\ x \to x,\ h \to h + h_0$: Growth rules do not depend explicitly on the height of the interface. Consequently, no explicit dependence of F on h is allowed. Derivatives are possible.

- inversion about the vertical direction (in higher dimensions this generalizes to rotation); $t \to t,\ x \to x' = -x,\ h \to h$: Left and right are equal, there is no sidewards drift of particles. This rules out terms with an uneven number of spatial derivatives, e.g., $\nabla^n h$ and $(\nabla h)^n$ with n uneven.

[8] Surface relaxation breaks symmetry under inversion of time, which is still present for ballistic deposition alone.

- inversion about the horizontal direction (up/down symmetry for h); $t \to t$, $x \to x$, $h \to -h$: protrusions and indentations of the surface relax and build up in an identical fashion, which is a typical equilibrium property (as found, e.g., for a polymer or membrane in a heat bath). This forbids, e.g., all terms proportional to $(\nabla h)^n$, where n is even, since these terms do not change sign under the transformation.

The above analysis allows terms proportional to $(\nabla^{2n} h)(\nabla h)^{2m}$. We are interested in scaling properties on large scales ($x \to \infty$) and after transients have decayed ($t \to \infty$). In this limit derivatives of higher order are negligible, compared to lower order ones with the same symmetry properties, as can be argued for with scaling arguments or shown by renormalization group techniques.

We are left with the two terms of lowest order for F in equation (6.32), which are a constant, v, and a term proportional to $\nabla^2 h$. v corresponds to a mean growth velocity, and can be taken care of by considering a co-moving reference frame, $h \to h + vt$. This leaves us with the Edwards-Wilkinson (EW) equation

$$\dot{h}(x,t) = \nu \nabla^2 h + \eta(x,t) \; , \tag{6.35}$$

where ν is a parameter, which corresponds to the strength of "surface tension". From the Edwards-Wilkinson equation one can infer the exponents $\alpha = 1/2$ and $\beta = 1/4$ by scaling arguments, or since it is linear, through Fourier transformation.

6.3.2 Non-linear Models

Though the EW equation describes a universality class of surface models which have self-affine interfaces it is not very realistic. In fact, the examples introduced to motivate surface growth processes at the beginning of this section all differ, in that they grow perpendicular to the *local* orientation of the surface.

Ballistic Deposition

An extension of random deposition, called *ballistic deposition* incorporates growth normal to the surface by introducing sticky particles. These fall down a column until they encounter a particle at the site exactly below *or* left or right of them, in which case they stick. This generates sidewards growth and lateral correlations of the surface above a bulk containing many holes and overhangs. An example is given in Figure 6.4. For ballistic deposition one finds a different set of exponents ($\alpha = 1/2$, $\beta = 1/3$), since the sidewards growth breaks the up/down symmetry, which indicates that there might be additional relevant terms that have to be added to the EW equation (6.35). Before we elaborate this idea, we will take a look at another growth model, which is more relevant to this work and has the same set of exponents as ballistic deposition and is, therefore, member of the same universality class.

The Eden Model

The microscopic rules and appearance of the *Eden model* [153], which was devised to mimic the growth of bacterial colonies, are quite different from the models introduced so far in this

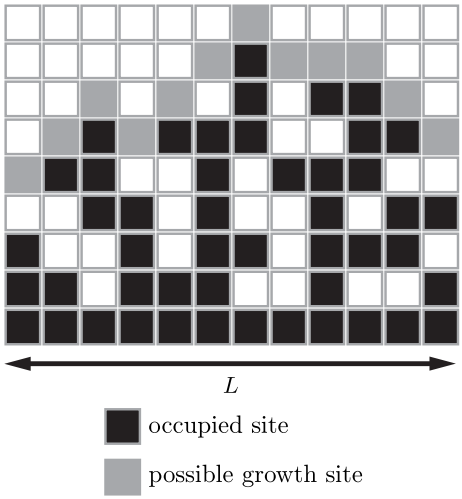


Figure 6.4: A realization of ballistic deposition on a lattice with linear extension $L = 12$. Particles are added to a random column and fall down until they encounter an occupied site below or to the right or left of them, in which case they stick. The possibility of sticking to a particle at a neighboring column introduces growth which is locally perpendicular to the surface. A present configuration is shown in black; in the next step a single particle will be added to one of the gray sites with equal probability.

section. Here an empty side, adjacent to an occupied site is chosen at random and a new particle is placed there. In this case the bulk is compact and covered by a thin layer with holes, which are rapidly filled. An example is given in Figure 6.5. There exist three versions of the Eden model, which differ in the relative probabilities with which a given possible growth site is chosen [193]:

- Version A: All empty sites with at least one next-neighboring occupied site have the same probability, choose one at random.
- Version B: Choose a connection/bond between an occupied and an empty site at random, with equal probability. Add a particle at the empty site.
- Version C: Choose any occupied site at the surface (i.e. next-neighboring to an empty site) at random with equal probability. Then choose randomly and with equal probability one of its empty next-nearest neighbors and add a particle there.

The difference in the update rules mainly influences how fast holes and overhangs are filled and also has some influence on the amplitude of the width of the front. The scaling behavior is the same for all three versions, when ignoring transients and the strength of finite size effects. Version C, which focusses on the particles which are already there, shows best scaling and is also closest to biology, where the particles are proliferating microbes. Interestingly, experiments with prokaryotes [154], found exponents different from those of the Eden model ($\alpha = 1/2$, $\beta = 1/3$), while a recent study with eukaryotic cells [155] is in excellent agreement with these values. The origin of this qualitative difference remains a challenging problem at the border between microbiology and statistical physics.

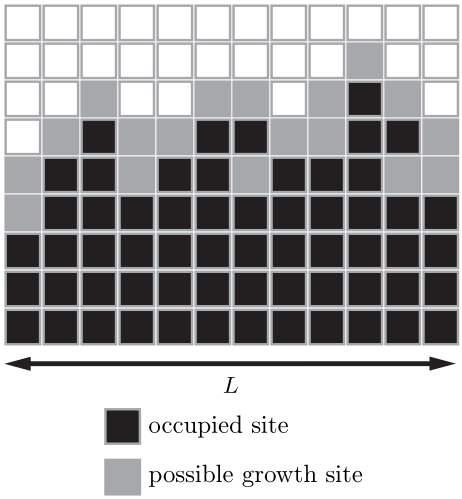


Figure 6.5: A realization of the Eden model on a lattice with linear extension $L = 12$. Particles are added to a random site with a next-neighboring occupied site. A present configuration is shown in black; in the next step a single particle will be added to one of the gray sites. The relative probabilities to choose a given growth site depend on the version of the Eden model (see main text). As for ballistic deposition growth is normal to the local surface.

The Kardar-Parisi-Zhang Equation

Though very different in detail, both ballistic deposition and the Eden model are members of the same universality class, which they share with a number of other models, e.g., solid-on-solid models. Due to normal growth they have a growth exponent $\alpha = 1/3$ in distinction to representatives of the EW universality class. The reason for this is that growth normal to the surfaces induces *lateral growth* of protrusions of the interface. This breaks symmetry under

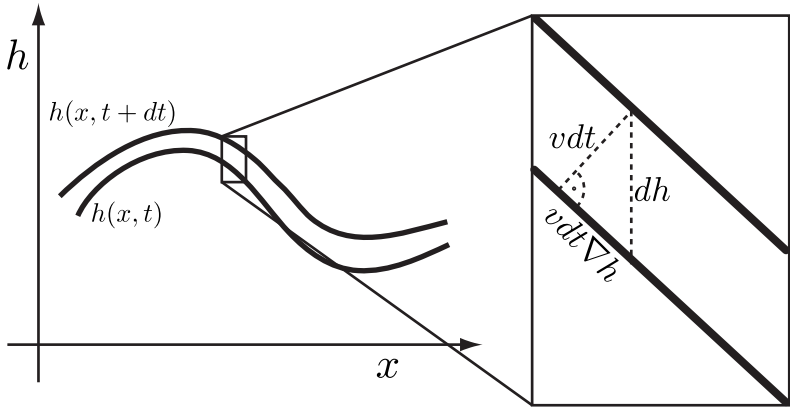


Figure 6.6: Growth normal to the local surface leads to an additional growth term proportional to $dh \sim (\nabla h)^2$.

inversion about the horizontal direction, and therefore requires a different coarse-grained

growth equation. From Figure 6.6 we infer

$$dh = \left[(vdt)^2 + (vdt\nabla h)^2\right]^{1/2} + \mathcal{O}(dt^2) \tag{6.36}$$

$$= vdt\left[1 + (\nabla h)^2\right]^{1/2} + \mathcal{O}(dt^2) \tag{6.37}$$

$$= \left[v + \frac{v}{2}(\nabla h)^2\right] dt + \mathcal{O}\left(dt^2, (\nabla h)^4\right) \ . \tag{6.38}$$

If the inclination of the surface is not too large[9], $\nabla h \ll 1$, this means we can incorporate growth normal to the surface by adding a term proportional to $(\nabla h)^2$ to the Edward-Wilkinson equation (6.35). This yields the famous Kardar-Parisi-Zhang (KPZ) equation [194]:

$$\dot{h}(x,t) = \nu\nabla^2 h + \frac{\lambda}{2}(\nabla h)^2 + \eta(x,t) \ . \tag{6.39}$$

The additional non-linear term in the equation is evidence for the non-equilibrium nature of the KPZ equation, as, e.g., the average height is no longer conserved in contrast to the EW equation (6.35). The non-linear term also prohibits analytical solution of the KPZ equation, and analysis is thus restricted to renormalization group techniques. In $1d$ one finds the exponents as measured in simulations of lattice models, $\alpha = 1/2$ and $\beta = 1/3$, which can also be argued for by more qualitative arguments.

There exists an extensive literature devoted to the study of the KPZ equation and its solutions as well as many extensions, see, e.g., [195, 152, 191]. Since this is not in the focus of the present work, we will not discuss it here. Instead we will examine some properties of multi-species Eden models.

6.4 Multi-Species Eden Models: Coupling of Bulk and Surface Dynamics

In Section 5.2.2 we shortly discussed the growth of microbial colonies composed of multiple strains or species. In such settings new interesting questions arise, which can be addressed by multi-species growth models.

6.4.1 Super-diffusion of Sector Boundaries

The surface roughness which is inherent to growth models like the Eden model has consequences when instead of just one there are two or more distinct kinds of particles. A two-species generalization of the Eden model [163], exemplifies coarsening of single-species domain at the surface. Coarsening happens as boundaries between the domains annihilate as they meet. The speed at which the number of domains decreased, in the case of equally fast reproducing species, can be explained by a super-diffusive meandering of the domain boundaries. This is the result of locally normal growth of the rough surface, which is imprinted in the

[9] This is always the case on large enough length scales, if the roughness exponent α is smaller than unity, since then the typical slope of the coarse grained surface decreases when zooming out.

trajectories of the domain boundaries, as has been discussed in Section 5.2.2 (see Figure 5.1). Experiments with fluorescently labeled bacteria corroborate these results [161].

This finding of strong, surface induced, boundary fluctuations becomes more important if a neutral two-species Eden model, which is started with different initial conditions, is considered [196]. It turns out that the scaling of the boundaries' trajectories is strongly dependent on the curvature of the front and can reach from sub-diffusive up to ballistic behavior. While one may argue that such initial conditions only have a transient influence on the dynamics, this must not be true for the non-neutral case, where domain boundaries may be correlated with local curvature.

6.4.2 Surface-Bulk-Coupling under the Influence of Selection and Mutations

Some studies of non-Eden growth models have found that surface roughness and domain dynamics influence each other [197, 198, 199, 200, 201, 202, 203, 204, 205, 206, 207, 208]. One of them [208], finds a phase transition to an absorbing state, which belongs to the DP universality class. However there is only limited literature on Eden models with multiple species [209, 160] besides that mentioned above, and, as it appears, none involving mutations.

In the manuscript "Range Expansion with Mutation: Dynamical Phase Transition in a Two-Species Eden model", by J.-T. Kuhr, M. Leisner, and E. Frey we introduce a two-species Eden model incorporating selection and unidirectional mutation. The irreversibility of mutations creates an absorbing state, where only mutants are at the front, while selection increases the surface roughness and entails correlations between domain boundaries and curvature of the surface. This interplay of roughness and domain dynamics alters properties of both the phase transition and of dynamic surface roughening, as manifest in critical exponents different from the expected DP and KPZ universality classes, respectively.

New Journal of Physics
The open-access journal for physics

Range expansion with mutation and selection: dynamical phase transition in a two-species Eden model

J-T Kuhr, M Leisner and E Frey[1]
Arnold Sommerfeld Center for Theoretical Physics (ASC) and Center for NanoScience (CeNS), Ludwig-Maximilians-Universität München, Theresienstraße 37, 80333 München, Germany
E-mail: frey@lmu.de

New Journal of Physics **13** (2011) 113013 (28pp)
Received 4 August 2011
Published 11 November 2011
Online at http://www.njp.org/
doi:10.1088/1367-2630/13/11/113013

Abstract. The colonization of unoccupied territory by invading species, known as range expansion, is a spatially heterogeneous non-equilibrium growth process. We introduce a two-species Eden growth model to analyze the interplay between uni-directional (irreversible) mutations and selection at the expanding front. While the evolutionary dynamics leads to coalescence of both wild-type and mutant clusters, the non-homogeneous advance of the colony results in a rough front. We show that roughening and domain dynamics are strongly coupled, resulting in qualitatively altered bulk and front properties. For beneficial mutations the front is quickly taken over by mutants and growth proceeds Eden-like. In contrast, if mutants grow slower than wild-types, there is an antagonism between selection pressure against mutants and growth by the merging of mutant domains with an ensuing absorbing state phase transition to an all-mutant front. We find that surface roughening has a marked effect on the critical properties of the absorbing state phase transition. While reference models, which keep the expanding front flat, exhibit directed percolation critical behavior, the exponents of the two-species Eden model strongly deviate from it. In turn, the mutation-selection process induces an increased surface roughness with exponents distinct from that of the classical Eden model.

[1] Author to whom any correspondence should be addressed.

Contents

1. Introduction

The spatial structure strongly influences the dynamics of evolving biological systems and often gives rise to qualitatively different outcomes as compared to well-mixed populations [1, 2]. Over the last few years, microbiological experiments have progressively been used to shed light on the dynamics of spatially extended systems [3–9], and have spurred theoretical investigations [10–16]. Since microbial colonies typically grow from some initial seed, a particularly interesting question to ask is: how are both a colony's morphology *and* its internal composition shaped by the growth rates of the different strains it is composed of and by the interactions between these strains. Already the simplest scenario, the spreading of two selectively neutral strains or species shows intriguing phenomena, which makes the evolutionary outcome quite distinct from well-mixed populations [17–22]. Experimental investigations of expanding *Escherichia coli* and *Saccharomyces cerevisiae* colonies containing two fluorescently labeled but selectively neutral strains have shown that the population differentiates along the growing front and thereby segregates into well-defined domains [18, 22]. This is caused by demographic fluctuations: since mainly cells at the leading front of the growing colony access nutrients and reproduce, the effective population size is small and neutral dynamics leads to local fixation of strains and thereby generates sectoring of the population [22].

In general, however, microbial communities are heterogeneous and composed of multiple strains, which may have different growth rates or show other kinds of distinct phenotypic features; cf e.g. [16, 23–28]. One well-known phenomenon, which is the focus of this paper, is bursts of new sectors of mutants during the growth of bacterial colonies [3, 22, 29–31]. In a growing bacterial colony, which initially consists of one phenotype (the wild-type) only, mutations of this strain may appear during the reproduction of individuals. If these mutations happen at the leading front and are beneficial, i.e. if the mutant strain has a larger growth rate than the wild-type strain, mutant regions along the front not only advance faster but also expand laterally. Consequently, mutant sectors take over ever larger parts of the front in a quasi-deterministic fashion [22, 32, 33]. While deleterious mutations are, for the same reason, handicapped by selection, they may still form large clusters along the front if they appear frequently enough. In this case mutant sectors are no longer spatially separated, but may

coalesce. At a critical mutation probability, the mutants' selective disadvantage is effectively balanced and mutants may take over the front [22]. If back-mutations are prohibited, the front remains trapped in an all-mutant state. In the language of non-equilibrium statistical mechanics, the critical mutation rate marks a phase transition between an active state, for which the front is composed of wild-types and mutants, and an absorbing homogeneous state, composed of mutants only [34–37].

What makes this absorbing state phase transition in a growing bacterial colony interesting is the intricate interplay between the morphology of the growing front and the evolutionary dynamics of the colony. Depending, among others, on the particular type of bacterial strain, nutrient concentration and softness of the agar surface, bacterial colonies exhibit a kaleidoscope of possible morphologies [3, 4, 6, 30, 38–40]. The Eden model [41] has been devised to describe the growth of bacterial cultures with a compact morphology on which we focus in this work. A hallmark of this model is the generation of rough fronts with characteristic features closely resembling recent experimental observations [42]. The roughness of the front directly affects the trajectory of interfaces between domains of different strains [32]: the undulations of the front are imprinted in the meandering of the domain boundaries on all length scales, as has recently also been observed experimentally [18, 22]. This surface-induced meandering speeds up coalescence of clusters, i.e. the roughness of the propagating front strongly affects the temporal evolution of the population's composition. This suggests that front roughening is highly relevant for the nature of the phase transition from a heterogeneous to a homogeneous population at the expanding front.

It is precisely this issue that we would like to address in this paper. To this end we study bacterial range expansion using a two-species Eden model, which incorporates surface roughness, selection and irreversible mutations. We intend to gain deeper insight into the interplay of these key features of the dynamics and their relative importance for the transition to the absorbing state. While our findings are mainly of general importance for a broader class of multi-species growth models, we also expect that real-world range expansions of bacterial colonies are subject to this coupling and carry its signature in the evolving patterns. In the remainder of the introduction, we give a concise overview of previous work on surface roughness and absorbing state phase transitions as relevant for this work. A more in-depth discussion and comparison with the results of our work is given in the final section.

Both, discrete numbers and roughness of the expanding front are intrinsic to surface growth models [43–45], which mimic the stochastic advance of particles into empty space. One particular growth model, the Eden model [41] (of which there are three, slightly different, variants [46]) has been devised to describe the growth of bacterial cultures. Mesoscopically its evolution is captured by the Kardar–Parisi–Zhang (KPZ) equation [47]. The KPZ equation constitutes a robust universality class which incorporates many surface growth models, like e.g. ballistic deposition and solid-on-solid models. There have been a number of generalizations to multi-species growth models [32, 48–60], notably one of the Eden model that incorporates selection [32]. The coupled influence of mutations and selection on kinetic surface roughening, which is one of the topics of this work, has not been analyzed in detail so far. Even when neglecting roughness, multi-species propagation is not treated easily in more than one dimension. The reason for this is the intricate interplay of creation, annihilation and merging of clusters, which contain only one kind of individual, at the leading front of the colony.

In the case of irreversible mutations, both analytical results and numerical simulations for range expansion with flat fronts (neglecting surface roughness) predict a transition to an

all-mutant absorbing state, which emerges even for deleterious mutations at a critical mutation probability [22]. For such models with flat fronts [22], the dynamics closely resembles that of a contact process [61]. As a consequence, it belongs to a broader class of absorbing state phase transitions whose main representative is directed percolation (DP) [62], a dynamic version of percolation [63]. The DP universality class of phase transitions to absorbing states has been found to display enormous robustness with respect to alterations of the microscopic update rules and is considered a paradigm of non-equilibrium statistical physics [34–37]. How roughness of the front may influence phase transitions to absorbing states has, to the best of our knowledge, not been addressed previously.

This paper is organized as follows. In section 2, we introduce a two-species growth process on a 2d lattice to analyze the general properties of range expansion of asexually reproducing microorganisms. It explicitly includes irreversible mutations from wild-types to mutants and selection between the two strains. The evolving system's morphology intimately depends on the antagonistic effects of new mutant domains being created at the front and others losing contact with it. For abundant, deleterious mutations a phase transition to an absorbing state exists, which changes both the evolutionary dynamics and the surface roughening behavior of the system qualitatively. We discuss the properties of both surface and bulk morphology and map out the phase diagram in section 3. The system's critical behavior near the transition is affected by the front's roughness, since the temporal evolution of the system is restricted to the growing front. This alters the critical properties of the absorbing state phase transition. In section 4.1 we determine its critical exponents, which are different from those of the DP class (which is most often found for flat systems [34–37]). In addition, the different birth rates of the two strains induce an enhanced width of the front near the phase transition. Close to the transition the roughness exponent of our model is severely enhanced compared to that of the Eden model [43–45]. In addition to KPZ behavior we identify and characterize a critical roughening regime in section 4.2. We conclude with a discussion of our results and a comparison to related models which study the coupling between surface roughening and domain dynamics.

2. The Eden model with mutations

Range expansion into hitherto unoccupied territory proceeds in a non-homogeneous manner on the length scale of individuals. Along the leading front local protrusions emerge randomly and subsequently expand, thereby creating a rough front and an overall forward movement. The main features of this growth process are well captured by the classical Eden model, which was developed to mimic the growth of microbial colonies [41, 44, 45]. While some multi-species extensions of the Eden model have been analyzed [32, 48–50], surface growth experiments with competing microorganisms have been performed only recently [18, 22]. They reveal intriguing, nontrivial patterns if the population is comprised of distinguishable sub-populations. Mutations and selection can alter the growth dynamics by giving some individuals a growth advantage or by introducing qualitatively different organisms.

In this work, we employ a lattice gas model to analyze the influence of mutations and selection on range expansion at rough, fluctuating fronts. We model microbial range expansion with mutations as a cellular automaton on a 2d semi-infinite square lattice of extensions $L \times \infty$ with periodic boundary conditions in the transverse direction (see figure 1).

At a given time t, each site (i, j), with $i \in \{1, \ldots, L\}$ and $j \in \mathbb{N}$, is either empty or occupied with an individual, which in turn can be either a wild-type or a mutant. We identify empty

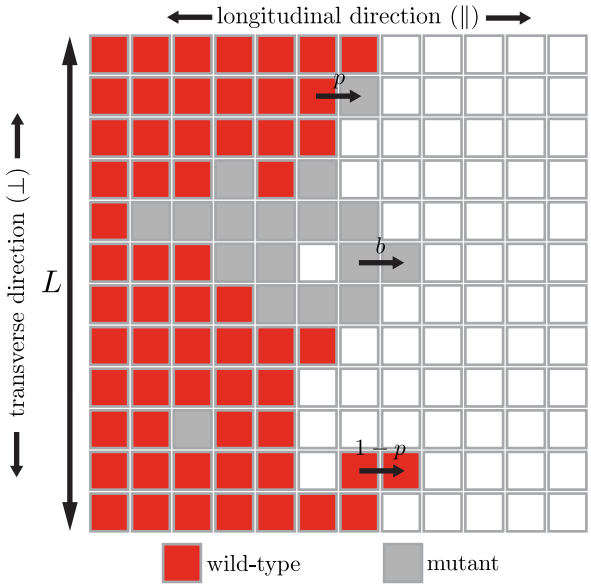


Figure 1. The Eden model with mutations. An initially line-shaped bacterial colony of length L, consisting of only wild-types, shown in dark gray (red), grows into an empty half-space. Wild-types reproduce at rate 1, given that they have a vacant nearest-neighboring lattice site. Individual offsprings are either wild-types or mutants (light gray) with probabilities $1-p$ and p, respectively. Mutants reproduce like wild-types but at birth rate b. Back-mutations are prohibited. In the L-direction (transverse) periodic boundary conditions apply.

sites, wild-types and mutants with state variables $s=0$, $s=1$ and $s=-1$, respectively. The state of the system at a given time is, therefore, specified by the set of occupation numbers $s_{i,j} \in \{-1, 0, 1\}$. The system evolves as individuals reproduce: empty sites with $s=0$ can change their state if an individual on a nearest-neighboring site reproduces. We assume that individuals do not die and hence any site with $s=\pm 1$ remains in its state indefinitely.

To implement random-sequential update of a configuration, we apply a simplified version of Gillespie's stochastic algorithm [64]. Only individuals at the front, defined as the set of occupied sites with at least one empty nearest-neighbor site, can reproduce. Of these, an individual is randomly, but proportional to its birth rate, chosen to reproduce: mutants reproduce with relative birth rate $b \geqslant 0$, while the birth rate of wild-types is 1 and thereby sets the timescale. Let N_{wt} and N_{mut} denote the number of wild-types and mutants with free neighbors, respectively. The overall birth rate of the population (at which production events happen) is given by $b_{\mathrm{tot}} := (N_{\mathrm{wt}} + bN_{\mathrm{mut}})$. To account for different birth rates, each wild-type individual with an empty neighbor is chosen to reproduce with probability $1/b_{\mathrm{tot}}$, while each mutant individual with an empty neighbor is chosen with probability b/b_{tot}. The new individual is placed on a random empty nearest-neighbor site of the individual which has been chosen to reproduce. During wild-type reproduction, a mutation may happen with probability p. Thus, if the reproducing individual is wild-type ($s=1$), the offspring is a wild-type with probability $1-p$ and a mutant with probability p. If the reproducing individual is a mutant ($s=-1$), the offspring is necessarily also a mutant. Note that by these rules back-mutations and multiple

mutations are prohibited and that all mutants are identical and reproduce with the same birth rate b. Assuming exponentially distributed reproduction times, the expectation time until the next reproduction event is b_{tot}^{-1}. Hence, we update time by $t \to t + b_{\mathrm{tot}}^{-1}$ whenever an individual reproduces. Since, after some initial transient period, the average front position moves at constant velocity, one may take the longitudinal coordinate j as a proxy for time t.

The model, as described above, is a generalization of version C of the Eden model as introduced by Jullien and Botet [46]. It is the biologically most realistic version, as it focuses on occupied sites, i.e. individuals, rather than on empty sites (version A) or bonds between adjacent occupied and empty sites (version B). In the limit of vanishing mutation rate our model reduces to the model of Saito and Müller-Krumbhaar [32]. While we consider the case of a homogeneous initial front with uni-directional mutations, they analyzed the temporal evolution of an initially heterogeneous front in the absence of mutations; see section 5 for more details. If not stated otherwise, we use a line of wild-type particles as the initial condition, i.e. $s_{i,j} = \delta_{j,1}$. Since diffusion is not included in the model, surface configurations become frozen in the bulk, as observed for patterns in range expansion experiments [18].

3. Phenomenology and phase diagram

3.1. Phenomenology

We now turn to a phenomenological description of the morphology of the evolving colony, as obtained from stochastic simulations; a representative realization of the range expansion process is shown in figure 2(a). Starting from an initial line of wild-types the growth front moves forward and, as a result of the stochastic individual birth processes, some parts of the front expand more rapidly than others. This leads to front *roughening* which first appears on length scales of the lattice constant, but as time progresses, the typical size of protrusions and indentations of the front grows both longitudinally and laterally.

As the range expansion proceeds, *mutation events* occur where a wild-type individual gives birth to a mutant, whose direct descendants create a new mutant sector growing between two wild-type domains. Mutant and wild-type domains are separated by domain boundaries. Since the reproduction rates of mutant and wild-type individuals differ in general, the boundary between their respective growth sectors performs a *biased random motion*. While for beneficial mutations, with $b > 1$, sectors consisting of mutants broaden on average, they have a tendency to decrease in size for deleterious mutations where $b < 1$. Fluctuations in the trajectory of the boundary arise mainly for two reasons: on short length scales, they are due to the intrinsic stochasticity of the birth events. On larger scales, roughening of the front drives a super-diffusive meandering of the sector boundaries [18, 32, 48]: as the population locally always expands normal to the front, the roughness of the front is imposed on the trajectories of the sector boundaries.

While the domain boundaries move transversely through the system, they may encounter other boundaries, resulting in mutual annihilation and an ensuing merging of domains. There are two distinct types of coalescence events of domain boundaries; cf figure 2(a): either boundaries of *different* mutant clusters meet such that they merge and form a larger mutant cluster, or two boundaries of the *same* cluster meet, in which case a mutant cluster loses contact with the growing front and is trapped by wild-types. Note that boundaries are always created and annihilated pairwise, and regions of wild-types and mutants alternate at the front.

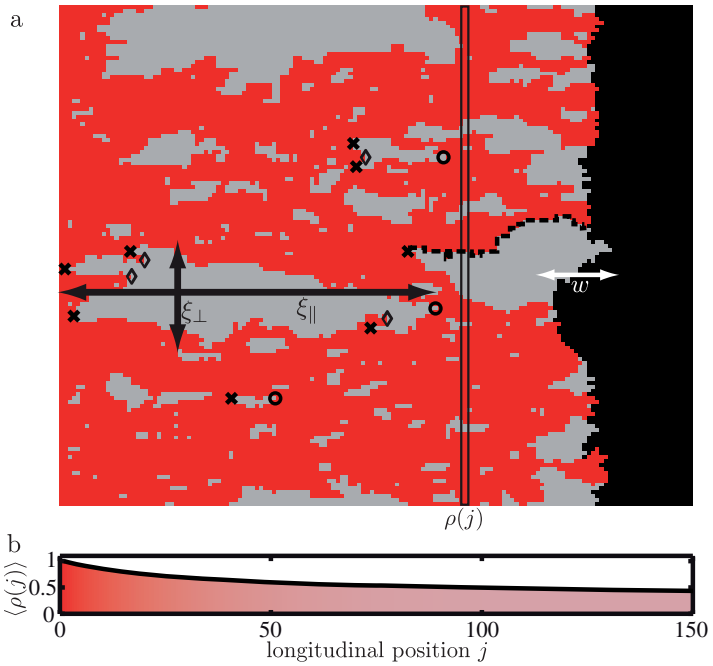


Figure 2. Phenomenology of the model. (a) Morphology. Wild-types are shown in dark gray (red); mutants in light gray. During range expansion mutants are created (examples indicated by crosses, $\times$), which in turn reproduce and thereby create mutant clusters. The lateral boundaries (an example is given by the dashed black line) of mutant clusters are created pairwise and perform biased, super-diffusive random walks. Mutant clusters are either outgrown by the wild-types (examples indicated by circles, $\circ$) or merge with other clusters (examples marked with diamonds, $\Diamond$), forming even bigger mutant clusters. In both cases a pair of boundaries annihilates. In the bulk, the presence of mutant clusters is discernible as a density of wild-type sites $\rho(j) < 1$ at longitudinal position j. Mutant clusters in the bulk are characterized by their typical longitudinal and transverse extensions, which are identified with the correlation lengths $\xi_\parallel$ and $\xi_\perp$, respectively. The surface roughness of the population front is characterized by the width w, defined as the standard deviation of the front position. For this realization on a lattice of length $L = 128$, we used birth rate $b = 0.9$ and mutation probability $p = 0.016$. (b) Averaged observables. By averaging over many independent realizations of the growth process, fluctuation effects are suppressed and mean observables can be defined. Shown here is the mean density $\langle\rho(j)\rangle$ as a function of longitudinal position j, averaged over 500 realizations for the parameters used in (a). For large j, the density of wild-type sites settles to a stationary value ρ_s, as the creation and annihilation of mutant clusters (and boundaries) at the front equilibrate.

The relative frequency of events creating and annihilating sector boundaries determines the ultimate fate of the expanding front. In wild-type dominated regions of the front, mutation events enhance phenotypic heterogeneity and create new boundaries. At the same time, merging of mutant clusters creates homogeneous mutant regions. If the selective advantage of

wild-types is too small to trap mutant clusters, coalescence events promote the growth of mutant clusters, leading to more uniform front populations. Since we do not allow for back-mutations, the expanding front may end up in an *absorbing state* where it is completely taken over by mutants. For a finite system, this will eventually always happen even for deleterious mutations. The main question to ask then is: how does the corresponding *fixation time* scale with the system size L?

Inspecting figure 2 one can discern several morphological features of the expanding population, which we discuss phenomenologically now and analyze quantitatively in the following sections. We discriminate properties of the front and of the bulk of the population. An important bulk observable is the *density* of wild-type sites,

$$\rho(j) := \frac{1}{L}\sum_i \delta_{s_{i,j},1}, \tag{1}$$

at longitudinal position j. As indicated in figure 2(b), the ensemble averaged density decays, for $L \to \infty$, towards some stationary value ρ_s, which serves as an order parameter. A value of $\rho_s = 0$ corresponds to the absorbing state where all individuals at the front are mutants. Any finite value indicates phenotypic heterogeneity with both mutants and wild-types present at the front of the expanding population. Mutant clusters are, in general, anisotropic and one has to distinguish between their extension parallel and perpendicular to the preferred direction of the range expansion: the corresponding longitudinal and transverse *correlation lengths* are denoted by $\xi_\parallel$ and $\xi_\perp$, respectively. The front of the population is characterized by its average speed and its roughness. A good measure for the latter is the *width* w, defined as the standard deviation of the front's position,

$$w(L,t) := \left(\frac{1}{L}\sum_{i=1}^{L}\left[h(i,t) - \bar{h}(t)\right]^2\right)^{1/2}, \tag{2a}$$

from its average value,

$$\bar{h}(t) := \frac{1}{L}\sum_{i=1}^{L} h(i,t). \tag{2b}$$

Here $h(i,t)$ is the local position of the front, defined as the largest j for which $s_{i,j} \neq 0$.

3.2. Phase behavior: active and inactive phases

The morphology of the expanding population, as discussed above, depends on the values of the mutation probability p and the relative reproduction rate b of the mutant individuals; cf figure 3. One can clearly identify two distinct phases: in what we call the *active phase* the population front is composed of both wild-types and mutants in a heterogeneous mixture. Here mutants are continuously created by mutation events and the ensuing mutant sectors are subsequently lost again by a coalescence process where sector boundaries meet and the domain of mutants loses contact with the front. This typically happens in a parameter regime where mutations are deleterious and rare. In contrast, if mutations become more frequent and/or their reproduction rate becomes larger, mutant clusters have a significant probability to merge and/or to sweep through the system and thereby completely take over the population front. This is termed the *inactive phase* since the absorbing state, with only mutants at the front, cannot be left anymore.

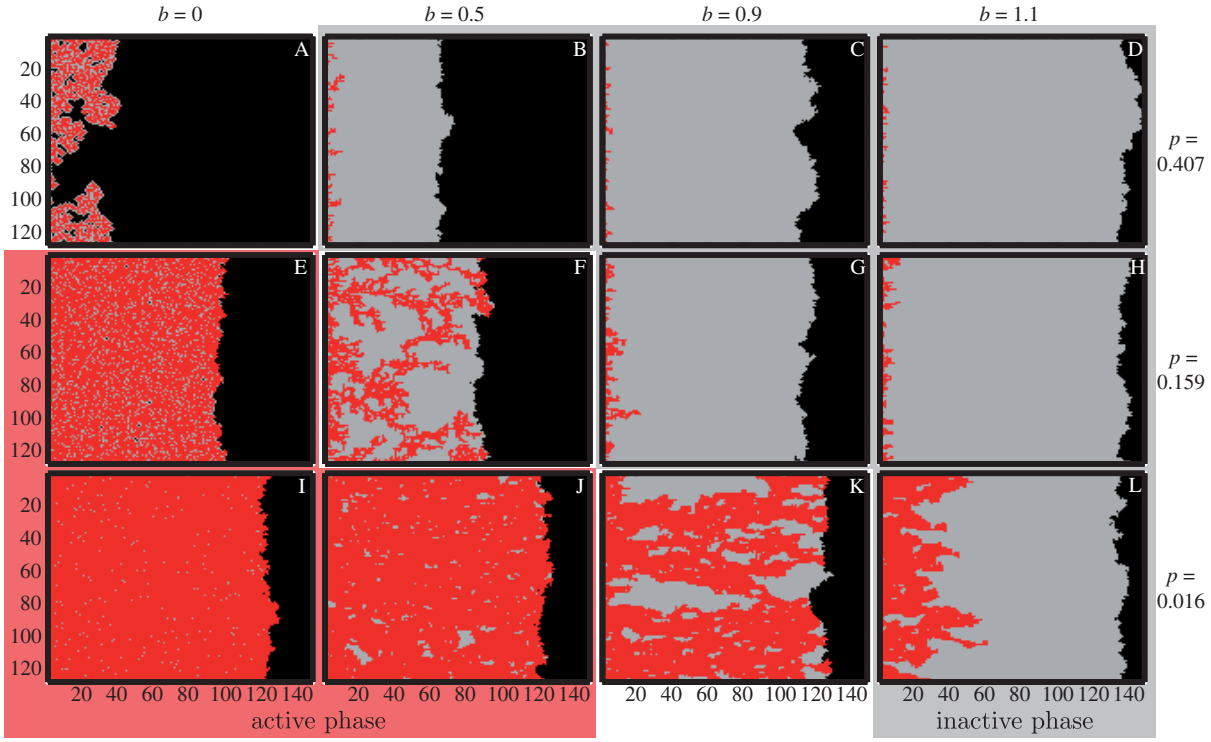


Figure 3. Parameter study of the Eden model with mutations. Typical realizations of the model in a finite system ($L = 128$) are shown for time $t = 100$ and various combinations of mutation rates p and relative birth rates b. Wild-types are shown in dark gray (red), mutants in light gray, and empty sites in black. For beneficial mutations (cf panels (D), (H) and (L)) mutant clusters expand quasi-deterministically. Together with panels (B), (C) and (G) these are examples for the inactive phase. Here the wild-type is lost in the bulk and the absorbing state, with only mutants at the front, is reached very fast, as mutants merge into extended clusters. As opposed to this, the active phase (cf panels (E), (I) and (J)) is characterized by long lasting heterogeneous fronts with both mutants and wild-types present. Here, the average spatial distance between distinct mutation events is larger than the extension of the created mutant clusters, which repeatedly lose contact to the front. A non-equilibrium phase transition separates the two phases, where low relative birth rates b and high mutation probabilities p balance and rich self-affine patterns evolve (cf panels (A), (F) and (K)). Going from panel (J) to panel (B) at constant birth rate $b = 0.5$, the system changes from a heterogeneous front of wild-types and mutants to a homogeneous all-mutant front. The transition is characterized by diverging length scales (mutant clusters), vanishing wild-type density and enhanced width of the population front, as discussed in the main text.

For all beneficial mutations, where $b > 1$, we are well in the inactive phase, independent of the probability p at which mutations appear. Here, a sector created from a single mutation has a finite opening angle and hence grows laterally on average, as observed in experiments [3, 22]. Therefore, already a single mutant can sweep through the population and take over

the population front, as the boundaries of the respective growth sector have, on average, a transverse velocity directed outwards [22, 32, 33]. This fixation process is further accelerated by the merging of neighboring mutant sectors; cf figure 3, panel (L).

For deleterious mutations, where $b < 1$, there is an antagonism between merging of different mutant clusters, creating large mutant domains, and closing of individual mutant clusters, leading to an enlarged fraction of wild-types at the front. Since mutations go from wild-type to mutant only, phenotypic heterogeneity is maintained as long as there are some wild-type individuals left at the front.

An isolated sector of mutants can survive only for a finite time interval due to stochastic effects. These enable mutants to 'surf' population waves in expanding populations [19, 65–67]: if mutations appear at the front of an expanding population they have a twofold advantage compared to mutations in spatially homogeneous settings. Firstly, the front can be seen as a perpetual population bottleneck, where demographic fluctuations are enhanced, which in turn reduces the effect of selection. Secondly, the offspring of mutations at the front can spread into unoccupied territory, where there is less competition. By this 'founder effect' mutants form clusters at the front that can reach much higher frequencies and evade extinction much longer than would be expected in spatially homogeneous settings. However, in the long run an isolated sector tends to lose contact with the front as it is outgrown by the wild-type as a result of the lower birth rate of mutant individuals; cf panels (E), (I) and (J) of figure 3. The phenomenology changes qualitatively when mutation events become more frequent. Then, nearby mutant sectors can merge and thereby counteract the loss of mutant sectors by the coalescence of sector boundaries [22]; cf panels (B), (C) and (G) of figure 3. The founder effect promotes the growth of individual mutations, but persistence of mutants is guaranteed by the merging of domains. As a consequence of this antagonism, one expects a phase boundary $p_c(b)$ between the passive and the active phase. Hallatschek and Nelson [22] considered mutations appearing at a flat front and found that for deleterious mutations there exists a critical mutation rate where mutants take over. We here consolidate this interesting finding and incorporate the roughness of the front. We analyze in detail the properties of the transition from a heterogeneous to an all mutant front in section 4.

3.3. Phase diagram

Since our model explicitly excludes back-mutations, the absorbing state, for which the wild-type strain has lost contact with the population front and only mutant individuals are present, cannot be left. For finite systems, $L < \infty$, this absorbing state is eventually always reached, as even for deleterious mutations mutant clusters of arbitrary size can appear through rare fluctuations. As noted above, the key quantity to analyze is the average time for this to happen as a function of mutation probability p, relative birth rate b and system size L. This mean fixation time $t_f(p, b, L)$ is defined as the mean time t when the number of wild-type individuals with empty neighbors, N_{wt}, becomes zero for the first time.

The mean fixation time generally diverges with growing system size, $L \to \infty$. The results from our stochastic simulations, shown in figure 4, allow us to distinguish three generic cases: (i) for large mutation probability p or beneficial mutations $b \geqslant 1$, we find that $t_f \sim \ln L$. This is the same result as for well-mixed populations [68, 69]. We take this asymptotic law as a hallmark for the inactive phase. Note that for isolated beneficial mutations (neglecting the merging of mutant clusters), one finds that the extinction time scales linearly in the system size [22]. Hence,

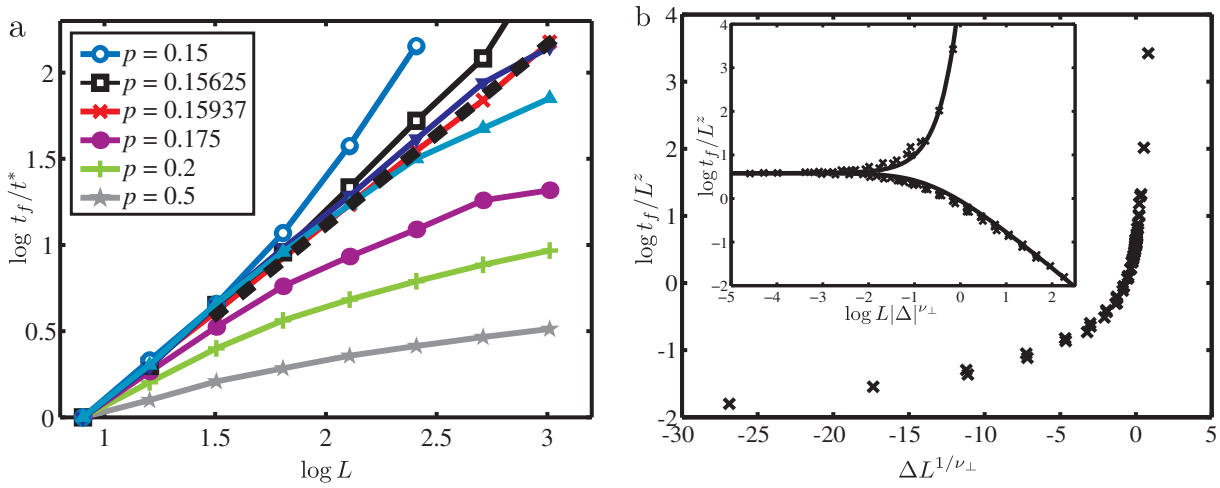


Figure 4. Fixation times t_f as a function of the system size L. (a) The fixation time t_f is the mean time until the absorbing state, where only mutants are at the expanding front, is reached. Asymptotically the fixation time grows with the system size L. Three cases, as discussed in the text and figure 3, can be distinguished: fast fixation, $t_f \sim \ln L$, exponentially slow fixation, $t_f \sim \exp cL$, and the marginal case $t_f \sim L^z$, where $z = 1.05 \pm 0.05$ (black dash line). Shown here are fixation times for $b = 0.5$, but for different birth rates the discrimination of three distinct regimes holds. For comparability, fixation times have been normalized by $t^* := t_f(L = 8)$. (b) The same data as in (a), but rescaled with system size L and distance to the phase transition Δ, according to finite size scaling as detailed in section 4.1. All data collapse onto a master curve as can be inferred from the semi-logarithmic plot. The inset is a double logarithmic plot of the same data, which depicts the characteristic scaling above and below criticality, $t_f = L^z \tau^{\pm}(L/\Delta^{-\nu_\perp})$. As a guide to the eye, we have included best fits (black lines) to these functions given by $\tau^+(x) = 3.8 \exp(10.5x)/(1+x)$ and $\tau^-(x) = 0.26 \ln(1+10.5x)/x$.

the logarithmic law arises from the merging events. (ii) In the active phase we have $t_f \sim \exp(cL)$, with some constant c. (iii) The phase boundary $p_c(b)$ between the active and the inactive phase is characterized by power law behavior of the mean extinction time: $t_f \sim L^z$ with a dynamical exponent z. This signature is well known from previous studies of phase transitions to absorbing states [36, 37] (see section 4). From our numerical data we estimate $z = 1.05 \pm 0.05$.

Since the phase transition from the active phase to the inactive phase is accompanied by a qualitative change in the L dependence of the mean fixation time t_f from logarithmic to exponential behavior, we may use it to map out the phase diagram; see figure 5.

We find a phase transition line $p_c(b)$, which ends at the points $(b, p) = (0, p^*)$ and $(b, p) = (1, 0)$. For $b = 0$ mutants do not reproduce at all and each mutant is created by a reproducing wild-type individual in an independent event with probability p. For low p the clusters have a typical size of one site and are obstacles around which the wild-types grow. For larger p more extended clusters are formed by mutation events that happen on neighboring sites. For the wild-types this corresponds to an Eden process on a *site percolation* network, where wild-type individuals are identified with occupied sites (present with probability $1 - p$)

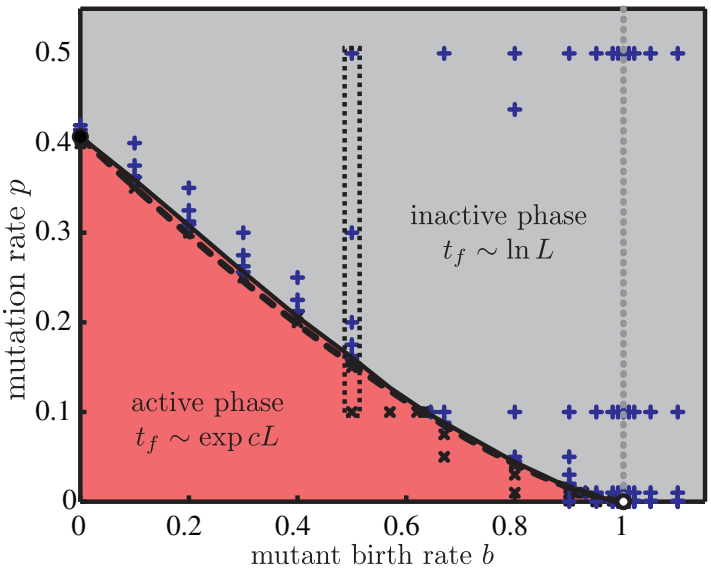


Figure 5. Phase diagram of the Eden process with mutations. The transition from the inactive phase (fast fixation, light gray) to the active phase (exponentially slow fixation, dark gray (red)) is a non-equilibrium phase transition (solid black line). Simulations can be categorized by the slope in a double-logarithmic t_F versus L plot for $L \gg 1$: plus signs (+ , blue) denote inactive phase behavior with slope <1; crosses ($\times$, black) denote active phase behavior with slope >1. Along the gray dotted line $b = 1$ mutations are neutral. The point ($p = 0$, $b = 1$) (black open circle) is not part of the transition line, as for $p = 0$ no mutants can appear and we recover the original Eden model. The phase transition line terminates on the axis $b = 0$, at critical mutation probability p^* (full black circle). This point corresponds to the critical density of isotropic site percolation; see the main text. The mean field transition line (black dashed line), given by equation (4), is in excellent agreement with our numerical results. To characterize the properties of the phase transition we explore how physical observables depend on the distance to the phase transition $\Delta := p_c - p$. To this end, simulations were carried out for $b = 0.5$ (without loss of generality), indicated by a black dotted box, for which we found $p_c = 0.159 \pm 0.001$.

and mutant individuals with empty sites (present with probability p). Thus $(b, p) = (0, p^*)$ is a multi-critical point of the transition line, below which the wild-type strain, for $L \to \infty$, keeps growing on the infinite percolating cluster. This implies that $p^* = 1 - p_c^{\text{isp}} \approx 0.407$, where p_c^{isp} is the percolation threshold of isotropic site percolation on a 2d square lattice [63].

For non-vanishing birth rates, $b > 0$, a newly formed mutant cluster can grow to some significant size. As a result clusters originating from more distant mutation events can merge before losing contact with the front. The phase transition line marks the set of mutation rates $p_c(b)$, where, for a given birth rate of mutants b, the typical spatial distance between two mutation events becomes comparable to the typical extension of individual mutant clusters. Thus, as we move to larger and larger birth rates, the critical mutation probability p_c declines.

For $b \to 1$, mutations become neutral, that is, the selection pressure vanishes and the random motion of sector boundaries is not biased anymore. Individual clusters can grow

unlimited by fluctuations, and consequently the smallest number of mutations suffices for the population to be in the inactive phase. If $p = 0$ no mutants are created and we recover the classical one-species Eden model composed of growing wild-types [41]. Thus, at $(b, p) = (1, 0)$ the front is composed of only wild-type individuals and the absorbing state is never reached. Our numerical data suggest that p_c approaches this point in a cusp-like singularity.

For $b > 1$ mutations are beneficial, which means mutations and selection are not antagonists anymore, and hence a phase transition is absent. Here, the mutants take over the system quasi-deterministically for all finite values of p.

To rationalize our findings for the phase diagram we consider a phenomenological mean field theory for the dynamics of the density of the wild-type strain at the front of the expanding population, $n = N_{\mathrm{wt}}/(N_{\mathrm{wt}} + N_{\mathrm{mut}})$. On average, wild-type individuals are lost by mutation at a rate p, and gained or lost through natural selection depending on the relative growth rates of the wild-type and mutant individuals. The effect of natural selection is proportional to $n(1-n)$, which vanishes if the front is composed of one strain only. This leads to the following rate equation:

$$\dot{n} = -pn + s(b)n(1-n), \tag{3}$$

where $s(b)$ is an effective selection strength. The functional form of $s(b)$ can be determined phenomenologically: the rate equation (3) has the fixed points $n_1 = 0$ and $n_2 = 1 - p/s(b)$. In the active phase, $n_1 = 0$ is unstable, while n_2 is stable. At the phase transition the two fixed points merge and interchange their stability in a transcritical bifurcation. Solving $n_1 = n_2$ gives the mean field transition line $p_c^{\mathrm{mf}}(b) = s(b)$. For $b \to 0$ the problem reduces to isotropic percolation and hence $s(0) = p^*$. Certainly the selection coefficient $s(b)$ has to vanish if mutants reproduce as fast as wild-types, $s(1) = 0$, which gives us the other end of the transition line. Close to it the transition line can be approximated by a power law, $p_c^{\mathrm{mf}}(b \to 1) \sim (1-b)^{\mu}$, reflecting the cusp singularity we found in the simulations. The simplest ansatz for the mean field transition line, which fulfills these constraints, is

$$p_c^{\mathrm{mf}}(b) = s(b) = p^*(1-b)^{\mu}. \tag{4}$$

A fit to our numerical derived phase transition line gives $\mu = 1.41 \pm 0.03$. The phenomenological mean-field transition line, depicted in figure 5, is in very good agreement with our numerical results.

4. Critical behavior

In this section, we investigate in depth the properties of the absorbing state phase transition from the active into the inactive phase. Without loss of generality we focus on a fixed value for the birth rate of mutants, $b = 0.5$, for which we found the critical mutation probability to be $p_c(b = 0.5) = 0.159 \pm 0.001$. The qualitative behavior of all observables stays the same along the transition line. Close to the special points $b = 0$ and $b = 1$ crossover effects become more pronounced, which we do not examine here.

4.1. Bulk properties

We first address bulk properties of the system, i.e. observables measured sufficiently far away from the rough growth front. Since we are dealing with a phase transition to an absorbing state,

we expect to find four independent critical exponents [34–37]. A common choice of observables is: the stationary density $\rho_s := \rho(j \to \infty)$ of active sites (wild-types); the survival probability P_s (survival, in this context, means the wild-type having contact with the growth front) of a single active site (wild-type) in a front of inactive sites (mutants); the longitudinal correlation length $\xi_\parallel$, and the transverse correlation length $\xi_\perp$. For $L \to \infty$, all these observables diverge like power laws in the control parameter $\Delta := p_c - p$ in the vicinity of the phase transition. The respective critical exponents are defined through

$$\rho_s \sim \Delta^\beta \quad \text{for } \Delta \geqslant 0, \tag{5a}$$

$$P_s \sim \Delta^{\beta'} \quad \text{for } \Delta \geqslant 0, \tag{5b}$$

$$\xi_\parallel \sim |\Delta|^{-\nu_\parallel}, \tag{5c}$$

$$\xi_\perp \sim |\Delta|^{-\nu_\perp}. \tag{5d}$$

The above observables are understood as ensemble averages. The stationary density ρ_s and the survival probability P_s are 0 in the inactive phase, $\Delta < 0$, since in this case any heterogeneous composition of the front is unstable.

For finite systems, $L < \infty$, both the stationary density ρ_s and the survival probability P_s are identical to 0 for the inactive *and* the active phase. Although it may be extremely rare, it is always possible—even in the active phase—that large enough mutant clusters appear and finally lead to a front consisting of mutants only, which is the absorbing state. Hence, ρ_s and P_s are not particularly useful observables for $L < \infty$, and one instead has to examine the time-dependent density $\rho(\Delta, t, L)$ and survival probability $P(\Delta, t, L)$, and also time-dependent correlation lengths $\xi_\parallel(\Delta, t, L)$ and $\xi_\perp(\Delta, t, L)$.

We measured the critical exponents, defined in equations (5*a*)–(5*d*), using different methods. Starting with a line of wild-types in the active phase we let the system evolve, disregarding realizations where the absorbing state was reached. This allows us to use stationary observables, instead of more involved dynamical ones. From each realization we extracted for each mutant cluster (i) the longitudinal extension $\ell_\parallel$, (ii) the transverse extension $\ell_\perp$ and (iii) the number of cluster sites or mass m. In analogy to percolation theory [63], the correlation lengths are then calculated by

$$\xi_\#^2(\Delta, L) = \frac{2\sum_k m_k \ell_{\#,k}^2}{\sum_k m_k}, \quad \# \in \{\parallel, \perp\}, \tag{6}$$

where k runs over all observed clusters. The computed correlation lengths depend on both the distance to the phase transition Δ and the system size L. However, close to the critical point all macroscopic observables are invariant under scaling transformations of the form [34–37]

$$\Delta \to c\Delta, \tag{7a}$$

$$x_\parallel \to c^{-\nu_\parallel} x_\parallel, \tag{7b}$$

$$x_\perp \to c^{-\nu_\perp} x_\perp, \tag{7c}$$

$$\rho \to c^\beta \rho, \tag{7d}$$

$$P \to c^{\beta'} P. \tag{7e}$$

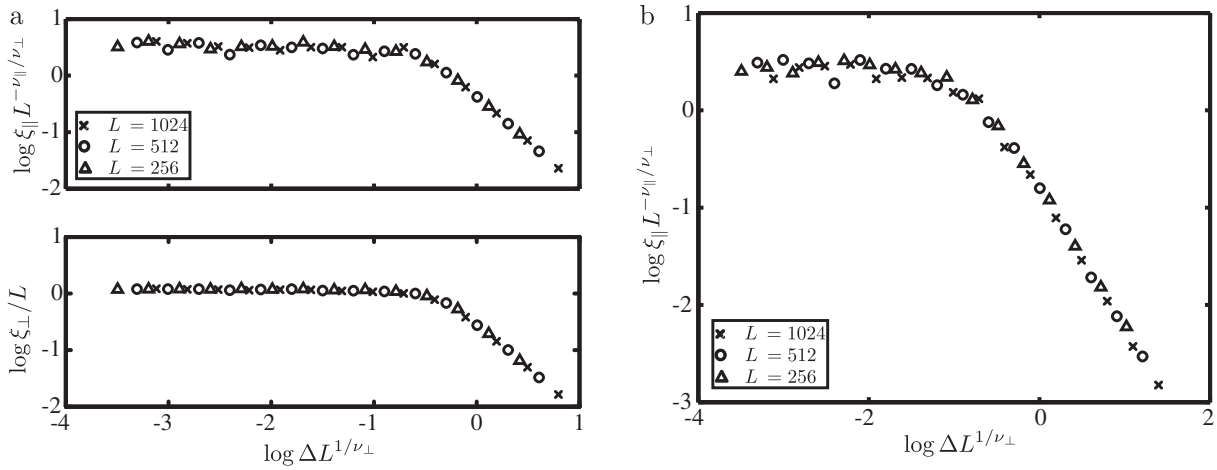


Figure 6. Rescaled correlation lengths from simulations started with a line of wild-types as the initial condition for different lattice sizes. (a) Rescaled correlation lengths from cluster-size distributions in the active phase. Longitudinal correlation length $\xi_\parallel$ (upper plot) and transverse correlation length $\xi_\perp$ (lower plot) have been computed from mutant cluster masses m and cluster extensions $\ell_\parallel$ and $\ell_\perp$ (see text). Correlation lengths depend on both L and distance to the phase transition Δ. Taking $\nu_\perp = \nu_\parallel = 1.6 \pm 0.1$ all data collapse to master curves under rescaling. (b) Rescaled longitudinal correlation length as found from decay of wild-type density in the inactive phase. The wild-type density ρ decays exponentially, with a decay length proportional to the longitudinal correlation length. Again all data collapse to a master curve for $\nu_\perp = \nu_\parallel = 1.6 \pm 0.1$.

where c is some positive rescaling factor. This implies for the correlation lengths the following scaling forms:

$$\xi_\#(\Delta, L) = L^{\nu_\#/\nu_\perp} f_\#(\Delta L^{1/\nu_\perp}), \tag{8}$$

where the $f_\#(x)$ are universal scaling functions with the asymptotic behavior

$$f_\#(x) \sim \begin{cases} 1, & x \ll 1, \\ x^{-\nu_\#}, & x \gg 1. \end{cases} \tag{9}$$

Indeed, after rescaling, our simulation data for both correlation lengths $\xi_\#(\Delta, L)$ collapse onto master curves for all L and Δ, if we take

$$\nu_\perp = \nu_\parallel = 1.6 \pm 0.1, \tag{10}$$

see figure 6(a). Note that this means that close to the transition the longitudinal and transverse correlation lengths show the same scaling behavior, $\xi_\perp \sim \xi_\parallel$, at least within the accuracy of our simulations. Simply put, a fluctuation of twice the size takes twice as long to decay, which explains that close to the phase transition fixation times are proportional to the system size, $t_\mathrm{f} \sim L$, which we found in section 3.3.

With the critical exponents $\nu_\perp$ and $\nu_\parallel$ at hand, rescaling of the fixation time $t_\mathrm{f}(L, \Delta)$ is easily achieved as well. Note that as a time-like observable t_f scales like a longitudinal distance.

Again employing the phenomenological scaling theory for absorbing state phase transitions, equations (7*a*)–(7*e*), we obtain

$$t_{\mathrm{f}}(L,\Delta) = L^{z}\tau^{\pm}(\Delta L^{1/\nu_{\perp}}), \tag{11}$$

where $z = \nu_{\parallel}/\nu_{\perp}$ is the dynamical critical exponent. The overall scaling function can be split into two distinct parts, τ^{+} and τ^{-}, which exhibit characteristic logarithmic and exponential behavior for $\Delta > 0$ and $\Delta < 0$, respectively. Using this scaling, all data for the fixation times collapse, as shown in figure 4(b).

In the inactive phase, the density $\rho(\Delta, j, L)$ decays exponentially with j and the absorbing state, with $\rho = 0$, is reached fast. The decay length is proportional to $\xi_{\parallel}$, which can easily be found by fitting an exponential function to the measured density profile. Applying finite size scaling, we again find data collapse for critical exponents $\nu_{\perp} = \nu_{\parallel} = 1.6 \pm 0.1$; see figure 6(b).

Further evidence for these values of the critical exponents is obtained by calculating the (active state) correlation functions

$$\Gamma_{\parallel}(r) = \langle s_{i,j}s_{i,j+r}\rangle - \langle s_{i,j}\rangle\langle s_{i,j+r}\rangle, \tag{12a}$$

$$\Gamma_{\perp}(r) = \langle s_{i,j}s_{i+r,j}\rangle - \langle s_{i,j}\rangle\langle s_{i+r,j}\rangle, \tag{12b}$$

in the bulk. Averages are with respect to all lattice indices i, j and independent realizations. The correlation functions are expected to behave as

$$\Gamma_{\#}(r) \sim r^{-\sigma_{\#}} \exp\left(-\frac{r}{\xi_{\#}}\right), \quad \# \in \{\parallel, \perp\}, \tag{13}$$

where $\sigma_{\#} = 2\beta/\nu_{\#}$. To obtain the correlation lengths we fitted this expression to our data with the fitting parameters $\xi_{\#}$. In plots of the correlation lengths, we again find good collapse of data for $\nu_{\perp} = \nu_{\parallel} = 1.6 \pm 0.2$ (not shown).

Next, we consider the stationary density $\rho_{\mathrm{s}}(\Delta)$, which equals the fraction of wild-type sites for large j, which is sufficiently far away from the initial line for the density to relax to its stationary value. As shown in figure 7(a), a double-logarithmic plot asymptotically gives power law behavior with the exponent

$$\beta = 0.50 \pm 0.02. \tag{14}$$

Deviations from the asymptotic power law are well described by the finite-size scaling form

$$\rho_{\mathrm{s}}(\Delta, L) = L^{-\beta/\nu_{\perp}} g(\Delta L^{1/\nu_{\perp}}), \tag{15}$$

where $g(x)$ is a universal scaling function; cf figure 7(b). Two characteristic regimes can be distinguished:

$$g(x) \sim \begin{cases} 1, & x \ll 1, \\ x^{\beta}, & x \gg 1, \end{cases} \tag{16}$$

corresponding to the scaling law $\rho_{\mathrm{s}} \sim \Delta^{\beta}$, for a given L and sufficiently far from criticality, and the scaling law, $\rho_{\mathrm{s}} \sim L^{-\beta/\nu_{\perp}}$, for a given distance Δ from the critical point and system sizes smaller than the transverse correlation length, $L \ll \xi_{\perp}$.

To measure β', the exponent associated with the survival probability when starting from a single seed, P_{s}, different initial conditions must be used. Instead of a line composed of only wild-types, a single wild-type is placed in a line of mutants. A typical realization for these initial conditions is shown in figure 8(a).

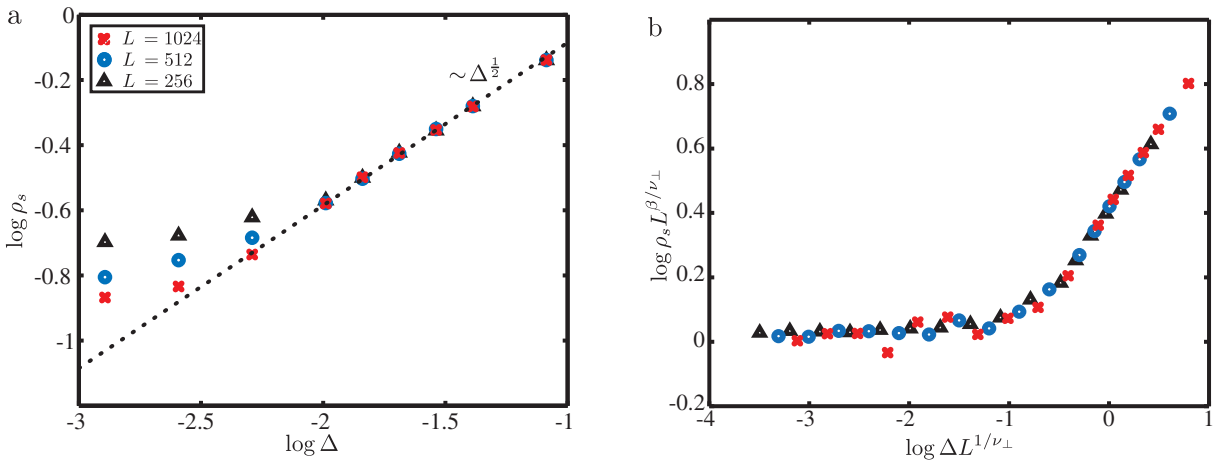


Figure 7. Wild-type stationary density in the active phase. (a) Simulations for different lattice sizes started with a line of wild-types as the initial condition in the active phase. We find a power law with critical exponent $\beta = 0.50 \pm 0.02$ that, for small Δ, holds better in larger systems, a typical finite size effect. (b) The same data as in (a) after finite size scaling. We find collapse of all data for critical exponents $\beta = 0.5$ and $\nu_\perp = 1.6$.

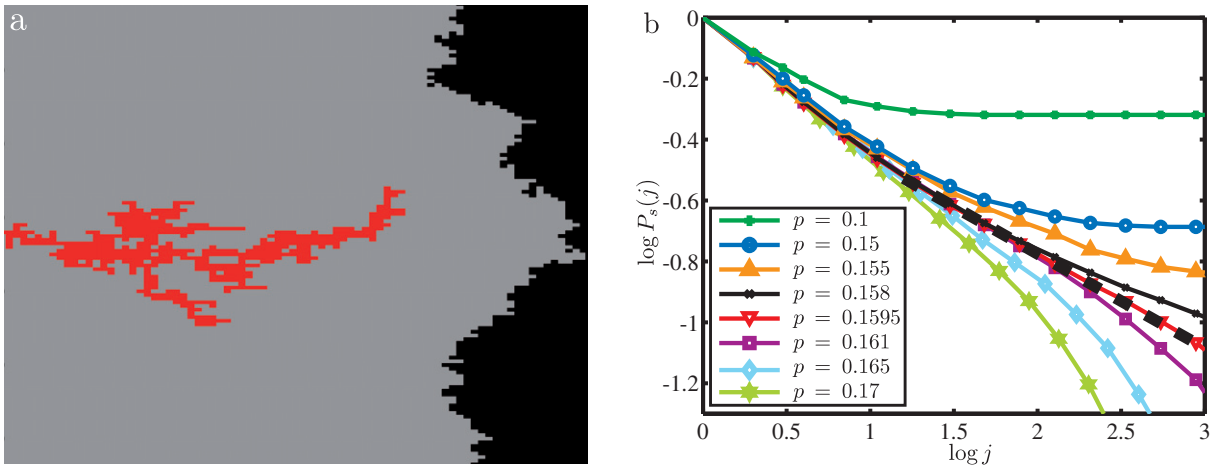


Figure 8. Single wild-type initial condition and related observables. (a) Typical realization for single-seed initial conditions. Wild-types are shown in dark gray (red), mutants in light gray. By averaging over many realizations of the process the survival probability of wild-types $P(j)$ and the number of wild-types $N(j)$ can be measured as a function of longitudinal position j. For this realization on a lattice of length $L = 128$ we used birth rate $b = 0.5$ and mutation probability $p = 0.15$. (b) Survival probability as a function of time. Simulations for different mutation rates p started with a single wild-type site as initial condition. For $p_c \approx 0.159$, we find a power law for $P_s(j)$ with critical exponent $\beta'/\nu_\parallel = -0.32 \pm 0.2$, indicated by the dashed line. Birth rate $b = 0.5$ and lattice size $L = 1024$ are the same for all shown data.

Table 1. Critical exponents for the non-equilibrium phase transition to an absorbing state of the two-species Eden model with mutations and selection.

β	β'	$\nu_\parallel$	$\nu_\perp$	z	θ
0.50 ± 0.02	0.51 ± 0.07	1.6 ± 0.1	1.6 ± 0.1	1.05 ± 0.05	0.32 ± 0.02

From an ensemble of realizations the survival probability $P_s(j)$ and the mean number of wild-types $N(j)$, both as functions of distance j to the initial line, can be extracted. After transients have decayed and sufficiently far away from the front (where there might exist empty lattice sites), we find power laws for both quantities when close to the phase transition. For P_s the exponent is given by $-\beta'/\nu_\parallel$, while for N it is denoted by θ. Figure 8(b) shows the results for $P_s(j)$, from where we obtain $\beta'/\nu_\parallel = -0.32 \pm 0.2$. Using our result for $\nu_\parallel$, we find

$$\beta' = 0.51 \pm 0.07. \tag{17}$$

Similarly, from the simulation data for $N(j)$ we obtain

$$\theta = 0.32 \pm 0.02. \tag{18}$$

θ is related to the other critical exponents by the generalized hyperscaling relation [70] in $d = 1$ dimension

$$\theta = \frac{\nu_\perp - \beta - \beta'}{\nu_\parallel}. \tag{19}$$

Using our previous results for the critical exponents we find good agreement with the above result, equation (18), within the error margin. The critical exponents for the phase transition to an absorbing state are summarized in table 1.

For systems with infinitely many absorbing states, like ours, it is known that critical exponents, especially those related to single seed initial conditions (in our case β' and θ), can vary significantly with the configuration away from the seed [71, 72]. The correct exponents, which satisfy the hyperscaling relation of equation (19), are obtained if the configuration away from the seed is a typical configuration of the absorbing state. In our case this amounts to a rough front. To check for this dependence of the critical exponents, we initialized single wild-type seeds at arbitrary front positions of all-mutant Eden models grown to saturation. Data extraction works similar, but due to roughness and the non-uniqueness of the growth direction the initial seed is not necessarily the one with the smallest j-value, which complicates the evaluation slightly. These simulations take more time since independent saturated fronts have to be generated for every realization, which makes it harder to obtain data with high precision. However, we find $\beta'/\nu_\parallel = -0.33 \pm 0.03$ and $\theta = 0.31 \pm 0.04$ (data not shown), which indicates that our system is robust with respect to the initial configuration. We are therefore confident that the critical exponents of the phase transition, see table 1, are indeed correct up to the given precision.

To check if the DP universality class is recovered if roughness is neglected, we also considered a simplified model with synchronous update rules. Here, an individual at site (i, j) has two possible parents, at sites $(i-1, j-1)$ and $(i, j-1)$. Depending on their states, we define the probabilities $P(s_{i,j}|s_{i-1,j-1}, s_{i,j-1})$ for the type of offspring $s_{i,j}$, incorporating mutations and selection: if both parents are mutants, the offspring is inevitable mutant as well; if both parents are wild-types the offspring is wild-type if no mutation happens; if one parent is

wild-type while the other is a mutant, the mutant reproduces with rate $b/(1+b)$ and the wild-type otherwise. In the latter case a mutation can happen, resulting in a mutant offspring. All events are summarized by the probabilities

$$P(1|-1,-1) = 0, \tag{20a}$$

$$P(1|1,-1) = P(1|-1,1) := p_1 = \frac{1-p}{1+b}, \tag{20b}$$

$$P(1|1,1) := p_2 = 1-p, \tag{20c}$$

and $P(-1|.,.) = 1 - P(1|.,.)$. It is easily seen that this corresponds to the Domany–Kinzel cellular automaton [73], with probabilities p_1 and p_2. Indeed, for $b = 0.5$ we find the critical mutation probability for the flat model $p_c^{\text{flat}} \approx 0.077$, corresponding to $p_1 \approx 0.62$ and $p_2 \approx 0.93$. This lies well on the transition line of the Domany–Kinzel automaton [74]. As one would expect, the critical exponents of the flat model are those of the DP universality class (data not shown). These conclusions are further consolidated by earlier studies for a two-species ballistic deposition model with kinetics that allows surface growth only at sites where one type of particles has an exposed position on top of the incidence or neighboring column [60]. These particular growth rules imply that the surface roughness does not affect the bulk dynamics such that it can be mapped onto a one-dimensional contact process [61], which is in the DP universality class [34–37]. In conclusion, the interplay between surface roughening and domain dynamics in our two-species Eden model is clearly responsible for the deviation of its critical behavior from that of the DP universality class.

4.2. Front properties

We now turn to characterizing the front of the expanding population, i.e. the surface roughness properties of the model. As argued in section 3.1, the presence of mutations changes the roughening behavior of the growth front qualitatively: close to the phase transition large mutant clusters form extended parts of the leading front and since mutants and wild-types reproduce with different rates, the front's roughness is strongly enhanced. In the case of detrimental mutations, $b < 1$, mutant-dominated regions trail behind compared to the average front; see, for example, figure 3, panel (F).

The width is the key observable in the analysis of kinetic surface roughening processes. These processes can be organized into universality classes, which are characterized by symmetry properties and the dimensionality of the process under consideration [43–45]. Each growth model's universality class has a unique set of two exponents: the growth exponent γ and the roughness exponent α, defined through

$$w(L,t) \sim \begin{cases} t^{\gamma}, & t \ll t_{\times}, \\ L^{\alpha}, & t \gg t_{\times}, \end{cases} \tag{21}$$

with the crossover time between these two regimes scaling as $t_{\times} \sim L^{\tilde{z}}$, where $\tilde{z} = \alpha/\gamma$ is the dynamic roughening exponent.

In the limit of the vanishing mutation rate, $p = 0$, our model reduces to the Eden model, which falls into the well-known KPZ universality class [47], with exponents $\gamma = 1/3$ and $\alpha = 1/2$. For $p > 0$, the time evolution of the width depends on whether the system is in the active phase or in the inactive phase. In the inactive phase we find a transiently enhanced width as compared to the Eden model and a subsequent return to it. We attribute this to mutants

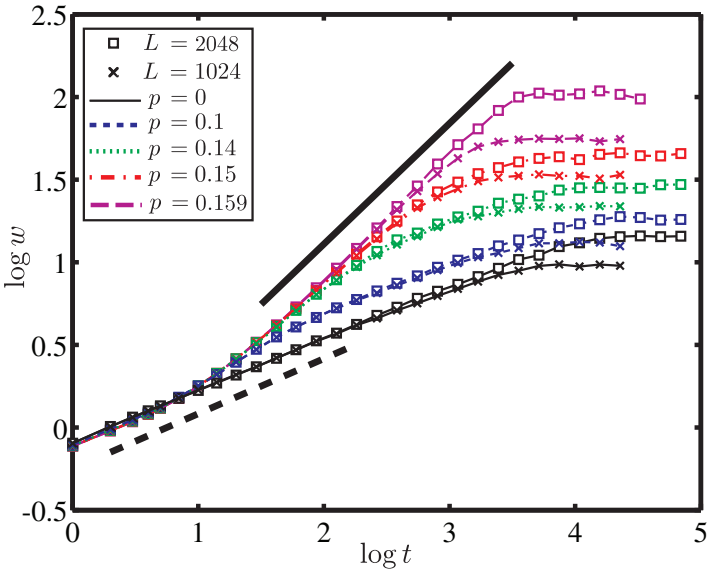


Figure 9. Width of the growth front as a function of time in the active phase. The temporal evolution of the width $w(t)$ of the growth front depends on both the system size L and the distance to the phase transition Δ. w initially grows Eden-like, $\gamma = 1/3$ (bold dashed line). Close to the phase transition (i.e. at small Δ), we find a crossover to a growth exponent $\gamma' = 0.73 \pm 0.03$ (bold straight line), as a result of different growth rates of mutants and wild-types. For intermediate times two scenarios can be distinguished, see the main text for details: (i) either the surface width saturates at a finite length-dependent value as a result of finite system size or (ii) mutant clusters at the front saturate at a size smaller than L. In the latter case, Eden growth is recovered on large length and time scales. As the width saturates we recover the Eden roughness exponent $\alpha = 1/2$ but with a p-dependent amplitude of the saturation width. In the former case, enhanced roughening proceeds until saturation. The saturated width then scales like $L^{\alpha'}$, with a different roughness exponent $\alpha' = 0.91 \pm 0.01$. All data shown are for $b = 0.5$, but results hold for different birth rates b as well.

outcompeting wild-types rapidly, as is typical for the inactive phase, followed by Eden growth of the resulting mutant front (data not shown). As shortly discussed in section 3, the active phase is characterized by a heterogeneous front. In finite systems, this is the case until a mutant cluster covers the whole system size L, which then takes the system into the absorbing state with subsequent return to Eden-like growth. Therefore, the roughness characteristics of the heterogeneous front in the active phase can only be analyzed if the absorbing state has not yet been entered. Hence, for the following analysis, realizations that reach the absorbing state are stopped, and simulation data are used only up to this point in time.

We now discuss the different regimes and scenarios of surface roughening in the active phase. Figure 9 displays our numerical results for a range of mutation rates and two different system sizes. Since we start with an initial condition where the lattice only contains wild-types, the initial surface roughening is that of a single species, i.e. Eden growth: $w \sim t^{\gamma}$ with $\gamma = 1/3$. As time passes, mutants are created, leading to growing mutant clusters. Their

extension in growth direction evolves like the correlation length $\xi_{\parallel}(\Delta, t, L)$, which initially grows proportional to t, as follows from scaling. Since these clusters grow slower than wild-type clusters this leads to additional surface roughening. As for small t both contributions to roughening are independent of Δ and L, the width gets dominated by differential expansion velocities at a crossover time $t_{\times,1} = \mathcal{O}(1)$.

As a result of the ensuing *strong coupling* between domain dynamics and surface roughness, the interface width then grows more rapidly than in the Eden model. We observe a regime with an altered growth exponent $\gamma' = 0.73 \pm 0.03$, which becomes more extended upon approaching the phase transition to the inactive phase, $\Delta \to 0$. There are two distinct scenarios by which this enhanced roughening regime may end: (i) either the surface width saturates at a finite length-dependent value as a result of finite system size or (ii) mutant clusters at the front saturate and their typical size reaches the asymptotic extension $\xi_{\perp}^{\infty} \sim \Delta^{-\nu_{\perp}} \leqslant L$.[2]

The first case requires that one is sufficiently close to the critical point such that the cluster size $\xi_{\perp}$ does not saturate at $\Delta^{-\nu_{\perp}} \leqslant L$, but mutant clusters continue to grow during the whole roughening process. Then, our simulations show that there is a direct crossover from enhanced surface roughening, $w \sim t^{\gamma'}$, to saturation, $w \sim L^{\alpha'}$, with a roughness exponent $\alpha' = 0.91 \pm 0.01$. The corresponding scaling form close to criticality reads

$$w_{\mathrm{c}}(L, t) = L^{\gamma'} \hat{w}_{\mathrm{c}}(t/L^{\tilde{z}'}) \tag{22}$$

with the dynamic critical exponent for surface roughening $\tilde{z}' = \alpha'/\gamma' = 1.25 \pm 0.05$; cf figure 10(a). (Note that deviations from the critical scaling behavior are all due to initial transient Eden growth: all systems were started with an initial condition of wild-type individuals only.) A roughening exponent α' close to unity indicates a very jagged surface. For models with large roughness exponents, especially for super-roughening with exponents larger than 1 (see, e.g., [75, 76]), it is known that they may exhibit anomalous scaling [77, 78], in the sense that the roughness exponent as determined from $w(L, t \to \infty)$ may only be an effective exponent. The actual exponent can be obtained from the structure factor or power spectrum [45, 79, 80]

$$S(k, t) := \langle h(k, t) h(-k, t) \rangle, \tag{23}$$

where $h(k, t)$ is the Fourier transform of the height fluctuations. From equation (21) it follows that the structure factor of the Eden model scales as

$$S(k, t) = t^{(d+2\alpha)/\tilde{z}} s(k t^{1/\tilde{z}}), \tag{24}$$

with the scaling function

$$s(x) \sim \begin{cases} 1, & x \ll 1, \\ x^{-d-2\alpha}, & x \gg 1. \end{cases} \tag{25}$$

The rescaled structure factor of our model is shown by the inset of figure 10(a). For the trivial case $p = 0$, we observe data collapse as expected for the Eden model. At criticality $\Delta = 0$ ($p = 0.159$), the scaling function for the power spectrum changes with time due to the growth dynamics of mutant clusters, i.e. the time dependence of the cluster size $\xi_{\perp}(t)$. For asymptotically large times, when a single mutant cluster spans the system, a broad regime emerges with a roughness exponent $\alpha' = 0.92 \pm 0.03$ corroborating our results obtained from analyzing the interface width.

[2] A third scenario where the typical lateral extension of mutant clusters reaches the extension of the system $\xi_{\perp} \sim L$ is also possible but excluded by our sampling method.

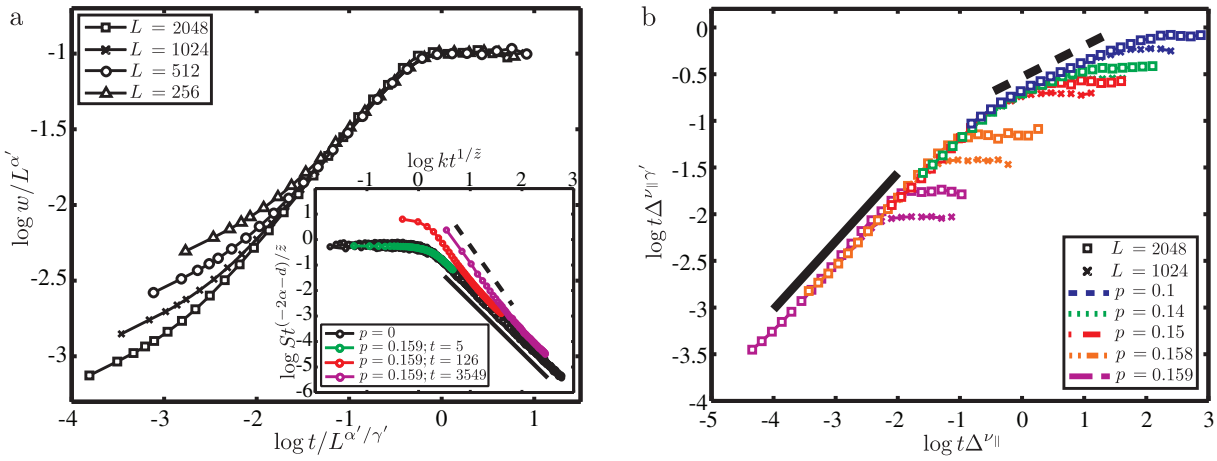


Figure 10. Scaling plots for surface roughening in the Eden model with mutation. (a) *Critical scaling*. Rescaled width versus time at criticality ($p = 0.159$) for a series of system sizes as indicated in the graph. Neglecting initial transients, all data collapse for $\alpha' = 0.91 \pm 0.01$ and $\tilde{z}' = \alpha'/\gamma' = 1.25 \pm 0.05$. Both growth exponent γ' and roughening exponent α' are markedly different from their KPZ counterparts. The inset shows the rescaled structure factor $S(k, t)$ for system size $L = 512$. For the Eden model all data collapse, and from the black solid line, proportional to k^{-2}, we recover $\alpha = 1/2$ from equation (25). For the critical system, $\Delta = 0$, we find an extended scaling regime with $S \sim k^{-2.85 \pm 0.05}$ (black dashed line) corresponding to $\alpha' = 0.92 \pm 0.03$. (b) *Crossover scaling*. Neglecting initial transients, data can be rescaled with respect to the intrinsic crossover time $t_{\times,2} \sim \Delta^{-\nu_\parallel}$, where the mutant cluster extension $\xi_\perp$ saturates at $\Delta^{-\nu_\perp}$. If the system size L is much smaller than $\xi_\perp^\infty$, the front width grows to saturation with the new growth exponent $\gamma' = 0.73 \pm 0.03$ (black solid line), whereas for $\xi_\perp^\infty \ll L$ the saturation width is reached after return to Eden-like roughening with $\gamma = 1/3$ (black dashed line). Saturation, due to finite size, happens where curves deviate from the common envelope and become horizontal.

This clearly shows that, for $\Delta \to 0$, the Eden model with mutations and selection is no longer in the KPZ universality class, but exhibits different asymptotic roughening behavior. We attribute this behavior to the strong coupling between critical domain growth and surface roughening.

Moving further away from the critical point, typical domains may saturate at their asymptotic value $\Delta^{-\nu_\perp} \leqslant L$ within a crossover time $t_{\times,2} \sim \Delta^{-\nu_\parallel}$. Then, on length scales larger than $\xi_\perp^\infty$, roughening proceeds, but slower than before, and one recovers Eden-like roughening with $w \sim t^\gamma$. This regime is most pronounced if $L \gg \xi_\perp^\infty$, and finally saturates in a constant surface width with $w \sim L^\alpha$. The crossover between critical roughening and return to Eden roughening is determined by a time scale inherent to the system, which only depends on the distance to the critical point but not on the system size. We may therefore rescale time by the crossover time, $t_{\times,2} = \Delta^{-\nu_\parallel}$, and rescale width w by the corresponding value $w(t_{\times,2}) = t_{\times,2}^{\gamma'} = \Delta^{-\gamma'\nu_\parallel}$, cf figure 10(b). All curves collapse to a crossover scaling function, from which they only deviate due to finite size effects.

5. Conclusion and outlook

A generalization of the Eden model to a two-species model, with uni-directional mutations and selection due to different reproduction rates, has interesting properties from the viewpoints of both dynamical phase transitions and kinetic surface roughening. We find that surface roughness has a marked effect on the critical properties of the absorbing state phase transition. While reference models [22], which keep the expanding front flat, exhibit DP critical behavior, the exponents of our generalized Eden model strongly deviate from it. In turn, the mutation-selection process in the bacterial colony induces an increased surface roughness with exponents distinct from that of the KPZ universality class.

For our model the critical exponents of the longitudinal and transverse correlation lengths, $\nu_\parallel$ and $\nu_\perp$, are identical within the error margin. As the aspect ratio $\xi_\parallel/\xi_\perp$ can still be different from unity, this does not make our system isotropic. However, it indicates that there is no longer a rigorous distinction between a time-like longitudinal and a space-like transverse direction as for DP. To some degree, this might have been expected from the microscopic update rules of the Eden model, which have no preferential direction of growth. Indeed, for vanishing birth rate of mutants, the system can be mapped to *isotropic* percolation, and it loses any preferential growth direction. One may also argue that front roughening in concert with locally normal growth of the interface leads to 'mixing' of transversal and longitudinal directions, and thereby promotes the same scaling behavior of both directions. It remains to be seen how this result can be obtained from a renormalization group calculation.

The critical exponents summarized in table 1 are significantly different from those of the DP universality class, which are observed for a broad range of systems exhibiting phase transitions to absorbing states [34–37]. The stability of the DP universality class is reflected in the DP hypothesis, a conjecture formulated by Janssen [81] and Grassberger [82]. It asserts that models characterized by a positive one-component order parameter, which exhibit a phase transition to a unique absorbing state, belong to the DP universality class, provided that interactions are short-ranged and no additional symmetries or special kinds of disorder or fluctuations are present [34–37]. Strictly speaking, our model is not in contradiction with the DP hypothesis as it does not fulfill the condition of having a *unique* absorbing state. Instead, there are infinitely many configurations with only mutants at the front. However, even for models with infinitely many absorbing states one often finds DP-class universal behavior [71, 72]. What distinguishes our model is that growth persists in the fluctuating absorbing state of an all mutant system. Most importantly, there is interplay between surface roughness and bulk properties: even though we analyzed bulk properties of the system, the dynamics are restricted to the front and therefore convey imprints of the rough surface. This might also be the reason why it is difficult to find actual experimental systems that obey DP-like behavior: any non-flat system may carry signatures of surface roughening in the numerical values of the critical exponents of the corresponding non-equilibrium phase transition.

The influence of surface roughness on domain boundaries, without the additional complication of mutations, selection and of merging clusters, is a rather complex problem on its own. The morphology of the boundary between two distinct but equally fast growing Eden clusters has previously been analyzed by Derrida and Dickman [48]. Numerically, they found that the dynamics of the domain boundary depends on the local curvature of the front. It may perform a subdiffusive or superdiffusive random walk and even move ballistically. For our model, this implies enhanced complexity, because boundaries are preferentially created

at protrusions, where wild-types are more prone to be, while merging of domain boundaries usually occurs at indentations of the front. Coarsening of domains in a two-species Eden model is the focus of the work by Saito and Müller-Krumbhaar [32], who also considered differential reproduction rates of the two species. As mutations are absent in their model, the faster growing species outgrows the slower one exponentially fast. In the neutral case, they show that domain boundaries exhibit super-diffusive scaling and thereby explain the domain coarsening kinetics of their simulations. It would be interesting to find a suitable generalization of their analysis to include the case of irreversible mutations.

There are also growth models where surface roughness and particle configuration at the front mutually affect each other [49, 50, 56–59]. A theoretically well-studied system is vapor deposition of binary films containing two different kinds of molecules which have a tendency to phase separate [58, 59]. Similar to the dynamics of the two-species Eden model, there is an interplay between surface roughening and phase ordering kinetics: the dynamics of the domain boundaries and the surface are coupled through the growth kinetics. One main difference between the models is that for binary films the non-equilibrium roughening process is coupled to the coarsening dynamics of a thermodynamic model (Ising model) exhibiting detailed balance, while in our model it is coupled to the far-from-equilibrium dynamics of an absorbing state phase transition. In the binary film, particles of one type attach to domains containing particles of mainly the other type. If this is interpreted as 'mutation', then mutations are bi-directional and symmetric in the binary film. In addition, the growth rules differ in some important aspects. While in our two-species Eden model, the two species have different growth rates, in binary film growth, particle attachment at domain boundaries and within domains differs [49, 50, 56–59]. A common finding of these studies on binary films is that phase ordering kinetics increases surface roughness on length scales comparable to domain size; this is phenomenologically similar to our findings but the critical exponents are quite different. The reverse coupling of surface roughness to phase ordering kinetics is subtle and depends on the details of the kinetic growth rules. If growth at domain boundaries is faster than within domains, the surface roughness imprints its scaling properties on the domain boundaries [59], as in our case. Else, it appears that the domain boundaries perform random walks or even show non-universal behavior [48, 58, 59].

The enhanced roughness, which is induced by differential front velocities, bears some similarities to pinning models [45]. There, inhomogeneities (e.g. obstructions) of the medium locally reduce the growth speed, similar to mutant clusters in our model. The presence of these heterogeneities changes the scaling of the interface, since it introduces a length scale $\xi_\perp$, up to which one finds enhanced roughening. It is found that in the pinned phase, many pinning models can be mapped to DP, and the roughness of the surface is dominated by the DP critical exponents [83, 84]. In contrast to pinning models, where the disorder is quenched, the mutant clusters in the Eden model with mutation are dynamically generated and strongly correlated with the growth dynamics. We regard the latter as the main reason why the ensuing roughness and growth exponents are not related to DP.

Summarizing, we here argue that for multi-species range expansion, surface roughness and domain dynamics interfere with each other, which qualitatively changes both bulk and front properties. If absorbing front states exist, phase transitions to absorbing states of a new universality class are possible. More research is needed to explore the coupling between surface roughening and evolutionary dynamics of range-expansion scenarios. It would be especially interesting to look at systems where the reproduction rate of an individual depends on the local composition of the front as might be the case in growing biofilms [24]. Such systems,

where growth depends on a population's composition, have been analyzed in the context of one-dimensional wave propagation [85], the dilemma of cooperation in spatial settings [86], games with cyclic dominance [87] and structured populations [88]. A discrete 2d growth model, which incorporates these effects, would be most interesting to analyze with the methods developed in this paper. Moreover, the universality class of our model should be tested for different lattices, and critical exponents should be determined for higher-dimensional setups.

Acknowledgments

Financial support from the Deutsche Forschungsgemeinschaft, grant numbers FR850/9-1 and RA655/5-1, and the Elite Network of Bavaria (International Doctorate Program NanoBioTechnology) is gratefully acknowledged.

References

[1] Durrett R and Levin S 1994 The importance of being discrete (and spatial) *Theor. Population Biol.* **46** 363–94
[2] Murray J D 2007 *Mathematical Biology II* 3rd edn (Berlin: Springer)
[3] Shapiro J A 1995 The significances of bacterial colony patterns *BioEssays* **17** 597–607
[4] Ben-Jacob E, Aharonov Y and Shapira Y 1999 Bacteria harnessing complexity *Biofilms* **1** 239–63
[5] Kerr B, Riley M A, Feldman M W and Bohannan B J M 2002 Local dispersal promotes biodiversity in a real-life game of rock–paper–scissors *Nature* **418** 171–4
[6] Ben-Jacob E 2003 Bacterial self-organization: co-enhancement of complexification and adaptability in a dynamic environment *Phil. Trans.* A **361** 1283–312
[7] Buckling A, Maclean R C, Brockhurst M A and Colegrave N 2009 The Beagle in a bottle *Nature* **457** 824–9
[8] Be'er A, Ariel G, Kalisman O, Helman Y, Sirota-Madi A, Zhang H P, Florin E L, Payne S M, Ben-Jacob E and Swinney H L 2010 Lethal protein produced in response to competition between sibling bacterial colonies *Proc. Natl Acad. Sci. USA* **107** 6258–63
[9] Zhang H P, Be'er A, Florin E L and Swinney H L 2010 Collective motion and density fluctuations in bacterial colonies *Proc. Natl. Acad. Sci. USA* **107** 13626–30
[10] Reichenbach T, Mobilia M and Frey E 2007 Mobility promotes and jeopardizes biodiversity in rock–paper–scissors games *Nature* **448** 1046–9
[11] Reichenbach T, Mobilia M and Frey E 2008 Self-organization of mobile populations in cyclic competition *J. Theor. Biol.* **254** 368–83
[12] Wakano J, Maenosono S, Komoto A, Eiha N and Yamaguchi Y 2003 Self-organized pattern formation of a bacteria colony modeled by a reaction diffusion system and nucleation theory *Phys. Rev. Lett.* **90** 258102
[13] Cates M E, Marenduzzo D, Pagonabarraga I and Tailleur J 2010 Arrested phase separation in reproducing bacteria creates a generic route to pattern formation *Proc. Natl Acad. Sci. USA* **107** 11715–20
[14] Wakano J Y, Nowak M A and Hauert C 2009 Spatial dynamics of ecological public goods *Proc. Natl Acad. Sci. USA* **106** 7910–4
[15] Wakano J Y and Hauert C 2011 Pattern formation and chaos in spatial ecological public goods games *J. Theor. Biol.* **268** 30–8
[16] Melke P, Sahlin P, Levchenko A and Jönsson H 2010 A cell-based model for quorum sensing in heterogeneous bacterial colonies *PLoS Comput. Biol.* **6** e1000819
[17] Klopfstein S, Currat M and Excoffier L 2006 The fate of mutations surfing on the wave of a range expansion *Mol. Biol. Evol.* **23** 482–90
[18] Hallatschek O, Hersen P, Ramanathan S and Nelson D R 2007 Genetic drift at expanding frontiers promotes gene segregation *Proc. Natl Acad. Sci. USA* **104** 19926–30

[19] Excoffier L and Ray N 2008 Surfing during population expansions promotes genetic revolutions and structuration *Trends Ecol. Evol.* **23** 347–51
[20] Korolev K S, Hallatschek O and Nelson D R 2010 Genetic demixing and evolution in linear stepping stone models *Rev. Mod. Phys.* **82** 1691–718
[21] Ali A and Grosskinsky S 2010 Pattern formation through genetic drift at expanding population fronts *Adv. Complex Syst.* **13** 349–66
[22] Hallatschek O and Nelson D R 2010 Life at the front of an expanding population *Evolution* **64** 193–206
[23] Alberti L and Harshey R M 1990 Differentiation of *Serratia marcescens* 274 into swimmer and swarmer cells *J. Bacteriol.* **172** 4322–8
[24] Nadell C D, Xavier J B and Foster K R 2009 The sociobiology of biofilms *FEMS Microbiol. Rev.* **33** 206–24
[25] Griffin A S, West S A and Buckling A 2004 Cooperation and competition in pathogenic bacteria *Nature* **430** 1024–7
[26] Wilson A E, Kaul R B and Sarnelle O 2010 Growth rate consequences of coloniality in a harmful phytoplankter *PLoS One* **5** e8679
[27] Tuchscherr L *et al* 2011 Staphylococcus aureus phenotype switching: an effective bacterial strategy to escape host immune response and establish a chronic infection *EMBO Mol. Med.* **3** 129–41
[28] Neulinger S C, Jarnegren J, Ludvigsen M, Lochte K and Dullo W C 2008 Phenotype-specific bacterial communities in the cold-water coral *Lophelia pertusa* (Scleractinia) and their implications for the coral's nutrition, health and distribution *Appl. Environ. Microb.* **74** 7272–85
[29] Koshland D, Kent J C and Hartwell L H 1985 Genetic analysis of the mitotic transmission of minichromosomes *Cell* **40** 393–403
[30] Ben-Jacob E, Cohen I and Gutnick D L 1998 Cooperative organization of bacterial colonies: from genotype to morphotype *Annu. Rev. Microbiol.* **52** 779–806
[31] Golding I, Cohen I and Ben-Jacob E 1999 Studies of sector formation in expanding bacterial colonies *Europhys. Lett.* **48** 587–93
[32] Saito Y and Müller-Krumbhaar H 1995 Critical phenomena in morphology transitions of growth models with competition *Phys. Rev. Lett.* **74** 4325–8
[33] Murray J D 2007 *Mathematical Biology I* 3rd edn (Berlin: Springer)
[34] Hinrichsen H 2000 Non-equilibrium critical phenomena and phase transitions into absorbing states *Adv. Phys.* **49** 815–958
[35] Hinrichsen H 2006 Non-equilibrium phase transitions *Physica* A **369** 1–28
[36] Ódor G 2004 Universality classes in nonequilibrium lattice systems *Rev. Mod. Phys.* **76** 663–724
[37] Henkel M, Hinrichsen H and Lübeck S 2009 *Non-Equilibrium Phase Transitions* vol 1 (Berlin: Springer)
[38] Budrene E O and Berg H C 1991 Complex patterns formed by motile cells of *Escherichia coli Nature* **349** 630–3
[39] Matsushita M, Wakita J, Itoh H, Watanabe K, Arai T, Matsuyama T, Sakaguchi H and Mimura M 1999 Formation of colony patterns by a bacterial cell population *Physica* A **274** 190–9
[40] Matsuyama T and Matsushita M 2001 Population morphogenesis by cooperative bacteria *Forma* **16** 307–26
[41] Eden M 1960 A two-dimensional growth process *Proc. 4th Berkeley Symp. on Mathematical Statistics and Probability* vol 4 pp 223–39
[42] Huergo M A C, Pasquale M A, Bolzan A E and Arvia A J 2010 Morphology and dynamic scaling analysis of cell colonies with linear growth fronts. *Phys. Rev.* E **82** 031903
[43] Krug J and Spohn H 1991 Kinetic roughening of growing surfaces *Solids Far from Equilibrium* ed C Godrèche (Cambridge: Cambridge University Press) chapter 6
[44] Halpin-Healy T and Zhang Y 1995 Kinetic roughening phenomena, stochastic growth, directed polymers and all that. Aspects of multidisciplinary statistical mechanics *Phys. Rep.* **254** 215–414
[45] Barabási A-L and Stanley E H 1995 *Fractal Concepts in Surface Growth?* (Cambridge: Cambridge University Press)
[46] Jullien R and Botet R 1985 Surface thickness in the Eden model *Phys. Rev. Lett.* **54** 2055

[47] Kardar M, Parisi G and Zhang Y 1986 Dynamic scaling of growing interfaces *Phys. Rev. Lett.* **56** 889–92
[48] Derrida B and Dickman R 1991 On the interface between two growing Eden clusters *J. Phys. A: Math Gen.* **24** L191–5
[49] Ausloos M, Vandewalle N and Cloots R 1993 Magnetic Eden model *Europhys. Lett.* **24** 629–34
[50] Ausloos M and Vandewalle N 1996 Growth models with internal competition *Acta Phy. Pol.* B **27** 737–46
[51] Pellegrini Y P and Jullien R 1990 Roughening transition and percolation in random ballistic deposition *Phys. Rev. Lett.* **64** 1745–8
[52] Wang W and Cerdeira H A 1993 Kinetic growth of randomlike and ballisticlike deposition models *Phys. Rev.* E **47** 3357–61
[53] Wang W and Cerdeira H A 1995 Surface growth of two kinds of particle deposition models *Phys. Rev.* E **52** 6308–13
[54] El-Nashar H F, Wang W and Cerdeira H A 1996 Surface growth kinetics and morphological structural transition in a (2 + 1)-dimensional deposition model *J. Phys.: Condens. Matter* **8** 3271–83
[55] El-Nashar H F and Cerdeira H A 1999 Growth kinetics and morphology of a ballistic deposition model that incorporates surface diffusion for two species *Phys. Rev.* E **60** 1262–8
[56] Kotrla M and Predota M 1997 Interplay between kinetic roughening and phase ordering *Europhys. Lett.* **39** 251–6
[57] Kotrla M, Slanina F and Predota M 1998 Scaling in a two-component surface-growth model *Phys. Rev.* B **58** 10003–11
[58] Drossel B and Kardar M 2000 Phase ordering and roughening on growing films *Phys. Rev. Lett.* **85** 614–7
[59] Drossel B and Kardar M 2003 Interplay between phase ordering and roughening on growing films *Eur. Phys. J.* B **36** 401–10
[60] Reis F D A A 2002 Dynamic transition in deposition with a poisoning species *Phys. Rev.* E **66** 027101
[61] Harris T E 1974 Contact interactions on a lattice *Ann. Probab.* **2** 969–88
[62] Broadbent S R and Hammersley J M 1957 Percolation processes *Math. Proc. Camb. Phil. Soc.* **53** 629–41
[63] Stauffer D and Aharony A 1994 *Introduction to Percolation Theory* (London: Taylor and Francis)
[64] Gillespie D T 1976 A general method for numerically simulating the stochastic time evolution of coupled chemical reactions *J. Comput. Phys.* **22** 403–34
[65] Edmonds C A, Lillie A S and Cavalli-Sforza L L 2004 Mutations arising in the wave front of an expanding population *Proc. Natl Acad. Sci. USA* **101** 975–9
[66] Travis J M J, Münkemüller T, Burton O J, Best A, Dytham C and Johst K 2007 Deleterious mutations can surf to high densities on the wave front of an expanding population *Mol. Biol. Evol.* **24** 2334–43
[67] Hallatschek O and Nelson D R 2008 Gene surfing in expanding populations *Theor. Popul. Biol.* **73** 158–70
[68] Frey E 2010 Evolutionary game theory: theoretical concepts and applications to microbial communities *Physica* A **389** 4265–98
[69] Cremer J, Reichenbach T and Frey E 2009 The edge of neutral evolution in social dilemmas *New J. Phys.* **11** 093029
[70] Mendes J F F, Dickman R, Henkel M and Marques M C 1994 Generalized scaling for modes with multiple absorbing states *J. Phys. A : Math. Gen.* **27** 3019–28
[71] Jensen I and Dickman R 1993 Nonequilibrium phase transitions in systems with infinitely many absorbing states *Phys. Rev.* E **48** 1710–25
[72] Muñoz M A, Grinstein G, Dickman R and Livi R 1996 Critical behavior of systems with many absorbing states *Phys. Rev. Lett.* **76** 451–4
[73] Domany E and Kinzel W 1984 Equivalence of cellular automata to Ising models and directed percolation *Phys. Rev. Lett.* **53** 311–4
[74] Lübeck S 2006 Crossover scaling in the Domany–Kinzel cellular automaton *J. Stat. Mech.* P09009
[75] Das Sarma S, Ghaisas S V and Kim J M 1994 Kinetic super-roughening and anomalous dynamic scaling in nonequilibrium growth models *Phys. Rev.* E **49** 122–5

[76] Lopez J M, Rodriguez M A and Cuerno R 1997 Superroughening versus intrinsic anomalous scaling of surfaces *Phys. Rev.* E **56** 3993–8
[77] Plischke M, Shore J, Schroeder M, Siegert M and Wolf D 1993 Comment on 'Solid-on-solid rules and models for nonequilibrium growth in 2+ 1 dimensions' *Phys. Rev. Lett.* **71** 2509–9
[78] Schroeder M, Siegert M, Wolf D E, Shore J D and Plischke M 1993 Scaling of growing surfaces with large local slopes *Europhys. Lett.* **24** 563–8
[79] Schmittbuhl J, Vilotte J-P and Roux S 1995 Reliability of self-affine measurements *Phys. Rev.* E **51** 131–47
[80] Siegert M 1996 Determining exponents in models of kinetic surface roughening *Phys. Rev.* E **53** 3209–14
[81] Janssen H K 1981 On the nonequilibrium phase transition in reaction–diffusion systems with an absorbing stationary state *Z. Phys.* B **42** 151–4
[82] Grassberger P 1982 On phase transitions in Schlögl's second model *Z. Phys.* B **47** 365–74
[83] Buldyrev S, Barabási A L, Caserta F, Havlin S, Stanley H and Vicsek T 1992 Anomalous interface roughening in porous media: experiment and model *Phys. Rev.* A **45** R8313–6
[84] Tang L-H and Leschhorn H 1992 Pinning by directed percolation *Phys. Rev.* A **45** R8309–12
[85] Hallatschek O 2011 Noise driven evolutionary waves *PLoS Comput. Biol.* **7** e1002005
[86] Nadell C D, Foster K R and Xavier J B 2010 Emergence of spatial structure in cell groups and the evolution of cooperation *PLoS Comput. Biol.* **6** e1000716
[87] Rulands S, Reichenbach T and Frey E 2011 Threefold way to extinction in populations of cyclically competing species *J. Stat. Mech.* L01003
[88] Cremer J, Melbinger A and Frey E Demographic fluctuations promote the evolution of cooperation (unpublished)

Bibliography

[1] H. H. McAdams and A. P. Arkin, It's a noisy business! Genetic regulation at the nanomolar scale, Trends Genet **15**, 65 (1999).

[2] C. V. Rao, D. M. Wolf, and A. P. Arkin, Control, exploitation and tolerance of intracellular noise, Nature **420**, 231 (2002).

[3] H. H. McAdams and A. P. Arkin, Stochastic mechanisms in gene expression, Proc Natl Acad Sci USA **94**, 814 (1997).

[4] M. Thattai and A. van Oudenaarden, Intrinsic noise in gene regulatory networks, Proc Natl Acad Sci USA **98**, 8614 (2001).

[5] E. M. Ozbudak, M. Thattai, I. Kurtser, A. D. Grossman, and A. van Oudenaarden, Regulation of noise in the expression of a single gene, Nat. Genet. **31**, 69 (2002).

[6] A. Einstein, Über die von der molekularkinetischen Theorie der Wärme geforderte Bewegung von in ruhenden Flüssigkeiten suspendierten Teilchen, Annalen der Physik **322**, 549 (1905).

[7] P. Langevin, On the Theory of Brownian Motion (Sur la théorie du mouvement brownien), Comptes Rendus de l'Académie des Sciences **146**, 530 (1908).

[8] M. von Smoluchowski, Zur kinetischen Theorie der Brownschen Molekularbewegung und der Suspensionen, Annalen der Physik **326**, 756 (1906).

[9] E. Frey and K. Kroy, Brownian motion: a paradigm of soft matter and biological physics, Annalen der Physik **14**, 20 (2005).

[10] M. Delbrück, Statistical Fluctuations in Autocatalytic Reactions, The Journal of Chemical Physics **8**, 120 (1940).

[11] S. Luria and M. Delbrück, Mutations of bacteria from virus sensitivity to virus resistance, Genetics **28**, 491 (1943).

[12] L. Reese, A. Melbinger, and E. Frey, Crowding of molecular motors determines microtubule depolymerization, arXiv **physics.bio-ph** (2011).

[13] W. Moebius and U. Gerland, Quantitative Test of the Barrier Nucleosome Model for Statistical Positioning of Nucleosomes Up- and Downstream of Transcription Start Sites, PLoS Comput Biol **6**, (2010).

[14] L. Bintu, N. E. Buchler, H. G. Garcia, U. Gerland, T. Hwa, J. Kondev, T. Kuhlman, and R. Phillips, Transcriptional regulation by the numbers: applications, Curr Opin Genet Dev **15**, 125 (2005).

[15] L. Bintu, N. E. Buchler, H. G. Garcia, U. Gerland, T. Hwa, J. Kondev, and R. Phillips, Transcriptional regulation by the numbers: models, Curr Opin Genet Dev **15**, 116 (2005).

[16] G. Fritz, N. E. Buchler, T. Hwa, and U. Gerland, Designing sequential transcription logic: a simple genetic circuit for conditional memory, Syst Synth Biol **1**, 89 (2007).

[17] M. Leisner, J.-T. Kuhr, J. O. Rädler, E. Frey, and B. Maier, Kinetics of genetic switching into the state of bacterial competence, Biophys J **96**, 1178 (2009).

[18] G. Schwake, S. Youssef, J.-T. Kuhr, S. Gude, M. P. David, E. Mendoza, E. Frey, and J. O. Rädler, Predictive modeling of non-viral gene transfer, Biotechnol Bioeng **105**, 805 (2010).

[19] J. Halatek and E. Frey, A highly canalized MinD transfer and MinE sequestration explain the origin of robust MinCDE-protein dynamics, to be published .

[20] K. Alim and E. Frey, Quantitative predictions on auxin-induced polar distribution of PIN proteins during vein formation in leaves, Eur Phys J E **33**, 165 (2010).

[21] J.-T. Kuhr, M. Leisner, and E. Frey, Range Expansion with Mutation and Selection: Dynamical Phase Transition in a Two-Species Eden Model, arXiv **cond-mat.stat-mech** (2011), accepted at New Journal of Physics.

[22] J. Cremer, A. Melbinger, and E. Frey, Demographic fluctuations promote the evolution of cooperation, to be published .

[23] B. Alberts, A. Johnson, J. Lewis, M. Raff, K. Roberts, and P. Walter, *Molecular Biology of the Cell*, Garland Science, 4th edition, 2002.

[24] G. Mulder, Ueber die Zusammensetzung einiger thierischen Substanzen, Journal für Praktische Chemie (1839).

[25] R. Dahm, Discovering DNA: Friedrich Miescher and the early years of nucleic acid research, Hum. Genet. **122**, 565 (2008).

[26] R. D. Simoni, R. L. Hill, and M. Vaughan, The Structure of Nucleic Acids and Many Other Natural Products: Phoebus Aaron Levene, Journal of Biological Chemistry **277**, 23 (2002).

[27] J. Huxley, *Evolution: The Modern Synthesis*, Allen & Unwin, Lodon, 1942.

[28] E. Schrödinger, *What Is Life?*, The Physical Aspect of the Living Cell, The Macmillan Company, 1945.

[29] G. Mendel, *Versuche uber Pflanzen-Hybriden*, volume IV, Verhandlungen des naturforschenden Vereins Brünn, 1866.

[30] C. Darwin, *On the Origin of Species by Means of Natural Selection, or the Preservation of Favoured Races in the Struggle for Life*, John Murray, 1859.

[31] G. W. Beadle and E. L. Tatum, Genetic Control of Biochemical Reactions in Neurospora, Proc Natl Acad Sci USA **27**, 499 (1941).

[32] O. T. Avery, C. M. MacLeod, and M. McCarty, Studies on the Chemical Nature of the Substance Inducing Transformation of Pneumococcal Types Induction of Transformation by a Desoxyribonucleic Acid Fraction Isolated From Pneumococcus Type III, J Exp Med **79**, 137 (1944).

[33] A. D. Hershey and M. Chase, Independent Functions of Viral Protein and Nucleic Acid in Growth of Bacteriophage, J Gen Physiol **36**, 39 (1952).

[34] E. Chargaff, Chemical Specificity of Nucleic Acids and Mechanism of Their Enzymatic Degradation, Experientia **6**, 201 (1950).

[35] R. E. Franklin and R. G. Gosling, Molecular Configuration in Sodium Thymonucleate, Nature **171**, 740 (1953).

[36] M. H. F. Wilkins, A. R. Stokes, and H. R. Wilson, Molecular structure of deoxypentose nucleic acids, Nature **171**, 738 (1953).

[37] J. D. Watson and F. H. C. Crick, Molecular structure of nucleic acids, Nature **171**, 737 (1953).

[38] J. D. Watson and F. H. C. Crick, Genetical implications of the structure of deoxyribonucleic acid, Nature **171**, 964 (1953).

[39] F. Crick, *What Mad Pursuit: A Personal View of Scientific Discovery*, Basic Books, 1990.

[40] F. Crick, On protein synthesis, Symposia of the Society for Experimental Biology **12**, 138 (1958).

[41] G. Gamow, Possible Relation between Deoxyribonucleic Acid and Protein Structures, Nature **173**, 318 (1954).

[42] F. H. C. Crick, F. R. S. Leslie Barnett, S. Brenner, and R. J. Watts-Tobin, General Nature of Genetic Code for Proteins, Nature **192**, 1227 (1961).

[43] J. H. Matthaei and M. W. Nirenberg, Characteristics and stabilization of DNAase-sensitive protein synthesis in E. coli extracts, Proc Natl Acad Sci USA **47**, 1580 (1961).

[44] M. W. Nirenberg and J. H. Matthaei, The dependence of cell-free protein synthesis in E. coli upon naturally occurring or synthetic polyribonucleotides, Proc Natl Acad Sci USA **47**, 1588 (1961).

[45] R. S. Gardner, A. J. Wahba, C. Basilio, R. S. Miller, P. Lengyel, and J. F. Speyer, Synthetic polynucleotides and the amino acid code. VII, Proc Natl Acad Sci USA **48**, 2087 (1962).

[46] A. J. Wahba, R. S. Gardner, C. Basilio, R. S. Miller, J. F. Speyer, and P. Lengyel, Synthetic polynucleotides and the amino acid code. VIII, Proc Natl Acad Sci USA **49**, 116 (1963).

[47] M. W. Nirenberg, P. Leder, M. Bernfield, R. Brimacombe, J. Trupin, F. Rottman, and C. O'Neal, RNA codewords and protein synthesis, VII. On the general nature of the RNA code, Proc Natl Acad Sci USA **53**, 1161 (1965).

[48] H. P. Ghosh, D. Söll, and H. G. Khorana, Studies on polynucleotides. LXVII. Initiation of protein synthesis in vitro as studied by using ribopolynucleotides with repeating nucleotide sequences as messengers, Journal of Molecular Biology **25**, 275 (1967).

[49] F. Crick, Central dogma of molecular biology, Nature **227**, 561 (1970).

[50] C. H. Waddington, *The Strategy of the Genes: A discussion of some aspects of theoretical biology*, Macmillan, 1957.

[51] V. E. A. Russo, R. A. Martienssen, and A. D. Riggs, editors, *Epigenetic Mechanisms of Gene Regulation*, Cold Spring Harbor Laboratory Press, 1996.

[52] S. Prusiner, Prions, Proc. Natl. Acad. Sci. U.S.A. **95**, 13363 (1998).

[53] T. R. Cech, The RNA Worlds in Context, Cold Spring Harb Perspect Biol (2011).

[54] F. Jacob and J. Monod, Genetic regulatory mechanisms in the synthesis of proteins, Journal of Molecular Biology **3**, 318 (1961).

[55] U. Alon, *An Introduction to Systems Biology: Design Principles of Biological Circuits*, Chapman and Hall/CRC, 1st edition, 2006.

[56] M. Akerfelt, R. I. Morimoto, and L. Sistonen, Heat shock factors: integrators of cell stress, development and lifespan, Nat. Rev. Mol. Cell Biol. **11**, 545 (2010).

[57] S. L. Porter, G. H. Wadhams, and J. P. Armitage, Signal processing in complex chemotaxis pathways, Nat Rev Micro **9**, 153 (2011).

[58] M. B. Miller and B. L. Bassler, Quorum sensing in bacteria, Annu. Rev. Microbiol. **55**, 165 (2001).

[59] R. G. Foster and L. Kreitzman, *Rhythms of Life: The Biological Clocks that Control the Daily Lives of Every Living Thing*, Yale University Press, 2005.

[60] E. Davidson and M. Levin, Gene regulatory networks, Proc Natl Acad Sci USA **102**, 4935 (2005).

[61] S. B. Carroll, *Endless Forms Most Beautiful: The New Science of Evo Devo*, W. W. Norton & Company, 2006.

[62] E. Purcell, Life at low Reynolds number, Am. J. Phys **45**, 3 (1977).

[63] T. Lucretius Carus, *On the Nature of Things - De Rerum Natura*, Project Gutenberg, about 60 BC.

[64] M. B. Elowitz, A. J. Levine, E. D. Siggia, and P. S. Swain, Stochastic gene expression in a single cell, Science **297**, 1183 (2002).

[65] A. P. Arkin, J. Ross, and H. H. McAdams, Stochastic kinetic analysis of developmental pathway bifurcation in phage lambda-infected Escherichia coli cells, Genetics **149**, 1633 (1998).

[66] A. Eldar and M. B. Elowitz, Functional roles for noise in genetic circuits, Nature **467**, 167 (2010).

[67] O. Shimomura, F. H. Johnson, and Y. Saiga, Extraction, purification and properties of aequorin, a bioluminescent protein from the luminous hydromedusan, Aequorea, J Cell Compar Physl **59**, 223 (1962).

[68] M. Chalfie, Y. Tu, G. Euskirchen, W. W. Ward, and D. C. Prasher, Green fluorescent protein as a marker for gene expression, Science **263**, 802 (1994).

[69] R. Heim, D. C. Prasher, and R. Y. Tsien, Wavelength mutations and posttranslational autoxidation of green fluorescent protein, Proc Natl Acad Sci USA **91**, 12501 (1994).

[70] M. Ormö, A. B. Cubitt, K. Kallio, L. A. Gross, R. Y. Tsien, and S. J. Remington, Crystal structure of the Aequorea victoria green fluorescent protein, Science **273**, 1392 (1996).

[71] R. Heim, A. B. Cubitt, and R. Y. Tsien, Improved green fluorescence, Nature **373**, 663 (1995).

[72] S. H. Strogatz, *Nonlinear Dynamics And Chaos: With Applications To Physics, Biology, Chemistry, And Engineering*, Westview Press, 1 edition, 2001.

[73] S. Wiggins, *Introduction to Applied Nonlinear Dynamical Systems and Chaos*, Springer, 2nd edition, 2010.

[74] J. D. Murray, *Mathematical Biology I. An Introduction*, volume 1, Springer, 3rd edition, 2007.

[75] C. W. Gardiner, *Handbook of Stochastic Methods for Physics, Chemistry and the Natural Sciences*, Springer, 2nd edition, 1997.

[76] N. G. Van Kampen, *Stochastic Processes in Physics and Chemistry*, North-Holland Personal Library, 3rd edition, 2007.

[77] N. T. J. Bailey, *The Elements of Stochastic Processes with Applications to the Natural Sciences*, Wiley-IEEE, 1990.

[78] D. T. Gillespie, A General Method for Numerically Simulating the Stochastic Time Evolution of Coupled Chemical Reactions, Journal of Comutational Physics **22**, 403 (1976).

[79] D. T. Gillespie, Exact stochastic simulation of coupled chemical reactions, The Journal of Physical Chemistry (1977).

[80] H. Risken and T. Frank, *The Fokker-Planck Equation: Methods of Solutions and Applications*, Springer, 2nd edition, 1996.

[81] V. J. Freeman, Studies on the virulence of bacteriophage-infected strains of Corynebacterium diphtheriae, Journal of Bacteriology **61**, 675 (1951).

[82] K. Ochiai, T. Yamanaka, K. Kimura, and O. Sawada, *Inheritance of drug resistance (and its transfer) between Shigella strains and Between Shigella and E. coli strains*, volume 1861, Hihon Iji Shimpor, 1959.

[83] M. Syvanen, Cross-species gene transfer; implications for a new theory of evolution, Journal of Theoretical Biology **112**, 333 (1985).

[84] J. Lederberg and E. L. Tatum, Gene recombination in Escherichia coli, Nature **158**, 558 (1946).

[85] F. Griffith, The Significance of Pneumococcal Types, The Journal of Hygiene **27**, 113 (1928).

[86] J. Lederberg, E. M. Lederberg, N. D. Zinder, and E. R. Lively, Recombination analysis of bacterial heredity, Cold Spring Harbor Symposia on Quantitative Biology **16**, 413 (1951).

[87] C. Mereschkowsky, Über Natur und Ursprung der Chromatophoren im Pflanzenreiche, Biol. Centralbl. **25**, 593 (1905).

[88] I. Wallin, The mitochondria problem, American Naturalist **57**, 255 (1923).

[89] N. M. Calvin and P. C. Hanawalt, High-efficiency transformation of bacterial cells by electroporation, Journal of Bacteriology **170**, 2796 (1988).

[90] Y. Song, T. Hahn, I. P. Thompson, T. J. Mason, G. M. Preston, G. Li, L. Paniwnyk, and W. E. Huang, Ultrasound-mediated DNA transfer for bacteria, Nucleic Acids Res **35**, e129 (2007).

[91] M. Tsukakoshi, S. Kurata, Y. Nomiya, Y. Ikawa, and K. T, A Novel Method of Dna Transfection by Laser Microbeam Cell Surgery, Appl Phys B **35**, 135 (1984).

[92] T. E. McKnight, A. V. Melechko, D. K. Hensley, D. G. J. Mann, G. D. Griffin, and M. L. Simpson, Tracking Gene Expression after DNA Delivery Using Spatially Indexed Nanofiber Arrays, Nano Lett. **4**, 1213 (2004).

[93] A. Grässmann and M. Grässmann, Über die Bildung von Melanin in Muskelzellen nach der direkten Übertragung von RNA aus HARDING-PASSEY-Melanomzellen, Hoppe-Seyler's Zeitschrift für physiologische Chemie **352**, 527 (1971).

[94] T. M. Klein, E. D. Wolf, R. Wu, and J. C. Sanford, High-velocity microprojectiles for delivering nucleic acids into living cells, Nature **327**, 70 (1987).

[95] F. Scherer, M. Anton, U. Schillinger, J. Henke, C. Bergemann, A. Krüger, B. Gänsbacher, and C. Plank, Magnetofection: enhancing and targeting gene delivery by magnetic force in vitro and in vivo, Gene Ther. **9**, 102 (2002).

[96] F. Graham and van der Eb A J, A new technique for the assay of infectivity of human adenovirus 5 DNA, Virology **52**, 456 (1973).

[97] P. L. Felgner, T. R. Gadek, M. Holm, R. Roman, H. W. Chan, M. Wenz, J. P. Northrop, G. M. Ringold, and M. Danielsen, Lipofection: a highly efficient, lipid-mediated DNA-transfection procedure, Proc Natl Acad Sci USA **84**, 7413 (1987).

[98] O. Boussif, F. Lezoualc'h, M. A. Zanta, M. D. Mergny, D. Scherman, B. Demeneix, and J. P. Behr, A versatile vector for gene and oligonucleotide transfer into cells in culture and in vivo: polyethylenimine, Proc Natl Acad Sci USA **92**, 7297 (1995).

[99] S. Brunner, T. Sauer, S. Carotta, M. Cotten, M. Saltik, and E. Wagner, Cell cycle dependence of gene transfer by lipoplex, polyplex and recombinant adenovirus, Gene Ther. **7**, 401 (2000).

[100] I. Mortimer, P. Tam, I. MacLachlan, R. W. Graham, E. G. Saravolac, and P. B. Joshi, Cationic lipid-mediated transfection of cells in culture requires mitotic activity, Gene Ther. **6**, 403 (1999).

[101] W. C. Tseng, F. R. Haselton, and T. D. Giorgio, Mitosis enhances transgene expression of plasmid delivered by cationic liposomes, Biochim. Biophys. Acta **1445**, 53 (1999).

[102] D. A. Jackson, R. H. Symons, and P. Berg, Biochemical Method for Inserting New Genetic Information into DNA of Simian Virus 40: Circular SV40 DNA Molecules Containing Lambda Phage Genes and the Galactose Operon of Escherichia coli, Proc Natl Acad Sci USA **69**, 2904 (1972).

[103] S. N. Cohen and A. C. Chang, Recircularization and autonomous replication of a sheared R-factor DNA segment in Escherichia coli transformants, Proc Natl Acad Sci USA **70**, 1293 (1973).

[104] R. Jaenisch and B. Mintz, Simian virus 40 DNA sequences in DNA of healthy adult mice derived from preimplantation blastocysts injected with viral DNA, Proc Natl Acad Sci USA **71**, 1250 (1974).

[105] J. C. Avise, *The Hope, Hype, and Reality of Genetic Engineering: Remarkable Stories from Agriculture, Industry, Medicine, and the Environment*, Oxford University Press, USA, 2004.

[106] M. B. Elowitz and S. Leibler, A synthetic oscillatory network of transcriptional regulators, Nature **403**, 335 (2000).

[107] J. J. Collins, T. S. Gardner, and C. R. Cantor, Construction of a genetic toggle switch in Escherichia coli, Nature **403**, 339 (2000).

[108] E. Andrianantoandro, S. Basu, D. K. Karig, and R. Weiss, Synthetic biology: new engineering rules for an emerging discipline, Mol Syst Biol **2**, 2006.0028 (2006).

[109] D. G. Gibson et al., Complete Chemical Synthesis, Assembly, and Cloning of a Mycoplasma genitalium Genome, Science **319**, 1215 (2008).

[110] D. G. Gibson et al., Creation of a Bacterial Cell Controlled by a Chemically Synthesized Genome, Science **329**, 52 (2010).

[111] C. M. Varga, T. J. Wickham, and D. A. Lauffenburger, Receptor-mediated targeting of gene delivery vectors: insights from molecular mechanisms for improved vehicle design, Biotechnol Bioeng **70**, 593 (2000).

[112] C. M. Varga, K. Hong, and D. A. Lauffenburger, Quantitative analysis of synthetic gene delivery vector design properties, Mol. Ther. **4**, 438 (2001).

[113] C. Wiethoff and C. Middaugh, Barriers to nonviral gene delivery, J Pharm Sci **92**, 203 (2003).

[114] C. Varga, N. Tedford, M. Thomas, A. Klibanov, L. Griffith, and D. Lauffenburger, Quantitative comparison of polyethylenimine formulations and adenoviral vectors in terms of intracellular gene delivery processes, Gene Ther. **12**, 1023 (2005).

[115] C. G. Ehrenberg, Physikalische Abhandlungen der Koeniglichen Akademie der Wissenschaften zu Berlin aus den Jahren 1833–1835 , 145 (1835).

[116] P. Graumann, editor, *Bacillus: Cellular and Molecular Biology*, Caister Academic Press, 1 edition, 2007.

[117] M. Thattai and A. van Oudenaarden, Stochastic gene expression in fluctuating environments, Genetics **167**, 523 (2004).

[118] W. K. Smits, O. P. Kuipers, and J.-W. Veening, Phenotypic variation in bacteria: the role of feedback regulation, Nat Rev Micro **4**, 259 (2006).

[119] J. S. Griffith, *Mathematical Neurobiology*, Academic Press Inc, 1971.

[120] J. M. Raser and E. K. O'Shea, Control of stochasticity in eukaryotic gene expression, Science **304**, 1811 (2004).

[121] M. Assaf, E. Roberts, and Z. Luthey-Schulten, Determining the Stability of Genetic Switches: Explicitly Accounting for mRNA Noise, Phys Rev Lett **106** (2011).

[122] M. Leisner, K. Stingl, J. O. Rädler, and B. Maier, Basal expression rate of comK sets a 'switching-window' into the K-state of Bacillus subtilis, Mol Microbiol **63**, 1806 (2007).

[123] C. Stringer, Human evolution: Out of Ethiopia, Nature **423**, 692 (2003).

[124] V. Macaulay et al., Single, rapid coastal settlement of Asia revealed by analysis of complete mitochondrial genomes, Science **308**, 1034 (2005).

[125] H. Liu, F. Prugnolle, A. Manica, and F. Balloux, A Geographically Explicit Genetic Model of Worldwide Human-Settlement History, The American Journal of Human Genetics **79**, 230 (2006).

[126] M. Rasmussen et al., An Aboriginal Australian Genome Reveals Separate Human Dispersals into Asia, Science **334**, 94 (2011).

[127] National Invasive Species Information Center: http://www.invasivespeciesinfo.gov/.

[128] Global Invasive Species Database: http://www.issg.org/database.

[129] R. W. Griffiths, D. W. Schloesser, J. H. Leach, and W. P. Kovalak, Distribution and Dispersal of the Zebra Mussel (Dreissena polymorpha) in the Great Lakes Region, Canadian Journal of Fisheries and Aquatic Sciences **48**, 1381 (1991).

[130] D. Pimentel, L. Lach, R. Zuniga, and D. Morrison, Environmental and Economic Costs of Nonindigenous Species in the United States, BioScience **50**, 53 (2000).

[131] W. F. Ruddiman, The Anthropogenic Greenhouse Era Began Thousands of Years Ago, Climatic Change **61**, 261 (2003).

[132] C. Parmesan and G. Yohe, A globally coherent fingerprint of climate change impacts across natural systems, Nature **421**, 37 (2003).

[133] S. R. Loarie, P. B. Duffy, H. Hamilton, G. P. Asner, C. B. Field, and D. D. Ackerly, The velocity of climate change, Nature **462**, 1052 (2009).

[134] I. C. Chen, J. K. Hill, R. Ohlemuller, D. B. Roy, and C. D. Thomas, Rapid Range Shifts of Species Associated with High Levels of Climate Warming, Science **333**, 1024 (2011).

[135] S. Ramachandran, Support from the relationship of genetic and geographic distance in human populations for a serial founder effect originating in Africa, Proc. Natl. Acad. Sci. U.S.A. **102**, 15942 (2005).

[136] M. DeGiorgio, M. Jakobsson, and N. A. Rosenberg, Out of Africa: Modern Human Origins Special Feature: Explaining worldwide patterns of human genetic variation using a coalescent-based serial founder model of migration outward from Africa, Proc. Natl. Acad. Sci. U.S.A. **106**, 16057 (2009).

[137] J. H. Gillespie, *Population Genetics: A Concise Guide*, The Johns Hopkins University Press, 2nd edition, 2004.

[138] W. J. Ewens, *Mathematical Population Genetics: Theoretical introduction*, Springer Verlag, 2nd edition, 2004.

[139] E. Mayr, *Systematics and the origin of species*, Columbia University Press, 1942.

[140] J. A. Shapiro, The significances of bacterial colony patterns, Bioessays **17**, 597 (1995).

[141] E. Ben-Jacob, I. Cohen, O. Shochet, I. Aranson, H. Levine, and L. Tsimring, Complex bacterial patterns, Nature **373**, 566 (1995).

[142] E. Ben-Jacob, I. Cohen, and D. L. Gutnick, Cooperative organization of bacterial colonies: from genotype to morphotype, Annu. Rev. Microbiol. **52**, 779 (1998).

[143] E. B. Jacob, Y. Aharonov, and Y. Shapira, Bacteria harnessing complexity, Biofilms **1**, 239 (1999).

[144] E. Ben-Jacob, Bacterial self-organization: co-enhancement of complexification and adaptability in a dynamic environment, Philos Transact A Math Phys Eng Sci **361**, 1283 (2003).

[145] J. A. Shapiro, Organization of Developing Escherichia coli Colonies Viewed by Scanning Electron Microscopy, Journal of Bacteriology **169**, 142 (1987).

[146] H. P. Zhang, A. Be'er, E. L. Florin, and H. L. Swinney, Collective motion and density fluctuations in bacterial colonies, Proc. Natl. Acad. Sci. U.S.A. **107**, 13626 (2010).

[147] M. Matsushita, J. Wakita, H. Itoh, K. Watanabe, T. Arai, T. Matsuyama, H. Sakaguchi, and M. Mimura, Formation of colony patterns by a bacterial cell population, Physica A **274**, 190 (1999).

[148] T. Matsuyama and M. Matsushita, Population morphogenesis by cooperative bacteria, Forma **16**, 307 (2001).

[149] M. Matsushita and H. Fujikawa, Diffusion-limited growth in bacterial colony formation, Physica A: Statistical and Theoretical Physics **168**, 498 (1990).

[150] T. Witten and L. Sander, Diffusion-Limited Aggregation, a Kinetic Critical Phenomenon, Phys Rev Lett **47**, 1400 (1981).

[151] H. E. Stanley, A. Coniglio, S. Havlin, J. Lee, S. Schwarzer, and M. Wolf, Diffusion limited aggregation: a paradigm of disorderly cluster growth, Physica A: Statistical and Theoretical Physics **205**, 254 (1994).

[152] A.-L. Barabási and H. E. Stanley, *Fractal concepts in surface growth*, Cambridge University Press, 1st edition, 1995.

[153] M. Eden, A two-dimensional growth process, Proc of the Fourth Berkeley Symposium on Mathematical Statistics and Probability **4**, 223 (1960).

[154] T. Vicsek, M. Cserzo, and V. Horvath, Self-affine growth of bacterial colonies, Physica A **167**, 315 (1990).

[155] M. Huergo, M. Pasquale, A. Bolzán, A. Arvia, and P. González, Morphology and dynamic scaling analysis of cell colonies with linear growth fronts, Phys Rev E **82**, 031903 (2010).

[156] M. Kimura, *The Neutral Theory of Molecular Evolution*, Cambridge University Press, 1985.

[157] J. Cremer, T. Reichenbach, and E. Frey, The edge of neutral evolution in social dilemmas, New J. Phys. **11**, 093029 (2009).

[158] D. Koshland, J. C. Kent, and L. H. Hartwell, Genetic analysis of the mitotic transmission of minichromosomes, Cell **40**, 393 (1985).

[159] I. Golding, I. Cohen, and E. Ben-Jacob, Studies of sector formation in expanding bacterial colonies, Europhys Lett **48**, 587 (1999).

[160] O. Hallatschek and D. R. Nelson, Life at the front of an expanding population, Evolution **64**, 193 (2010).

[161] O. Hallatschek, P. Hersen, S. Ramanathan, and D. R. Nelson, Genetic drift at expanding frontiers promotes gene segregation, Proc Natl Acad Sci USA **104**, 19926 (2007).

[162] K. S. Korolev, O. Hallatschek, and D. R. Nelson, Genetic demixing and evolution in linear stepping stone models, Rev. Mod. Phys. **82**, 1691 (2010).

[163] Y. Saito and H. Müller-Krumbhaar, Critical phenomena in morphology transitions of growth models with competition, Phys Rev Lett **74**, 4325 (1995).

[164] M. Eigen, Selforganization of matter and the evolution of biological macromolecules, Naturwissenschaften **58**, 465 (1971).

[165] R. Fisher, The wave of advance of advantageous genes, Ann Eugenic **7**, 355 (1937).

[166] A. Kolmogorov, I. Petrovsky, and N. Piscounoff, *Etude de l'équation de la diffusion avec croissance de la quantité de matiere et son applicationa un probleme biologique*, volume 1, Mosc. Univ. Bull. Math, 1937.

[167] N. Shigesada and K. Kawasaki, *Biological Invasions: Theory and Practice (Oxford Series in Ecology & Evolution)*, Oxford University Press, USA, 1st edition, 1997.

[168] E. Brunet and B. Derrida, Shift in the velocity of a front due to a cutoff, Phys Rev E **56**, 2597 (1997).

[169] O. Hallatschek and K. S. Korolev, Fisher waves in the strong noise limit, Phys Rev Lett **103**, 108103 (2009).

[170] M. Kimura and G. H. Weiss, The Stepping Stone Model of Population Structure and the Decrease of Genetic Correlation with Distance, Genetics **49**, 561 (1964).

[171] K. S. Korolev and D. R. Nelson, Competition and cooperation in one-dimensional stepping-stone models, Phys Rev Lett **107**, 088103 (2011).

[172] C. A. Edmonds, A. S. Lillie, and L. L. Cavalli-Sforza, Mutations arising in the wave front of an expanding population, Proc Natl Acad Sci USA **101**, 975 (2004).

[173] J. M. J. Travis, T. Münkemüller, O. J. Burton, A. Best, C. Dytham, and K. Johst, Deleterious mutations can surf to high densities on the wave front of an expanding population, Mol Biol Evol **24**, 2334 (2007).

[174] L. Excoffier and N. Ray, Surfing during population expansions promotes genetic revolutions and structuration, Trends Ecol Evol **23**, 347 (2008).

[175] J. R. Miller, Survival of mutations arising during invasions, Evolutionary Applications **3**, 109 (2010).

[176] S. Klopfstein, M. Currat, and L. Excoffier, The fate of mutations surfing on the wave of a range expansion, Mol Biol Evol **23**, 482 (2006).

[177] J. Costerton, Introduction to biofilm, International Journal of Antimicrobial Agents **11**, 217 (1999).

[178] R. D. Monds and G. A. O'Toole, The developmental model of microbial biofilms: ten years of a paradigm up for review, Trends Microbiol **17**, 73 (2009).

[179] C. D. Nadell, J. B. Xavier, and K. R. Foster, The sociobiology of biofilms, FEMS Microbiol Rev **33**, 206 (2009).

[180] J. B. Xavier and K. R. Foster, Cooperation and conflict in microbial biofilms, Proc Natl Acad Sci USA **104**, 876 (2007).

[181] C. D. Nadell, K. R. Foster, and J. B. Xavier, Emergence of Spatial Structure in Cell Groups and the Evolution of Cooperation, PLoS Comput Biol **6**, e1000716 (2010).

[182] S. Mitri, J. B. Xavier, and K. R. Foster, Social evolution in multispecies biofilms, Proc. Natl. Acad. Sci. U.S.A. **108**, 10839 (2011).

[183] K. Huang, *Statistical Mechanics*, Wiley, 2nd edition, 1987.

[184] R. K. Pathria and P. D. Beale, *Statistical Mechanics*, Academic Press, 3rd edition, 2011.

[185] F. Schwabl, *Statistical Mechanics*, Springer, 2nd edition, 2006.

[186] M. Henkel, H. Hinrichsen, and S. Lübeck, *Non-equilibrium phase transitions: Absorbing phase transitions*, volume 1, Springer, 1st edition, 2008.

[187] G. Ódor, Universality classes in nonequilibrium lattice systems, Reviews of modern physics **76**, 663 (2004).

[188] J. L. Cardy, *Scaling and renormalization in statistical physics*, Cambridge Univ Pr, 1996.

[189] D. Stauffer and A. Aharony, *Introduction To Percolation Theory*, Taylor & Francis, 2nd edition, 1994.

[190] P. Grassberger and A. de la Torre, Reggeon field theory (Schlögl's first model) on a lattice: Monte Carlo calculations of critical behaviour, Annals of Physics **122**, 373 (1979).

[191] J. Krug and H. Spohn, Kinetic roughening of growing surfaces, in *Solids far from Equilibrium*, edited by C. Godrèche, Cambridge University Press, 1992.

[192] F. Family and T. Vicsek, Scaling of the active zone in the Eden process on percolation networks and the ballistic deposition model, J. Phys. A **18** (1985).

[193] R. Jullien and R. Botet, Surface thickness in the Eden model, Phys Rev Lett **54**, 2055 (1985).

[194] M. Kardar, G. Parisi, and Y.-C. Zhang, Dynamic scaling of growing interfaces, Phys Rev Lett **56**, 889 (1986).

[195] T. Halpin-Healy and Y.-C. Zhang, Kinetic roughening phenomena, stochastic growth, directed polymers and all that. Aspects of multidisciplinary statistical mechanics, Physics Reports **254**, 215 (1995).

[196] B. Derrida and R. Dickman, On the interface between two growing Eden clusters, J Phys A-Math Gen **24**, L191 (1991).

[197] M. Ausloos, N. Vandewalle, and R. Cloots, Magnetic Eden Model, Europhys Lett **24**, 629 (1993).

[198] M. Ausloos and N. Vandewalle, Growth models with internal competition, Acta Phys Pol B **27**, 737 (1996).

[199] Y. P. Pellegrini and R. Jullien, Roughening Transition and Percolation in Random Ballistic Deposition, Phys Rev Lett **64**, 1745 (1990).

[200] W. Wang and H. A. Cerdeira, Kinetic growth of randomlike and ballisticlike deposition models, Phys Rev E **47**, 3357 (1993).

[201] W. Wang and H. A. Cerdeira, Surface growth of two kinds of particle deposition models, Phys Rev E **52**, 6308 (1995).

[202] H. F. El-Nashar, W. Wang, and H. A. Cerdeira, Surface growth kinetics and morphological structural transition in a (2+ 1)-dimensional deposition model, J Phys Cond Matter **8**, 3271 (1996).

[203] H. F. El-Nashar and H. A. Cerdeira, Growth kinetics and morphology of a ballistic deposition model that incorporates surface diffusion for two species, Phys Rev E **60**, 1262 (1999).

[204] M. Kotrla and M. Predota, Interplay between kinetic roughening and phase ordering, Europhys Lett **39**, 251 (1997).

[205] M. Kotrla, F. Slanina, and M. Predota, Scaling in a two-component surface-growth model, Phys Rev B **58**, 10003 (1998).

[206] B. Drossel and M. Kardar, Interplay between phase ordering and roughening on growing films, Eur Phys J B **36**, 401 (2003).

[207] B. Drossel and M. Kardar, Phase ordering and roughening on growing films, Phys Rev Lett **85**, 614 (2000).

[208] F. D. A. A. Reis, Dynamic transition in deposition with a poisoning species, Phys Rev E **66**, 027101 (2002).

[209] A. Ali and S. Grosskinsky, Pattern formation through genetic drift at expanding population fronts, Adv. Complex Syst. **13**, 349 (2010).

Danksagung

Zu allererst will ich mich bei Prof. Erwin Frey bedanken. Ohne seine Offenheit für Ideen, ohne sein breites Wissen und seine Intuition, aber auch ohne seine Kritik wäre diese Arbeit in dieser Weise niemals möglich gewesen. Wenn es nicht recht weitergehen wollte, hatte er immer Zeit zu diskutieren, zu ermutigen und mir neue Ansatzpunkte zu zeigen. Nicht zuletzt konnte ich, dank seiner großzügigen Unterstützung, interessante Konferenzen besuchen und ein sehr angenehmes Leben führen.

Prof. Joachim R. O. Rädler gebührt im gleichen Maße Dank. Er begeisterte mich für die Biophysik, war immer interessiert an meiner Arbeit und ließ sich davon begeistern. Auch hatte ich die Möglichkeit, in seinem Labor selbst Experimente durchzuführen und verlor dadurch, während der theoretischen Beschreibung, die biologische Realität nicht aus den Augen.

Für finanzielle Unterstützung, Hilfe in Rat und Tat, sowie natürlich für die Organization wunderbarer informeller Veranstaltungen bedanke ich mich beim Internationalen Doktorandenkolleg NanoBioTechnologie, insbesondere bei Marilena Pinto und Marie-Christine Blüm.

Meinen Kollaborationspartnern Madeleine Leisner, Berenike Maier, Gerlinde Schwake, Simon Youssef, Sebastian Gude, Maria Pamela David und Eduardo Mendoza möchte ich hier ausdrücklich für die angenehme Zusammenarbeit und die, daraus erwachsenen, Resultate und Veröffentlichungen danken.

Meine Büromitbewohner Anna Melbinger und Jonas Cremer sind über die Jahre weit mehr als Kollegen geworden: Diskussionspartner im Bereich der Wissenschaft, der Politik und des sozialen Verhaltens, weiterhin Teilzeitpsychiater und Flurfunkflüsterer — mit einem Wort Freunde. Ich werde Euch vermissen.

Tatsächlich sind mir alle lieben Menschen am Lehrstuhl Frey ans Herz gewachsen. Durch Euch habe ich neue Themen kennengelernt, gut diskutiert und auch abseits der Physik viel Spaß gehabt. Ich hatte immer das Gefühl, dass jede(r) einzelne mir jederzeit helfen würde (wie es auch viele getan haben). Insbesondere haben das kürzlich Louis Reese, Brendan Osberg, Jonas Cremer, Anna Melbinger, Christoph Weber, Jan Heyder und Anton Winkler bewiesen, die sich als Korrekturleser dieser Arbeit beziehungsweise Lernhelfer für die mündliche Prüfung verdient gemacht haben.

An den Judge (a.k.a. Gold Servers), Jeff Goldboom, Ary Goldman und John Goldtrain, gehen, davon könnt ihr mal ausgehen, die dicksten Props. Für all dieses Zeuch und Kram und so, wisst schon.

Da „Mens sana in corpore sano“ für mich intuitiv Sinn macht, muss ich mich bei Professor Cabeludo und den anderen Capoeirista aus der Ballettschule bedanken. Die Gruppe ist, davon bin ich überzeugt, etwas ganz Besonderes, in ihrer Vielfalt und ihrem Zusammenhalt,

ihrer Stärke und ihrer Ungezwungenheit. Bei Euch konnte ich alle Dinge herauslassen, die in meinem Kopf der Fertigstellung dieser Arbeit im Weg standen — verzeiht mir, wenn ich machmal ruppig war. In diesem Zusammenhang sei auch Foguinha für alles gedankt, was ihrer Natur entspricht.

Auch meine Kletterpartner Birgit Reiserer und David Klingl (und einige andere) haben viel dabei geholfen den Blick von abstrakten, verzwickten Problemen zu lösen und lieber auf den nächsten Griff zu richten, denn das ist der einzige Weg nach oben.

Fast ganz am Ende, obwohl sie doch so wichtig ist, steht hier meine Familie. Dieter, Till, Elisabeth: Ohne Euch ging es nicht. Nichts kann Euch ersetzen, auf Euch ist immer Verlass. Meiner Mutter Hildegard danke ich für eine wunderbare Kindheit und mehr als ich hier schreiben kann. Gerne hätte ich mit Dir gefeiert und Dich hoffentlich stolz gemacht.

Vergessen habe ich nicht alle anderen, wenn auch namentlich unerwähnten, Freunde, die mir in den letzten, langen, kurzen fünf Jahren immer wieder viel Freude bereitet und geholfen haben. Danke an alle.