

Statistical Physics of Dynamical Networks and Morphogenesis

Dissertation
zur Erlangung des Doktorgrades
der Mathematisch-Naturwissenschaftlichen Fakultät
der Christian-Albrechts-Universität
zu Kiel

vorgelegt von
Thimo Rohlf

Kiel 2004

Referent: Prof. Dr. S. Bornholdt
Korreferent: Prof. Dr. H. G. Schuster
Tag der mündlichen Prüfung: 27. April 2004
Zum Druck genehmigt: Kiel, den 28. April 2004

gez. W. Depmeier (Dekan)

*Alle Gestalten sind ähnlich, und keine gleicht der
anderen. Und so deutet das Chor auf ein gemeinsames
Gesetz, ein heiliges Rätsel.*

*All shapes are similar, yet do not look alike. And so
everything points at a common law, a holy mystery.*

*Johann Wolfgang von Goethe (1798), Die
Metamorphose der Pflanzen (The Metamorphosis of
Plants)*

Research for this thesis was carried out between April 2001 and August 2002 at the Institute of Theoretical Physics and Astrophysics, Christian-Albrechts-University Kiel, and between August 2002 and March 2004 at the Interdisciplinary Center for Bioinformatics (IZBI), University of Leipzig, in partial fulfillment of the requirements for the Ph.D. degree of the faculty of mathematics and sciences at Christian-Albrechts-University Kiel.

I wish to express my gratitude to my advisor, Prof. Dr. Stefan Bornholdt, for his encouragement and support of this demanding project, and for inspiring and stimulating discussions, yet leaving me much freedom to plan and realize research projects independently. Special thanks also go to Prof. Dr. T. C. G. Bosch and his working group, for pointing at interesting questions in the context of *Hydra* development, providing the motivation for the part of this thesis dealing with morphogenesis and pattern formation.

I gratefully acknowledge a Ph.D. scholarship granted to me by the *Studienstiftung des deutschen Volkes* (German Merit Foundation) , giving me as well the opportunity to realize this project as the chance to participate in an inspiring Summer Academy in the nice surroundings of Rovinj, Croatia in autumn 2001. Particular thanks go to Prof. Dr. Dietrich Kurt Hoffmann and Dr. Sigur v. Boletzky, who provided an excellent introduction to developmental biology of marine organisms during that academy.

I would also like to thank the *Wilhelm und Else Heraeus-Stiftung* for financial support, Prof. Dr. Heinz Georg Schuster for the invitation to the Workshop *Concepts*

for Complex Adaptive Systems at the Hanse Institute of Advanced Study in spring 2002, and the Santa Fe Institute (New Mexico, USA) for the opportunity to participate in the *Complex Systems Summer School* in June 2002 and a generous travel grant.

Special thanks are due to the people at the Institute of Theoretical Physics in Kiel, who provided a stimulating and supportive working environment for the first part of this thesis. In particular, I am very grateful to Dr. Jens Christian Claussen, Dr. Holger Ebel, Dr. Christel Kamp, Norbert Lüthje, Lutz-Ingo Mielsch, Torsten Röhl, Prof. Dr. Heinz Georg Schuster, and Arne Traulsen. The second part of this thesis largely profited from the warm acceptance provided by my colleagues at the Interdisciplinary Center for Bioinformatics (IZBI) in Leipzig, I would like to thank - among many others - Gil Benkö, Dr. Dirk Drasdo, Claudia Fried, Sylvia Henger, Dr. Konstantin Klemm, Sonja Prohaska, Jörg Reichardt, Prof. Dr. Peter Stadler and Jens Steuck.

Last but not least I would like to thank my parents for their support and encouragement throughout this project.

March 2004

T. R.

Abstract

Complex networks of many interacting units occur in diverse areas as, for example, gene regulation, neural networks, economic interactions, and the organization of the internet. Many of these networks exhibit complex, non-Hamiltonian dynamics that strongly depends on network topology. In addition, ensembles of dynamical networks communicating with each other often show emergent behavior - for example, in development (morphogenesis) of multicellular organisms cell-cell communication can coordinate the dynamics of gene regulatory networks, leading to complex spatial patterns.

The purpose of this thesis is threefold: first, to gain insight into the interdependence between network dynamics and -topology; second, to explore how this interdependence, together with local rewiring events, can contribute to evolution of network topologies with properties as, e.g., observed for gene regulatory networks; third, to investigate how local interactions between coupled networks can lead to robust and reproducible emergence of spatial patterns and solve functional tasks in morphogenesis.

This train of thoughts is pursued in the following steps: (i) First, the critical connectivity of Random Threshold Networks is calculated analytically by combination of a new combinatorial approach to “damage spreading” with an annealed approximation. A non-trivial dependence of damage propagation on the in-degree of network nodes is identified that may have important consequences for the evolution of network topologies. (ii) Then, a model of topological network evolution is studied in the context of evolving gene regulatory networks. This model leads to evolution of network topologies close to criticality, however, with a degree-distribution and activity pattern that deviate from random networks, in good agreement with statistical data obtained for real gene networks. (iii) Thereafter, morphogenesis by coupled regulatory networks is studied. Starting from experimental observations in *Hydra*, a gene network model capable of *de novo pattern formation* and *regulation of pattern proportions* is developed, providing a network-based, alternative scenario to gradient-based explanations of these morphogenetic processes. (iv) Robustness of this model with respect to two type of perturbations often found in biological organisms, stochastic update errors and cell flow, is studied. It is shown that noise-induced control contributes to pattern stabilization, even under cell flow. A first order phase transition is found at vanishing noise, a second order phase transition at increased cell flow. (v) Finally, a two-dimensional extension of the morphogenetic model is studied. Here, a competition between initial symmetry breaking and neighborhood-dependent state changes of cells leads to sharp and localized boundaries of spatial activity domains, as required for robust self-organization of position information in a two-dimensional tissue.

Kurzfassung

Komplexe Netzwerke aus vielen wechselwirkenden Untereinheiten findet man in verschiedenen Bereichen, wie z.B. der Regulation von Genaktivität, neuronalen Netzen, ökonomischen Beziehungen, und der Organisation des Internets. Viele dieser Netzwerke zeigen komplexe, nicht-Hamiltonsche Dynamik, die stark von der Netzwerktopologie abhängt. In Ensembles von gekoppelten dynamischen Netzwerken findet man emergentes Verhalten - so kann z.B. die Kommunikation zwischen Zellen in der Entwicklung (Morphogenese) vielzelliger Organismen die Dynamik regulatorischer genetischer Netzwerke koordinieren, was zur Entstehung komplexer räumlicher Muster führt.

Die Zielstellung dieser Arbeit umfasst drei Bereiche: als erstes soll ein Einblick in die Abhängigkeiten zwischen Netzwerkdynamik und -topologie gewonnen werden, zweitens soll untersucht werden, wie diese Abhängigkeiten, zusammen mit lokalen Umverdrängungen, zur Evolution von Netzwerktopologien mit Eigenschaften, wie sie z.B. in Genregulationsnetzwerken beobachtet werden, beitragen können; drittens soll analysiert werden, wie lokale Wechselwirkungen zwischen gekoppelten Netzwerken zur robusten und reproduzierbaren Entstehung räumlicher Muster und zur Lösung funktioneller Aufgaben in der Morphogenese führen.

Dieser Gedankengang wird in folgenden Schritten entwickelt: (i) Als erstes wird die kritische Konnektivität zufälliger Schwellenwertnetzwerke mit Hilfe einer Kombination eines neuen, kombinatorischen Ansatzes mit einer Annealed-Näherung analytisch berechnet. Es wird eine nicht-triviale Abhängigkeit der Störungspropagation vom Knotengrad gefunden, die Auswirkungen auf die Evolution von Netzwerktopologien haben könnte. (ii) Danach wird ein Modell topologischer Netzwerkevolution untersucht, das zur Entstehung von Netzwerktopologien nahe dem kritischen Punkt führt, jedoch mit nicht-zufälligen Knotengradverteilungen und Aktivitätsmustern, in guter Übereinstimmung mit Daten für reale Genregulationsnetzwerke. (iii) Im nächsten Schritt wird Morphogenese durch gekoppelte regulatorische Netzwerke untersucht. Ausgehend von Beobachtungen bei *Hydra*, wird ein Netzwerkmodell entwickelt, das zur *de-novo Musterbildung* und zur *Regulation von Musterproportionen* imstande ist. (iv) Hiernach wird die Robustheit dieses Modells analysiert in Bezug auf zwei Arten von Störungen: stochastische Fluktuationen der Dynamik und Zellfluß. Als Ergebnis findet man eine Stabilisierung des räumlichen Musters durch rauschinduzierte Kontrolle, sogar unter Zellfluß. (v) Schließlich wird eine zweidimensionale Erweiterung untersucht, bei der Konkurrenz zwischen anfänglicher Symmetriebrechung und nachbarschaftsabhängiger Zelldifferenzierung zu scharfen und lokalisierten Grenzen zwischen Aktivitätsdomänen führt, wie es für die robuste Selbstorganisation von Positionsinformation in einem zweidimensionalen Gewebe erforderlich ist.

Contents

I	Overview	1
1	Introduction	2
1.1	Motivation	2
1.2	Outline	6
2	Biological background	8
2.1	Introduction	8
2.2	Basic mechanisms of animal morphogenesis	10
2.2.1	Introduction	10
2.2.2	Gene regulatory networks	10
2.2.3	Cell-cell communication and spatial pattern formation	12
2.2.4	High-level processes in morphogenesis	13
2.3	<i>Hydra</i> - developmental biology of a basic metazoan	14
3	Statistical Physics of Dynamical Networks	16
3.1	Network topology	16
3.2	Network dynamics	19
3.2.1	Introduction	19
3.2.2	Definitions	20
3.2.2.1	Boolean Networks	20
3.2.2.2	Threshold Networks	20
3.2.3	Global dynamics: transients and attractors	21
3.2.4	Criticality in RBN and RTN	22
3.2.5	Local measures of network dynamics	25
3.2.5.1	Average activity	26
3.2.5.2	Average correlation	26
3.2.5.3	Properties of the two measures in RTN	26

4	Statistical Physics of Cellular Automata (CA)	28
4.1	Introduction	28
4.2	Definitions	29
4.2.1	One-dimensional CA	29
4.2.2	Two-dimensional CA	30
4.3	Statistical and Computational Mechanics of Cellular Automata	31
5	Mathematical models of Morphogenesis	36
5.1	Reaction-diffusion models and morphogenetic gradients	36
5.1.1	Turing-type models as a paradigm of pattern formation	36
5.1.2	Example: Gierer-Meinhardt model of Hydra pattern formation	38
5.1.3	Limits of reaction-diffusion models	39
5.2	Chemical networks and pattern formation	41
5.3	Coupled regulatory networks and local information transfer	43
II	Dynamics and evolution of regulatory networks and morphogenesis: numerical and analytical studies	48
6	Criticality in Random Threshold Networks: Annealed Approximation and Beyond	49
6.1	Introduction	49
6.2	Random Threshold Networks	50
6.3	Damage Spreading in RTN: dependence on the in-degree k	51
6.4	Average probability for damage spreading	54
6.5	Calculation of K_c by Derrida's Annealed Approximation	56
6.6	Beyond the Annealed Approximation: Complexity in RTN	59
6.7	Discussion	61
7	Gene Regulatory Networks: A Discrete Model of Dynamics and Topological Evolution	64
7.1	Introduction	64
7.2	Evolution of network topologies based on local dynamical rules	66
7.2.1	The modeling framework: discrete dynamical networks	66
7.2.2	Definition of the threshold network	66
7.2.3	Network dynamics	66
7.2.4	An algorithm for evolving network topology	67
7.2.5	Results	69

7.3	Discussion	70
8	Morphogenesis by coupled regulatory networks	72
8.1	Introduction	72
8.2	Problem and Model	74
8.2.1	A test scenario: Hydra foot formation	75
8.2.2	One dimensional cellular automata: Definitions	76
8.2.3	Translation into spatially coupled Boolean networks	77
8.2.4	Translation into spatially coupled threshold networks	77
8.3	Results	78
8.3.1	Cellular Automata Model	78
8.3.2	Interaction topology of the minimal network	80
8.3.2.1	Boolean representation	80
8.3.2.2	Threshold network implementation	82
8.4	Discussion	83
9	Noise-induced control, phase transitions, and cell flow in pattern formation	86
9.1	Introduction	86
9.2	Dynamics under noise	87
9.3	Noise-induced pattern control under cell flow	93
9.4	Summary	94
10	A two-dimensional model of pattern formation	96
10.1	Introduction	96
10.2	Problem and model	97
10.3	Results	100
10.3.1	Stochastic time scale separation	100
10.3.2	Explicit time dependence of p_h	102
10.4	Summary	103
11	Conclusions and Outlook	104
A	Evolution of Cellular Automata by Genetic Algorithms (GA)	108
A.1	Definition of the GA	108
A.2	Evolution of cellular automata	109
B	Statistical analysis of CA solutions	112

CONTENTS	IV
Bibliography	115
Curriculum Vitae	132
Erklärung	133

Part I

Overview

Chapter 1

Introduction

1.1 Motivation

How is the intriguing complexity of biological organisms encoded in their genes? Almost fifty years after the discovery of the genetic code and the structure of DNA by Watson and Crick (1953), the first draft of the complete human genome sequence was obtained and led to the startling estimate, that humans have only between 25000 and 30000 genes (International Human Genome Sequencing Consortium, 2001; Claverie, 2001) - not even twice as many as the nematode worm *C. Elegans* (Reboul *et al.*, 2001). It seems hard to perceive how organs as complex as the human brain can be encoded by such a limited set of genes – what is the answer to this problem? For a long time, reductionist concepts as, for example, “one gene - one phenone” relationships or “selfish genes” (Dawkins, 1976) provided a sound working ground in molecular genetics, here, however, we are pointed at the limits of these ideas.

More recent experimental observations, that revealed an unexpected complexity of regulatory interactions in genomic systems, point at a possible solution. These regulatory interactions, mostly at the level of gene transcription, constitute complex *dynamical networks*, leading to a self-regulated, context-dependent, combinatorial retrieval of the information encoded in the genes. One can expect that new experimental techniques as, e.g., DNA microarrays, will provide even larger amounts of data in this context, hence, it seems time to think of new methods and theoretical concepts that may contribute to an understanding of the dynamics and evolution of *gene regulatory networks*, and how they solve complex functional tasks in living beings.

What can Statistical Physics contribute to this emerging “network perspective” of biological organisms? It seems quite natural to view molecular networks as stochastic many-particle systems (Sasai and Wolynes, 2003), however, with properties very dis-

tinct from the systems studied in classical Statistical Mechanics, mainly due to lack of a Hamiltonian in most cases. Furthermore, usually neither experimental data available today nor current theoretical concepts and computational capacities allow for detailed dynamical models.

Facing these current limitations to a “Statistical Mechanics of molecular networks”, different research strategies have been pursued that, in some sense, are complementary to each other: (i) detailed models of small sub-units or sub-networks, (ii) strongly simplified models of network dynamics, to gain insight into generic dynamical properties of networks of many interacting sub-units, (iii) toy-models for the evolution of network topologies and (iv) models for the solution of functional tasks by regulatory networks, e.g., spatial pattern formation during development of a multicellular organism.

In this thesis, we shall focus on strategies (ii) to (iv). In the following, we start with some considerations on dynamics and evolution of complex networks.

Kauffman (1969, 1993) introduced Random Boolean Networks (RBN) as simplified models for the dynamics of gene regulatory networks. RBN show complex non-Hamiltonian dynamics; whereas the dynamics of signal propagation, including a phase transition at a critical average connectivity in the limit of large system sizes (Derrida and Pomeau, 1986), is understood quite well, an understanding of global dynamical properties as, for example, the scaling of attractor periods (Albert and Barabási, 2000) or the number of different dynamical attractors (Samuelson and Troein, 2003) with network connectivity and system size, just starts to emerge. Random Threshold Networks (RTN) are a subset of RBN (Kürten, 1988b), but numerically much easier to investigate. However, analytical insight into the dynamics of RTN, concerning e.g. critical dynamical behavior, could be obtained only in some limiting cases (Kürten, 1988a). The first aim of this thesis is to contribute to an increased understanding of the dynamics of RTN, including more general network topologies. For this purpose, propagation of perturbations in RTN is studied by means of combinatorial methods, which allows for an analytical calculation of the critical mean connectivity K_c . It is found that propagation of perturbation depends on the in-degree of network nodes, which may have implications for the evolution of network topologies.

Evolution of complex networks has been studied in many different approaches. To explain the frequent observation of degree-distributions following power-laws in many social and technological networks, models based on preferential attachment (Barabási and Albert, 1999) have been suggested, whereas for similar observations in biological networks, e.g. protein-protein interactions or gene regulatory networks, models based on gene duplications may be a more favorable choice (Chung *et al.*, 2003). However, for the study of the evolution of regulatory networks, network dynamics, constituting the “phenotype” natural selection acts on, cannot be neglected any more. Only few

models of network evolution based on an interplay of network dynamics and selective “rewiring” of interactions between network nodes have been studied so far (Wagner, 1994; Bornholdt and Sneppen, 1998; Bornholdt and Rohlf, 2000). The second aim of this thesis is to provide a contribution to this discussion by a study of a simplified version of the model proposed by Bornholdt and Rohlf (2000), with a focus on self-organization of non-random topological and dynamical properties, and comparing the findings to statistical data available for gene regulatory networks of real organisms.

How does network dynamics affect the emergence of complexity at higher levels of organization, for example, during development of a multicellular organism? This fascinating question leads to the third strand of ideas followed in this thesis. In multicellular organisms, cells self-organize into complex social ensembles due to an interplay of diverse dynamical processes summarized in the term *morphogenesis*. Although all cells contain identical genetic information, cell differentiation into diverse cell types as well as their arrangement in complex spatial patterns are universal features of the development of multicellular organisms, and both can be explained by a *combinatorial use* (differential expression) of genetic information.

Until recently, however, most theoretical concepts that tried to explain these experimental observations adapted the rather “static” view common to the early genomic era. The typical picture in these approaches was the following: due to a Turing instability in initial conditions (Turing, 1952), chemical dynamics of a reaction-diffusion system containing at least two chemical species, an “activator” and an “inhibitor” with different diffusion constants, establishes a stationary pattern (e.g., a concentration gradient) of a hypothetical “morphogen”; genes then passively “respond” to this pattern by concentration-dependent activation or inhibition of transcription (Gierer and Meinhardt, 1972; Koch and Meinhardt, 1994). Gene regulation by morphogen gradients has been observed in several organisms, but in most cases these gradients do not originate from reaction-diffusion systems; furthermore, in such systems parameter fine-tuning is needed, including a non-trivial hierarchy (separation by orders of magnitude) of diffusion constants, which imposes severe constraints on applications to biological systems (Lander, Nie and Wan, 2002).

Recent biological research often has shifted the focus towards a “network perspective” of development, with complex (often short-range) cascades of signalling molecules coordinating regulatory dynamics of adjacent cells; genes no longer passively respond to “prepatterns” - regulatory networks actively take part in pattern formation processes. Theoretical concepts that adapt this picture of “coupled regulatory networks” now become available (e.g. Jackson, Johnson and Nash, 1986; Furusawa and Kaneko, 2000; Salazar-Ciudad, Garcia and Solé, 2000), and one may wonder if basic morphogenetic processes, as, for example, self-organization of *position information* (Wolpert, 1969),

that have been explained by gradient-based models so far, could arise as a consequence of this regulatory dynamics and local cell-cell communication.

As a first approach to this question, we start from two basic experimental observations made in the polyp *Hydra*: de-novo pattern formation from randomized cell aggregates (Technau *et al.*, 2000) and proportion regulation, i.e. scale independence, of (position-dependent) spatial gene expression patterns. Using the framework of cellular automata and coupled discrete threshold networks, a simple gene network model is developed that reproduces these experimental observations, with the only ingredient of (biased) local information transfer between cells, but without the need for parameter tuning as observed in many models based on reaction-diffusion systems.

Morphogenesis exhibits considerable robustness against perturbations acting on a large range of scales - hence, for the sake of biological plausibility, any theoretical model of the underlying dynamical processes must exhibit this feature, too. For this purpose, robustness of the proposed model with respect to two type of perturbations is studied: stochastic update errors and directed cell flow, as often found in biological organisms. It is shown that noise-induced control, based on soliton-like excitations (“particles”) that travel through the system, contributes to pattern stabilization. A first order phase transition is found at vanishing noise, a second order phase transition at increased cell flow.

Finally, the issue of *dimensionality* in morphogenesis is approached by extending the proposed model into two dimensions.

1.2 Outline

Let us now briefly summarize the train of thoughts pursued in this thesis.

In the remaining chapters of part I, a brief summary of the biological background is given (chapter 2), followed by surveys of the Statistical Physics of discrete dynamical networks (chapter 3) and of cellular automata (chapter 4), providing the methodological framework for subsequent studies on dynamics and evolution of complex networks and morphogenesis by coupled regulatory networks. In chapter 5, a survey of theoretical concepts and mathematical models of morphogenesis is provided, paving the way from older diffusion-based models to recent approaches based on dynamics of regulatory networks.

In part II of this thesis, the concepts introduced in part I are applied and advanced in the following numerical and analytical studies:

- In **chapter 6**, the critical connectivity of random threshold networks is calculated analytically by combination of a new combinatorial approach to “damage spreading” with an annealed approximation. A non-trivial dependence of damage propagation on the in-degree of network nodes is identified; this “damage suppression” at network hubs may have important consequences for the evolution of network topologies.
- In **chapter 7**, a simplified version of the model of topological network evolution introduced by Bornholdt and Rohlf (2000) is studied in the context of evolving gene regulatory networks. It is shown that this model not only leads to evolution of network topologies close to criticality, but also to a degree-distribution and activity pattern that deviates from random networks. These findings are compared to recent statistical data obtained for real gene networks.
- In **chapter 8**, morphogenesis by coupled regulatory networks is studied. Starting from experimental observations in *Hydra*, in three steps a gene network model capable of *de novo pattern formation* and *regulation of pattern proportions* is developed, providing a network-based, alternative scenario to gradient-based explanations of these morphogenetic processes.
- In **chapter 9**, robustness of the model introduced in chapter 8 in an implementation as a stochastic cellular automaton is studied. It is shown that the model is robust with respect to two type of perturbations: stochastic update errors and directed cell flow, as often found in biological organisms. Noise-induced control contributes to pattern stabilization. A first order phase transition is found at vanishing noise, a second order phase transition at increased cell flow.

-
- In **chapter 10**, a two-dimensional extension of the morphogenetic model is studied. It is shown that a competition between initial symmetry breaking and neighborhood-dependent state changes of cells leads to sharp and localized boundaries of spatial activity domains in 2D. Interference between both dynamical processes is studied and the findings are compared to experimental observations.
 - In **chapter 11**, a conclusion and an outlook are given.
 - In the **appendix**, the genetic algorithm that has been applied for the development of the model introduced in chapter 8 is explained. Statistical properties of different solutions are compared and possible implications for the evolution of similar mechanisms in biological organisms are discussed.

Chapter 2

Biological background

2.1 Introduction

Depending on its level (or complexity) of organization, life on earth can be divided into several big "kingdoms". The main divisions are:

- Prokaryotes (archaebacteria and bacteria)
- Eukaryotes (protista, algae, plants, animals and fungi (mushrooms)),

where only the latter possess a *cell nucleus* and organelles, as for example mitochondria and ribosomes. Eukaryotes themselves show two different levels of complexity: *unicellular organization* or *multicellular organization*. Algae, plants, animals and fungi have developed multicellular strains during evolution, however, plants and, in particular, animals have developed multicellular structures much more complex than those found in algae and fungi.

In particular, the occurrence of multicellular animals (metazoa) provides one of the great puzzles in evolutionary biology: there is some evidence in the fossil record that all animal phyla appeared during a very short period of evolution, apparently at the beginning of the Cambrium around 550 million years ago; in literature, this period of evolution is mostly quoted as "Cambrian explosion" (Solé, Salazar-Ciudad and Garcia-Fernandez, 2000; Arthur, 1997, pp.3 seq.). However, there is an ongoing debate about the precise timing, concerning both the starting point and the duration of the evolution of the first phyla of multi-cellular animals (Arthur, 1997). In any case, compared to the long period of time evolution of life remained at the unicellular level, multicellular organization obviously developed on a short time scale. Maynard Smith and Szathmáry (1995) argue that three basic problems had to be solved for development of a multicellular organism, with differentiated cells:

1. *Gene regulation.* In multicellular organisms – with few exceptions – all genes are present in all tissues, and differentiation of cells arises because different sets of genes are active. How are different genes switched on and off in different cells?
2. *Cell heredity, and the dual inheritance system.* Not only are different genes switched on in different cells, but the states of differentiation are heritable through cell division. There is a "dual inheritance system": The primary system, depending on the nucleotide sequence of the DNA, determines resemblance between parents and offspring, the secondary system requires the transmission, through cell division, of states of gene activity. How does the secondary system work?
3. *Spatial patterns.* Differentiated cells are arranged in space in a specific and repeatable pattern. How does this come about?

Although some arguments also seem to speak in favor of external triggers of the Cambrian explosion, e.g., a certain level of oxygen concentration in the atmosphere and in the oceans had to be reached before multicellular organization could "pay off" (Derry *et al.*, 1992), most biologists agree that the three problems formulated above form the "central part" of the puzzle (Solé, Salazar-Ciudad and Garcia-Fernandez, 2000). The probably most intriguing aspect of the three problems is that they are not isolated from each other, but have strong interdependencies: cell heredity relies on gene regulation, spatial pattern formation on cell heredity, and gene regulation is influenced by spatial pattern formation, as, for example, cell rearrangements and intercellular communication. It has been argued by theoretical biologists, that - because of these strong interdependencies - Darwinian evolution, i.e. the interplay of mutations and selection, does not suffice to explain the appearance of multicellularity, and that, in addition, there exist fundamental principles of self-organization, which - once they had developed - allowed for fast evolutionary "jumps" towards higher levels of complexity (Kauffman, 1993; Solé, Salazar-Ciudad and Garcia-Fernandez, 2000), in contrast to the gradual character of evolutionary changes as postulated by Darwin (1859).

The intention of this thesis is by no means to find a solution of this deep puzzle, but rather to contribute to an increased theoretical understanding of the first and the third problem formulated above, by definition and study of simple modelling frameworks that can be analyzed with methods of Statistical Physics. However, before these theoretical frameworks are developed (see chapters. 3–5 and part II of this thesis), the basic mechanisms underlying development, or, equivalently, *morphogenesis* of multicellular organisms are discussed in more detail in Sec. 2.2, and in Sec. 2.3, a specific example of animal development, the fresh water polyp *Hydra*, is introduced.

2.2 Basic mechanisms of animal morphogenesis

2.2.1 Introduction

In this section, a short review of developmental mechanisms common to all multicellular animals (metazoa) is given; for a more detailed description, see e.g. (Alberts *et al.*, 2002, pp. 1157 seq.).

Most animals and plants start their life as a single cell - a fertilized egg. During morphogenesis, this cell divides many times to produce many different cells in a final pattern of astonishing complexity and precision. How does the genome determine this pattern?

The genome is normally identical in every cell. Cells do not differ because they contain different genetic information, but because they *express different sets of genes*. There are four essential processes by which the embryo is constructed, that are under control of selective gene expression:

1. *cell proliferation*, producing many cells from one
2. *cell spezialization*, creating cells with different characteristics at different positions
3. *cell interactions*, coordinating the behavior of one cell with that of its neighbors
4. *cell movement*, rearranging the cells to form structured tissues and organs.

During morphogenesis, all these processes are happening *at once*. Indeed, development of multicellular organisms is a paradigmatic example of a *spatially extended decentralized system*: there is no central instance that coordinates development, rather each single cell has to make its own decisions, according to its own copy of the genetic program and the circumstances it meets during its developmental path. Cell differentiation relies on the remarkable fact that cells have a *memory*: the genes a cell expresses depend the cell's past as well as its present environment. The molecular machinery leading to this astonishing feature of context-dependent information processing in cells will be sketched in the next section, however, we will restrict ourselves to gene regulation at *transcription level*.

2.2.2 Gene regulatory networks

Cell memory as well as differential expression of genes mainly relies on the activity of *gene regulatory networks* at transcription level (Arnone and Davidson, 1997; Davidson, 2001): transcription factors, i.e. proteins produced by regulatory genes, can bind to

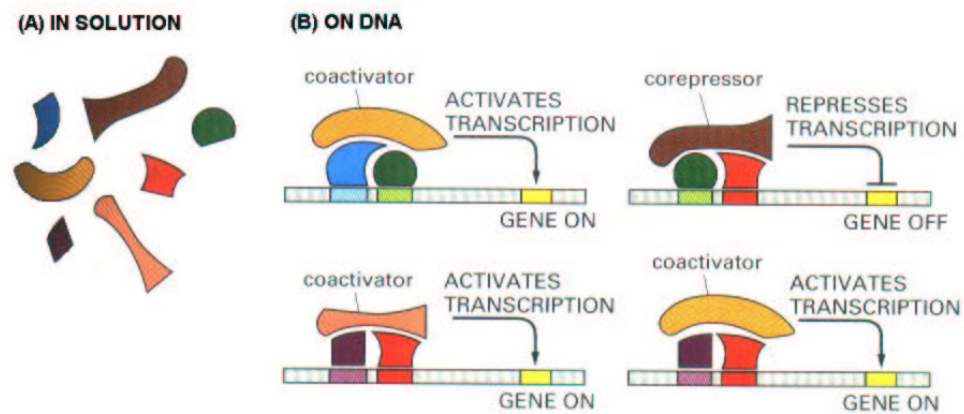


Figure 2.1: Eukaryotic gene regulatory proteins, that can act as transcription activators or repressors, often assemble into complexes on DNA. In panel (A), seven different regulatory proteins in solution are shown, that assemble into various complexes on DNA and thereby initiate or inhibit transcription of some genes (B). In this way, regulatory interactions can implement complex logical networks (after Alberts *et al.*, 2002).

certain regions of the DNA and either activate or repress transcription of a (down-stream) gene. In eucaryotes, regulatory proteins can assemble into complexes on the DNA (cf. Fig. 2.1), meaning that they in principle can implement networks with complex logical rules for gene expression control, as envisioned by Kauffman (1969, 1993). In such circuits, memory can be created, e.g., by a positive feedback loop (Alberts *et al.*, 2002, pp. 419 seq.), as sketched in Fig. 2.2. Several gene regulatory proteins that are involved in establishing the *Drosophila* body plan, for example, stimulate their own expression, thereby creating a positive feedback loop (Lawrence, 1992; Alberts *et al.*, 2002). By subsequent activation of different regulatory circuits, cells can be driven to different "fates" and different *cell lineages* are created in the course of development. Interestingly, in many cases the direction and composition of this sequence cannot be changed - it is said that such cells are *determined* for their fate (Alberts *et al.*, 2002, p. 1163); this introduces the notion of *irreversibility* to many developmental pathways (Kaneko and Furusawa, 1999). Concerning the global structure of regulatory networks, for most organisms the molecular data available today do not suffice to make conclusive predictions, however, for some organisms preliminary results have been obtained by means of statistical methods, raising the question for the evolutionary processes that could lead to networks with the observed properties (cf. chapter 7).

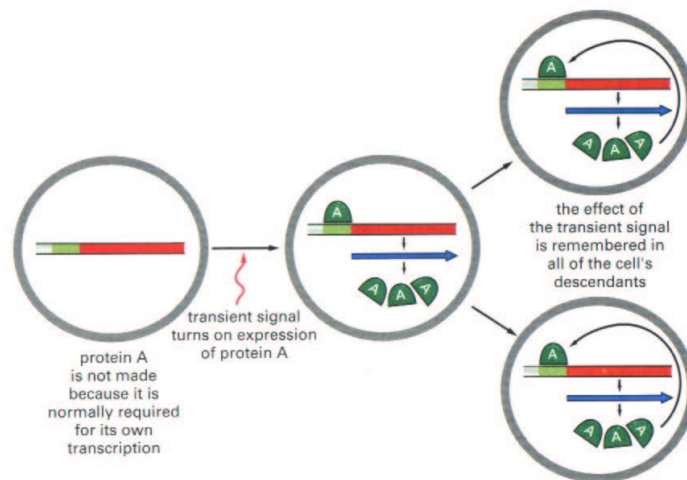


Figure 2.2: Creation of cell differentiation and cell memory by autoregulatory feedback. The gene coding for protein A normally is not transcribed, but a transient signal during development turns on its expression. Because protein A autoregulates its own expression via a positive feedback loop, the effect of the transient signal is remembered in all of the cell's descendants (after Alberts *et al.*, 2002).

2.2.3 Cell-cell communication and spatial pattern formation

For creating a "body plan", cells not only have to differentiate, but also have to do so in a *spatially ordered manner* - this requires *cell interactions* for establishing "position information" (a notion introduced by Wolpert, 1969). Although sister cells also may become different via, e.g., asymmetric divisions (this is, for example, observed in the Sea Urchin embryo, Hörstadius, 1935, 1973), mainly two mechanisms are known that can lead to self-organization of position information:

- inductive interactions
- morphogen gradients,

where morphogens, yet, may be regarded as "long range inducers" (Alberts *et al.*, 2002, p. 1166). In the strict sense, however, induction means *short range* signals exerted by a cell on its neighbor cells. Quite often, such signals are transmitted via direct cell-cell-contacts, which are mediated through some well-established signalling pathways, using receptor-ligand binding; examples of such signalling pathways are the TGF β superfamily (using TGF β as ligand and TGF β receptors), the Wnt (wingless) pathway (using Wnt as ligand and Frizzled as receptor), and the Delta-notch pathway, using

Delta as ligand and Notch as receptor (Alberts *et al.*, 2002, p. 1167). Morphogens, on the contrary, are signalling molecules that can *diffuse*, often from a localized source and thereby creating a concentration gradient; genes are assumed to be activated in a concentration-dependent manner, leading to position-dependent differentiation of cells. Morphogen gradients are well-established for several organisms, however, the mechanism of diffusion in a cellular environment itself is discussed in a controversial way in recent research (for a review, cf. Sec. 5.1.3). Interestingly, pattern formation via inductive interactions in development has been neglected by theoreticians until recently, whereas much research effort has been invested into the study of reaction-diffusion systems (cf. chapter 5). Indeed, in many cases, one may wonder if "position information" formerly attributed to morphogen gradients could arise from induction-like interactions instead. In the case of the Sea Urchin, for example, for some results of grafting experiments that could be interpreted by morphogene gradients in an intriguingly elegant way (Hörstadius, 1935, 1973) it finally turned out, that they had to be attributed to the regulative capacities of certain inductive signalling cascades (Ransick and Davidson, 1993).

2.2.4 High-level processes in morphogenesis

Cell movements and programmed cell death (apoptosis) are "high level mechanisms" of morphogenesis that are ubiquitous in animal development. Cell movements may arise either by "passive" mechanisms, e.g., via cell proliferations or cell-cell adhesion, or by "active" mechanisms as, e.g., chemotaxis. The biophysics of cell movements due to cell-cell adhesion is well understood today (Graner and Glazier, 1992; Ruoslahti, 1997; Steinberg and Takeichi, 1994) and it has been shown that physical constraints, as, for example, minimization of surface tension, can lead to movements of whole cell ensembles (Odell and Oster, 1981) that can be coordinated by genetic control (Drasdo and Forgacs, 2000). Hence, it might be speculated that the spectacular, coordinated movements of tissues – e.g. during gastrulation – are the result of a delicate interplay between "generic" and genetic mechanisms (Hogeweg, 2000). A detailed discussion of this fascinating field of research is beyond the scope of this thesis, in chapter 9, however, a very simple form of cell movement - directed cell flow due to cell proliferations - is discussed in the framework of a model of spatial pattern formation.

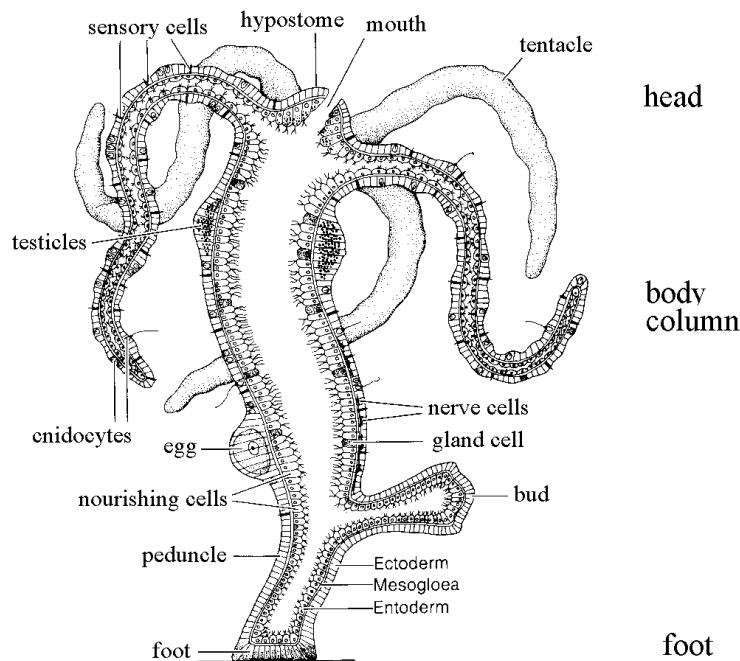


Figure 2.3: Schematic illustration of the *Hydra* body plan with its three distinct regions (head, body column and foot) along the body axis (adapted from Storch and Welsch, 1996).

2.3 *Hydra* - developmental biology of a basic metazoan

In the following, we will shortly review some basic aspects of the morphogenesis of a classical “model organism” in developmental biology, the fresh water polyp *Hydra*. Some of the observations made for *Hydra* development are taken as a starting point for the development of a model for pattern formation in chapter 8 of this thesis.

Hydra belongs to the phylum *Cnidaria*, a very basal group of multicellular animals (metazoa). Cnidaria are mostly marine organisms with radial symmetry and diploblastic organization, i.e., two layers of tissue - entoderm and exoderm - surround a hollow interior, the gastro-vascular space. The *body plan* is characterized by one central (basal-apical) body axis. *Hydra* shows the typical, simple body plan of a polyp with radial symmetry, with three distinct body regions along the symmetry axis: head, body and foot. The head is characterized by a mouth (hypostome) surrounded by tentacula, and the foot serves for the animal’s adherence to a substrate. *Hydra* reproduces asexually by *budding*. Buds typically arise in a localized “budding region” that roughly divides the body-axis in the ratio 1/2 (Fig. 2.3).

There are mainly two aspects of *Hydra* development that distinguish it from most other animals, and make it a very interesting model system as well for Developmental Biology as for Statistical Physics: *Hydra* has an astonishing capacity for *regeneration*, and the adult animal is in a dynamic (yet, stationary) state of *continuous growth and pattern formation* (*Hydra* as “eternal embryo”). *Hydra* has the ability of reboot-like “de-novo pattern formation”. After dissociation of its cells in a centrifuge, the animal can regenerate within five days from a small random aggregate of cells (Gierer *et al.*, 1972; Technau *et al.*, 2000). During this time, self-organization of a well-defined head-foot axis takes place, with establishment of the corresponding spatial gene expression patterns. Gierer and Meinhardt (1972) gave a possible physico-chemical explanation of these observations in terms of a Turing instability (Turing, 1952) in a reaction-diffusion system, however, recent careful experimental observations of axis formation have casted doubts on this explanation (Fütterer *et al.*, 2003; Ott *et al.*, 2004). The cells of *Hydra* epithelia proliferate during the animals whole lifetime, there is a steady *cell flow* from the proliferation zone in the middle of the body column towards head and foot (David and Campbell, 1972). To compensate for this dynamic tissue flow, the molecular machinery that control spatial pattern formation has to be constantly active in the adult *Hydra*.

Similar to higher animals (e.g., *Bilateria*), *Hydra* has a “organizer region”, the *hypostome* (cf. Fig. 2.3). In this region, many genes are expressed that are particularly important for pattern formation. In most cases, expression patterns of these genes are regionally confined and show sharp boundaries, an observation also found for other genes in other body regions; the nature of the mechanisms that lead to such sharp and confined patterns is unknown. However, the identification of a particularly large number of transcription factor binding sites at the head gene KS1 (Endl, Lohmann and Bosch, 1999), for example, as well as a number of taxon-specific genes found involved in *Hydra* development (Bosch and Khalturin, 2002), point at a complex gene regulatory network leading to self-organization of these expression patterns. For the genes Pedibin and Cn-NK2, that control pattern formation in the foot region, such a regulatory circuit has been identified (Thomsen, 2001; Thomsen *et al.*, 2003), which is presumably part of a much larger regulatory network.

Chapter 3

Statistical Physics of Dynamical Networks

3.1 Network topology

The mathematical discipline of *graph theory* provides the framework for a quantitative description of network topologies, (see e.g. West, 2001). In the following, we will summarize some basic concepts of graph theory, followed by some results for classical random graphs and their relation to statistical physics. Finally, some basic properties of network topologies as they are observed in natural, technological and social networks are discussed.

A **graph** is a triple $G = (V, E, f)$ consisting of a set of vertices V , a set of edges E , and a relation f that maps each edge to two vertices called its endpoints. Equivalent and more frequently used in the physical literature is the notion of a **network** of nodes (sites) and links. A **subgraph**, sometimes called cluster or subnetwork, is a graph J with $V(J) \subseteq V(G)$, $E(J) \subseteq E(G)$ and the same relation f for the assignment of links to endpoints as in G . A **path** is a simple graph (i.e., a graph with neither multiple links between a pair of nodes nor links from a node to itself) whose nodes can be ordered so that two nodes are adjacent if and only if they are consecutive in the list. A network is called **connected** if each pair of nodes belongs to a path. The **degree** k of a node is the number of links it is endpoint to. In Statistical Physics, the notion of connectivity is very often used as a synonym of degree. A convenient and important statistical measure for the characterization of network topologies is the degree distribution, defined as the probability $P(k)$ of the occurrence of a node with degree k in a given network. Furthermore, a network can be either unipartite or multipartite, meaning that it is composed of one or several kinds of nodes. It can

be **directed**, which means that the links have a direction pointing from one node to another, or undirected. For the study of dynamical networks, i.e. networks which are characterized not only by their topology, but also by some kind of dynamics taking place on them, in most cases directed, unipartite networks are considered, just as will be done in the remainder of this work.

In most complex networks in nature, a certain degree of randomness is found, which means that their topology cannot be mapped on some regular, low dimensional geometry. **Random graphs** can serve as a first approximation to stochasticity in networks (Erdős and Rényi, 1960).

In random graphs, links are assigned randomly to pairs of nodes with a constant probability p . Hence, the number of links L of a classical random network with N nodes is

$$L = \frac{1}{2} p N(N - 1), \quad (3.1)$$

leading to a mean degree of

$$\bar{K} = p (N - 1). \quad (3.2)$$

The random assignment of links translates into a binomial degree distribution,

$$P(k) = \binom{N - 1}{k} p^k (1 - p)^{N - 1 - k}. \quad (3.3)$$

If the number of nodes N is large and p is small, $P(k)$ can be approximated by a Poissonian:

$$P(k) = \frac{\bar{K}^k}{k!} e^{-\bar{K}}. \quad (3.4)$$

Most real-world networks, however, cannot be characterized by random graphs, as it turned out; concerning, for example, the degree distribution, usually clear deviations from a Poissonian are found. In particular, the degree distributions of many real-world networks approximately follow power laws (Barabási, 2002; Albert and Barabási, 2002):

$$P(k) \propto k^{-\gamma}. \quad (3.5)$$

Notice that such networks can be *directed*, in this case in-degree distribution and out-degree distribution may differ, e.g., they may follow power-laws with different exponents, or one of them may follow no power-law at all, but, e.g., an exponential distribution instead. Barabási and Albert (1999) established the notion of “scale-free” networks, in analogy to fractals, phase transitions, and other situations where power-laws are observed and no characteristic scale can be found. Since 1999, a large amount of research in the Statistical Physics community has been devoted to scale-free networks (for recent reviews cf. Albert and Barabási, 2002; Dorogovtsev and Mendes, 2002).

Scale-free network topologies have been observed, e.g., for the internet (Faloutsos *et al.*, 1999), the World Wide Web on the site level (Huberman and Adamic, 1999), citation networks of scientific papers (Redner, 1998), E-mail networks (Ebel *et al.*, 2002) and many other technological and social networks. For biological networks, the quality of databases in most cases does not suffice to make clear predictions, however, in some cases degree distributions roughly following a power-law have been identified, e.g., for the network of protein interactions in yeast (Kleinberg *et al.*, 1999; Dewey and Galas, 2001; Jeong *et al.*, 2001), for the metabolic network in *E. Coli* (Barabási and Albert, 1999; Jeong *et al.*, 2000), the yeast transcriptional regulatory network (Guelzim *et al.*, 2002) and functional interactions between genes (Snel, 2002). However, the in-degree distribution of several gene regulatory networks can be fitted much better with an exponential (Thieffry *et al.*, 1998; Guelzim *et al.*, 2002; Rohlf and Bornholdt, 2004a), which, on the other hand, is still an obvious deviation from the Poissonian, as it would be expected for a random graph.

A survey of existing results on the scaling exponents observed in large networks reveals a striking difference between biological (molecular and regulatory) networks and non-biological (social, technological and economic) networks: whereas the former mostly have scaling exponents in the range $1 < \gamma < 2$, the latter have γ values between 2 and 4 (Chung *et al.*, 2003). This leads to the assumption that the basic mechanisms that drive the *evolution* of network topologies are fundamentally different for both classes.

Besides the frequent observation of scale-free degree distributions, there exist a number of other non-trivial characteristics that are shared by many real-world networks. One of these is the so called *small world property* (Milgram, 1967), which, in mathematical terms, is defined as a network with (i) a clustering coefficient C much higher than expected for a random network of identical size (i.e., with identical numbers of nodes and links) and (ii) a small average length of the shortest path between two nodes ℓ (sometimes called **diameter** of the network) scaling logarithmically with network size (Watts and Strogatz, 1998; Amaral *et al.*, 2000). For the clustering coefficient C , there exist different definitions in literature, however, most frequently used is the one considered by (Barrat and Weigt, 2000; Newman *et al.*, 2001) in terms of the so-called triangle clustering coefficient based on the notion of “triples” in graphs. A triple is a connected subgraph containing exactly three nodes (or, equivalently, a pair of adjacent links), whereas a fully connected triple is a triangle of links. Then, the triangle clustering coefficient is

$$C_{\Delta} = \frac{3 \times (\text{number of fully connected triples})}{\text{number of triples}}. \quad (3.6)$$

The small world property has been observed not only in diverse social and technological networks (for a review cf., for example, Ebel, 2002), but also in biological networks as, for example, in the neural network of *C. Elegans* (Watts and Strogatz, 1998), and in the transcriptional regulatory network of yeast (Guelzim *et al.*, 2002). Whereas plausible models for the evolution of “small worlds” in social networks exist (Davidsen *et al.*, 2002), the evolutionary processes leading to small-world properties in biological networks are in most cases unknown.

A property that is very common in gene regulatory networks is their *modular* and *hierarchical* organization (Arnone and Davidson, 1997), meaning that they often decay into functional subnetworks (modules) with only sparse interconnections between them, and that these “modules” sometimes are controlled by a few “master genes” in a hierarchical manner. In the context of morphogenesis, such a hierarchical network will be discussed in chapter 8 of this thesis.

3.2 Network dynamics

3.2.1 Introduction

How can one define *dynamics* taking place on complex networks in a suitable way? Discrete Dynamical Networks (DDN) are networks, where the dynamics of the single elements (nodes) is *discrete in time and states*. DDN can serve as a first approximation to the dynamics of many natural networks. For example, dynamics of neurons, the elements that constitute *neural networks*, can be approximated by two different states, “silent” and “firing” (Schuster, 2001, p. 127); consequently, McCulloch and Pitts (1943) proposed a binary description of neural activity. For gene regulatory networks the situation is not that clear, however, there is empirical evidence that strong cooperative effects of transcriptional activation by individual transcription factors are mainly responsible for strong transcriptional activation (repression) of a target gene (Carey *et al.*, 1990; Wright and Gustafsson, 1991; Oliviero and Struhl, 1991); hence it has been argued that such transcriptional regulation can be approximated as a sum of weighted (regulatory) inputs with gene activity normalized to a certain interval, e.g. $[0, 1]$ (Wagner, 1994); here, 0 represents maximum repression and 1 maximum activation. It can be shown that restriction to a discrete state space does not change the essential dynamics in these models, consequently prototypes of such networks, namely Boolean (logical) networks (Kauffman, 1969, 1993) and threshold networks with discrete state space (Wagner, 1994; Bornholdt and Sneppen, 1998, 2000; Bornholdt and Rohlf, 2000) have been considered as simple theoretical models of gene regulatory networks.

In this thesis, we will follow this philosophy and consider only Random Boolean Networks and Random Threshold Networks. In the following, some basic definitions will be given.

3.2.2 Definitions

3.2.2.1 Boolean Networks

A **Boolean Network** is a directed network consisting of N binary sites (automata) with states $\sigma_i \in \{0, 1\}$. The sites interact with each other via logical (Boolean) rules and the states evolve discretely in time. The state of site i at time $t + 1$ is determined by the inputs it receives from k other sites, $\sigma_{j_i^m}(t)$ ($m = 1, 2, \dots, k$), at time t and the transition function f_i associated with the site i :

$$\sigma_i(t + 1) = f_i(\sigma_{j_i^1}(t), \sigma_{j_i^2}(t), \dots, \sigma_{j_i^k}(t)) \quad (3.7)$$

In *Random Boolean Networks (RBN)* the logical rules f_i are chosen at random from among all the 2^{2^k} possible Boolean rules of k inputs. Each of the k inputs is also chosen at random from among the N sites in the network.

3.2.2.2 Threshold Networks

A **Threshold network** is a directed network consisting of N binary sites (automata) with states $\sigma_i \in \{-1, 1\}$. For each site i , its state at time $t + 1$ is a function of the inputs it receives from other elements at time t :

$$\sigma_i(t + 1) = \text{sgn}(f_i(t)) \quad (3.8)$$

with

$$f_i(t) = \sum_{j=1}^N c_{ij} \sigma_j(t) + h. \quad (3.9)$$

The threshold parameter h is a real number. In *Random Threshold Networks*, each node receives its inputs from a random set of other nodes, and the interaction weights c_{ij} are sampled randomly from a given distribution $\rho(c_{ij})$. In the following we will always consider the discrete distribution

$$\rho(c_{ij}) = \{-1, 0, 1\}, \quad (3.10)$$

i.e. the interaction weights take discrete values $c_{ij} = \pm 1$, with $c_{ij} = 0$ if site i does not receive any input from element j .

There are close formal relationships between RBN and RTN; Kürten points out that - from an analytical point of view - it is plausible that a finite sequence of Boolean operations can be defined in terms of a neural network with weighted interactions and suitable thresholds (Kürten, 1988a). RTN can be considered as a subset of RBN, with similar dynamical properties, but numerically much easier to investigate.

3.2.3 Global dynamics: transients and attractors

Because - contrary to "classical" neural networks - RBN and RTN have *asymmetric* interconnections c_{ij} , they exhibit complex non-Hamiltonian dynamics (Kauffman, 1969, 1993; Kürten, 1988a; Bastolla and Parisi, 1996). The dimension of state space of a binary network of N elements is 2^N , thus after a maximum of 2^N time steps a previously occurred system state $\vec{\sigma} = (\sigma_1, \sigma_2, \dots, \sigma_N)$ must re-occur and the dynamics becomes trapped in a dynamical *attractor* or *limit cycle* of $L \leq 2^N$ ever-repeating states, defining the period L of the attractor. Starting dynamical runs with a random initial state vector $\vec{\sigma}_{ini}$, there follow always a number T of transient states the systems flushes through before a dynamical attractor is reached, defining the *transient* of length T . Usually a random network with an average coordination number $\bar{K} \geq 1$ has a number of different dynamical attractors, mostly with different periods. The median period length strongly depends on the average connectivity \bar{K} as we will see below. Each attractor has an entire set of transients leading to it from different initial state vectors, defining the *basin of attraction* of this limit cycle. Even for small networks (e.g. $N = 32$) the state space is very huge (here $2^{32} \approx 10^{10}$), making it hard to identify all attractors with their entire basins of attraction (Wuensche, 1998).

By identification of dynamical attractors with "cell types", Kauffman (1969) argued that RBN with a mean connectivity between 2 and 3 could serve as models of *gene regulatory networks*. In this parameter range, ensemble statistics gained in numerical simulations seemed to prove that the number of different attractors N_A scales with the square root of the network size N , similar to the observed number of different cell types in multicellular organisms as a function of the (estimated) number of genes. Later on, Kauffman refined this hypothesis and claimed that gene regulatory networks share properties of RBN near *criticality*, that will be discussed in the next section. However, recent research has rendered this postulate at least in part obsolete. Bilke and Sjunnesson (2001) observed, in an enumeration statistics of attractor periods, a scaling behavior $N_A \propto N$. Socolar and Kauffman (2003) stated that N_A presumably grows faster than linear at criticality, however, they argued, the true asymptotic behavior could only be observed for $N > 10^6$. Finally, Samuelson and Troein (2003) showed by an analytic calculation, that the number of attractors in Kauffman networks near

criticality grows faster than any power law, hence proving that the previous observations were artefacts of *biased undersampling* of state space in numerical simulations, i.e., because of the finite number of initial conditions that had been tested, only the attractors with the largest basins of attraction had been found.

3.2.4 Criticality in RBN and RTN

The dynamical behavior of RBN and RTN depends strongly on the average coordination number \bar{K} : sparse connections produce ordered nets, whereas high average coordination numbers lead to chaotic dynamics, i.e. transients and attractor periods diverging exponentially with system size (Kauffman, 1969, 1993; Kürten, 1988b). But the dynamics of RBN also depends on the way the Boolean functions are chosen: some of the rules are chaotic (XOR and EQUIVALENCE) in the sense that the dynamics exclusively due to them produces enormous periods while two others (CONTRADICTION and TAUTOLOGY) produce a constant output of either 1 or 0. Keeping a given connectivity \bar{K} fixed, a bias in the rules chosen can drive a network from ordered to chaotic behavior or vice versa. In RTN the threshold parameter h plays this role: a suitable increase of h forces networks into the ordered regime even at high connectivities (Kürten, 1988a,b).

In the following, let us consider a network where each node has the same number K of inputs and let us model a possible bias in the choice of the Boolean functions by the demand, that they are still chosen at random among all possible rules, but each of them produces a *biased output*: 0 with probability p and 1 with probability $1 - p$. One finds a transition from ordered to chaotic dynamics critically dependent on K and p ; in the limit $N \rightarrow \infty$ the position of the so-called *critical line* in the parameter space spanned by K and p is given by the equation (Flyvbjerg, 1988):

$$2p_c(1 - p_c) = \frac{1}{K} \quad (3.11)$$

The dynamical behavior of networks near the critical line is very complex, showing power-law distributions of various dynamical quantities (Bhattacharjya and Liang, 1996). Special interest was devoted to critical RBN with $p = 1/2$ and $K = K_c = 2$, which were introduced by Kauffman to model dynamical properties of genetic regulatory networks Kauffman (1969). In the literature, these networks with $p = 1/2$ are referred to as "Kauffman nets" (Derrida and Stauffer, 1986; Samuelson and Troein, 2003; Schuster, 2001).

Derrida and Pomeau (1986) introduced the notion of the so-called *annealed approximation* to the Statistical Physics of Kauffman networks, giving analytical evidence for a phase transition at $K_c = 2$ in the limit $N \rightarrow \infty$. In this approximation, in each time step to all sites (spins) K inputs and Boolean functions are reassigned randomly,

thus the network topology is *not* quenched. Then, dynamics is iterated and the time evolution of the *Hamming distance*¹ between two spin configurations is calculated, that initially are neighbors in state space. Let us consider two randomly chosen spin configurations $\vec{\sigma}_1$ and $\vec{\sigma}_2$ which have $d(\vec{\sigma}_1, \vec{\sigma}_2) = n$ at $t = 0$. One can calculate the probability $P_1(m, n)$ that the distance $d(\vec{\sigma}_1, \vec{\sigma}_2)$ of their images $\vec{\sigma}_1$ and $\vec{\sigma}_2$ at time $t = 1$ is $d(\vec{\sigma}_1, \vec{\sigma}_2) = m$. With the re-normalization $n/N = x$ and $m/N = y$ the authors show that $P_1(m, n)$ is very peaked around a value

$$y_1 = \frac{1 - (1 - x)^K}{2}. \quad (3.12)$$

Regarding the time evolution of $y = y_t$ of the two spin configurations for all t , this leads to the recurrence relation

$$y_t = \frac{1 - (1 - y_{t-1})^K}{2}, \quad (3.13)$$

y_1 given by 3.12. The authors also give a straight-forward generalization of formula (3.13) to networks where k is not constant:

$$y_{t+1} = \sum_k \rho_k \frac{1 - (1 - y_t)^k}{2}, \quad (3.14)$$

where ρ_k is the (in)degree distribution. For $K \leq 2$, the fixed point $x = 0$ is the only fixed point of the map (3.13) and it is attractive for any starting value x , $y_t \rightarrow 0$ as $t \rightarrow \infty$. So

$$\lim_{t \rightarrow \infty} \lim_{N \rightarrow \infty} y_t = \lim_{t \rightarrow \infty} \lim_{N \rightarrow \infty} \frac{d(\vec{\sigma}_1(t), \vec{\sigma}_2(t))}{N} = 0 \quad \text{for } K \leq 2 \quad (3.15)$$

For $K > 2$, the fixed point $x = 0$ is unstable and there appears a new fixed point y^* of (3.13) which is attractive. So one has

$$\lim_{t \rightarrow \infty} \lim_{N \rightarrow \infty} y_t = \lim_{t \rightarrow \infty} \lim_{N \rightarrow \infty} \frac{d(\vec{\sigma}_1(t), \vec{\sigma}_2(t))}{N} = y^* \quad \text{for } K > 2 \quad (3.16)$$

Thus the annealed approximation indeed confirms that there is a phase transition from “frozen” dynamics (convergence of dynamical trajectories) to “chaotic” dynamics (divergence of trajectories) at $K_c = 2$. Luque and Solé (1996) gave a simple interpretation of these results in terms of “damage spreading”, i.e. percolation of perturbations on the network, and generalized it to sets of coupled networks.

How does the argument of Derrida and Pomeau translate for the dynamics of RTN? Kürten (1988b) showed that for RTN, under the assumption that K remains finite (and

¹Consider two bit strings of length l . The *Hamming distance* then is defined as the number of bits that are different in both bit strings.

is the same for all network nodes, just as in Kauffman's original model), the Hamming distance of two states at time $t + \tau$, given their distance at time t , takes the form of a polynomial spline function of order K :

$$H_K(t + \tau) = F_K(H_K(t)) = \sum_{\nu=1}^K (-1)^{\nu+1} \binom{K}{\nu} a_\nu H_K^\nu(t) \quad (3.17)$$

with

$$a_\nu = 1 \sum_{m=1}^{\nu} (-1)^m \binom{m}{\nu} I_m^{(K)} \quad \nu = 1, \dots, K \quad (3.18)$$

and

$$I_m^{(K)}(\rho, h) = \int \dots \int dx_1 \dots dx_K \rho(x_1) \dots \rho(x_K) \cdot \Theta[x_{m+1} + \dots + x_K + h]^2 - (x_1 + \dots + x_m)^2], \quad (3.19)$$

where $\Theta(x)$ represents the Heaviside function and $\rho(c_{ij})$ is the statistical distribution of interaction weights. It can be shown that a possible critical point or critical line then is completely determined for $a_1 \neq 0$ by the equation

$$\left. \frac{dF_K}{dH_K} \right|_{H_K=0} = K a_1 = 1. \quad (3.20)$$

The main drawback of Kürten's method is that it can be applied only to networks with finite and constant K - a generalization for random networks, where the number K of inputs per node is allowed to vary, fails because of the complexity of the analytical expressions for large K . In chapter 6 of this thesis, we introduce a new approach that solves the problem of finding the critical point for this case with $h = 0$ (Rohlf and Bornholdt, 2002). The main idea is the following: First, by combinatorial considerations, the probability $p_s(k)$ for damage spreading at a given node as a function of the number k of its inputs is derived; it is found that for $k = 2, 4, 6, \dots$ it obeys the relation

$$p_s(k) = \frac{k \cdot \binom{k}{k/2} + (k+2) \cdot \binom{k}{(k+2)/2}}{k \cdot 2^{k+1}} \quad (3.21)$$

$$= p_s(k+1), \quad (3.22)$$

and for $k = 3, 5, 7, \dots$ one obtains

$$p_s(k) = \frac{(k+1) \cdot \binom{k}{(k+1)/2}}{k \cdot 2^k}. \quad (3.23)$$

For large k , in good approximation this scales as

$$p_s(k) = \frac{1}{k+1} \cdot \sqrt{\frac{2k}{\pi}} \cdot \left(\frac{k}{\sqrt{k^2-1}} \right)^k \approx \sqrt{\frac{2}{\pi}} \cdot \frac{1}{\sqrt{k}} \quad (3.24)$$

Assuming a Poisson distribution for the connectivity and taking the average over all nodes, one obtains the average probability for damage spreading at a given mean connectivity \bar{K} :

$$\langle p_s \rangle(\bar{K}) = e^{-\bar{K}} \left\{ 1 + \frac{1}{2} \sum_{i=1}^{\infty} \frac{1}{(i!)^2} \left(\frac{1}{2i} + \bar{K} \right) \cdot \left(\frac{\bar{K}}{2} \right)^{2i-1} \right\} \quad (3.25)$$

In chapter 6 we show that the time evolution of the (normalized) Hamming distance y then is given by the recurrence relation

$$y_{t+1} = \langle p_s \rangle(\bar{K}) (1 - e^{-\bar{K}y_t}). \quad (3.26)$$

A linear stability analysis of this map finally yields the critical mean connectivity

$$K_c = 1.849 \pm 0.001 \quad (3.27)$$

for this type of RTN, which, surprisingly, is smaller than the minimum critical value $K_c^{bool} = 2$ found in RBN.

Interestingly, Nakamura (2003) has applied the approach introduced in (Rohlf and Bornholdt, 2002) to RTN with scale-free network topology, and, introducing stochastic update errors into dynamics, found a non-monotonous dependence of damage spreading on the system temperature T . This result accords well with the notion of “damage suppression” at highly connected nodes (hubs) in RTN, that will be discussed in chapter 6.

3.2.5 Local measures of network dynamics

The simplest measures of network dynamics at the single-node level are the *average activity* of a network site i , which is the pendant of the magnetization per site in other spin systems, and the *average correlation* of the dynamical states of a pair (i, j) of sites. For RBN, some results based on numerical simulations are known, that connect the average activity with global dynamical properties as, for example, the number of “relevant nodes” (Bastolla and Parisi, 1998b), for RTN, however, only few results have been published (e.g. in Bornholdt and Rohlf, 2000).

In the following, the respective definitions of these two *local measures of network dynamics* for RTN are given, followed by a summary of their typical properties.

3.2.5.1 Average activity

The *average activity* $A(i)$ of a site i is defined as the average over all states $\sigma_i(t)$ site i takes in dynamical network evolution between two distinct points of time T_1 and T_2 :

$$A(i) = \frac{1}{T_2 - T_1} \sum_{t=T_1}^{T_2} \sigma_i(t) \quad (3.28)$$

“Frozen” sites i which do not change their states between T_1 and T_2 obviously have $|A(i)| = 1$, whereas sites occasionally flipping from $\sigma_i = 1$ to $\sigma_i = -1$ or vice versa have $0 \leq |A(i)| < 1$.

3.2.5.2 Average correlation

The *average correlation* $\text{Corr}(i, j)$ of a pair (i, j) of sites is defined as the average over the products $\sigma_i(t)\sigma_j(t)$ in dynamical network evolution between two distinct points of time T_1 and T_2 :

$$\text{Corr}(i, j) = \frac{1}{T_2 - T_1} \sum_{t=T_1}^{T_2} \sigma_i(t)\sigma_j(t) \quad (3.29)$$

If the dynamical activity of two sites i and j is correlated, i.e. if σ_i and σ_j always have either the same or the opposite sign, one has $|\text{Corr}(i, j)| = 1$. If the relationship between the signs of σ_i and σ_j occasionally changes, one has $0 \leq |\text{Corr}(i, j)| < 1$.

3.2.5.3 Properties of the two measures in RTN

The properties of the average activity in RTN as a function of the mean connectivity \bar{K} have been studied in detail by Rohlf (2000). One main result concerns the *frozen component* C , i.e. those network sites that “freeze” to a constant state in the asymptotic limit of dynamical evolution, and the *number of correlated pairs* C_{corr} . Both quantities show a phase transition near the critical connectivity K_c , that becomes sharp in the limit of large system sizes (Rohlf, 2000; Bornholdt and Rohlf, 2000). This distinct property, that relates local network dynamics and network topology to each other, qualifies $A(i)$ and $\text{Corr}(i, ij)$ as suitable measures for the study of network evolution based on local rewiring events. Other quantities of interest concern the sites which actively take part in the dynamics; one can study, e.g., the average activity $\langle A \rangle_{nf}$ of non-frozen sites or the average correlation $\langle \text{Corr} \rangle_{nc}$ of pairs of sites with at least one active site. These quantities can be related to the global network activity $\langle A \rangle_g$, defined as the average over the activity of all network sites, by the following relations (Rohlf, 2000):

$$\langle A \rangle_{nf} = \frac{\langle A \rangle_g - C}{1 - C} \quad (3.30)$$

and correspondingly

$$\langle \text{Corr} \rangle_{nc} = \frac{\langle A \rangle_g - C_{corr}}{1 - C_{corr}}. \quad (3.31)$$

It was shown that both quantities become maximal near the critical connectivity K_c (Rohlf and Bornholdt, 2002, cf. Sec. 6.6 of this thesis).

Chapter 4

Statistical Physics of Cellular Automata (CA)

4.1 Introduction

Cellular Automata (CA) are dynamical systems that are discrete in state, space and time. CA are paradigmatic examples of *decentralized spatially extended dynamical systems*: state changes (dynamics) depend only on the local neighborhood of a cell and some (usually very simple) update rules, there is no central control mechanism that coordinates dynamics. The concept of CA originally was introduced by von Neumann, after a suggestion of Ulam, to study the logical organization behind biological self-reproduction (von Neumann, 1966). Von Neumanns system was very complicated, it needed more than 30 different states and complex transition rules to work, however, it was soon recognized that CA with more simple update rules may serve as powerful models for the study of pattern formation and emergent complex behavior. CA became popular with Conways invention of the “Game of Life” in 1968 (cf. Gardner, 1983), a two-dimensional CA with very simple update rules, yet able to produce astonishingly complex patterns and, as it was shown later, even capable of universal computation. However, the beginning of systematic research on CA had to wait until the 1980s, and, to a large extent, was made possible by the fast development of computer technology that had taken place in the meantime. CA research now shifted to modeling natural and artificial systems in computer simulations (for a review, cf. for example Gerhardt and Schuster, 1995).

Traditionally, pattern formation in spatially extended systems has been modeled with (partial) differential equations (PDE); prominent examples are *reaction-diffusion systems*, that will be reviewed in Sec. 5.1 in the context of biological pattern for-

mation. Are cellular automata only some kind of “coarse grained approximation” to models based on differential equations? Indeed, Toffoli (1984) argued that CA rather represent a fully-fledged *alternative* to PDE-based modeling in physics. In many cases, they may be more suitable to define a modeling framework for a given problem than differential equations, e.g. for complex systems that are explicitly formed by discrete “agents”; problems of this type can be found in various areas as, for example, traffic simulations (Schreckenberg, 1995; Simon and Nagel, 1998), simulations of the immune system (Kamp and Bornholdt, 2002), or, in Statistical Physics, the Ising model of magnetism, which also can be interpreted as a CA with “update rules” defined by a Hamiltonian. For pattern formation in biological organisms, however, the situation is not that clear: on the one hand, the discrete cellular organization of (multi-cellular) organisms obviously seems to speak in favor of the application of CA, on the other hand, biological cells themselves are very complex objects with processes of information- and energy-transfer taking place on several different scales, ranging from single molecules over regulatory networks to motions of whole cell ensembles (cf. chapter 5). However, depending on the context, one may take a pragmatic view: cell differentiation into n different cell types, for example, could be interpreted in terms of CA with n different states, which then allows for the study of pattern formation on the “scale” of cell differentiation. A subsequent, more sophisticated interpretation in terms, e.g., of coupled regulatory networks, is possible and is studied in chapter 8. Similarly, additional levels of complexity in morphogenesis, e.g. cells movements, as studied in chapter 9, and cell-cell adhesion (Hogeweg, 2000) can be integrated in terms of suitably defined cellular automata.

In the following, a short introduction into the Statistical Physics of CA is provided, defining the framework for the development of a CA-based model of the emergence of position information and proportion regulation in biological pattern formation (Rohlf and Bornholdt, 2003).

4.2 Definitions

4.2.1 One-dimensional CA

A one-dimensional cellular automaton consists of a linear lattice (string) of identical cells, each of which can take one of a finite number n of states. The (local) state of cell i at time t is denoted by

$$\sigma_i(t) \in \Sigma = \{0, 1, \dots, n-1\} \quad \text{with} \quad t \in \mathbb{N}_0. \quad (4.1)$$

The (global) state $\vec{\sigma}(t)$ at time t then is defined as the state vector

$$\vec{\sigma}(t) = (\sigma_0(t), \sigma_1(t), \dots, \sigma_{N-1}(t)) \in \Sigma^N, \quad (4.2)$$

where N is the size of the lattice (the number of cells).

At each time step, all cells in the lattice are updated *in parallel* according to a *local update rule* f :

$$\sigma_i(t) = f(\sigma_{i-r}(t-1), \dots, \sigma_{i-1}(t-1), \sigma_i(t-1), \sigma_{i+1}(t-1), \dots, \sigma_{i+r}(t-1)) \quad (4.3)$$

with $f : \{0, 1, \dots, n\}^k \mapsto \{0, 1, \dots, n\}$, where $k = 2r + 1$ is the *neighborhoodsize* and r the *radius* of the CA. In this thesis, only rules with $k = 3$ ($r = 1$), i.e. strictly local update rules, are considered. f can be formulated as a *lookup table* (rule table), that explicitly maps each neighborhood state to a certain output state.

There exist n^k different neighborhood configurations, hence the state space has size $S = n^k$; this means that even for moderate n and k the number of possible states is already very large (e.g., for $n = 3$ and $k = 3$, there exist $3^{27} \approx 10^{12}$ different states). Consequently, simple update rules can lead to very complex CA dynamics. For finite N , boundary conditions have to be specified explicitly.

4.2.2 Two-dimensional CA

A two-dimensional cellular automaton consists of a two-dimensional lattice of identical cells, each of which can take one of a finite number n of states. The (local) state of the cell (i, j) at time t is denoted by

$$\sigma_{i,j}(t) \in \Sigma = \{0, 1, \dots, n-1\} \quad \text{with} \quad t \in \mathbb{N}_0. \quad (4.4)$$

The (global) state $\Sigma(t)$ at time t then is defined as the state matrix

$$\Sigma(t) = \begin{pmatrix} \sigma_{0,0}(t) & \sigma_{0,1}(t) & \cdots & \sigma_{0,N_2-1}(t) \\ \sigma_{1,0}(t) & \sigma_{1,1}(t) & \cdots & \sigma_{1,N_2-1}(t) \\ \vdots & \vdots & & \vdots \\ \sigma_{N_1-1,0}(t) & \sigma_{N_1-1,1}(t) & \cdots & \sigma_{N_1-1,N_2-1}(t) \end{pmatrix}, \quad (4.5)$$

if the CA has size $N = N_1 \times N_2$. In two-dimensional CA, the update neighborhood of a cell (i, j) can be defined in two different ways, either as *von Neumann neighborhood* or as *Moore neighborhood*; both possibilities are illustrated in Fig. 4.1. In this thesis,

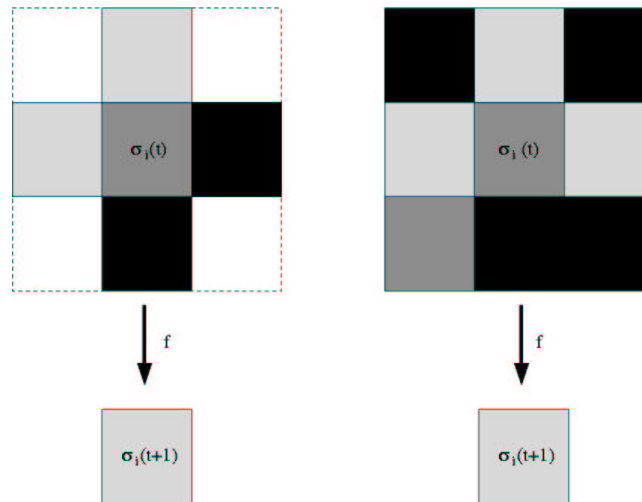


Figure 4.1: Schematic illustration of a two-dimensional CA with three different states (light grey, dark grey and black). Left panel: von Neumann neighborhood, update of the middle cell depends only on the states of the four nearest neighbors and its own state, right panel: Moore neighborhood, update of the middle cell depends on eight neighbor cells and its own state.

only two-dimensional CA with von Neumann neighborhood with radius $r = 1$ are considered, i.e., each cell (i, j) updates its state according to

$$\sigma_{i,j}(t) = f(\sigma_{i,j-1}(t-1), \sigma_{i,j+1}(t-1), \sigma_{i,j}(t-1), \sigma_{i-1,j}(t-1), \sigma_{i+1,j}(t-1)) \quad (4.6)$$

with $f : \{0, 1, \dots, n\}^5 \mapsto \{0, 1, \dots, n\}$.

For finite N , boundary conditions have to be specified explicitly (e.g., as periodic).

4.3 Statistical and Computational Mechanics of Cellular Automata

The simplest class of one-dimensional cellular automata was studied extensively by Wolfram (1983, 1984a,b). In particular, he considered CA with $n = 2$ and $k = 3$. Starting dynamics from simple (e.g., periodic) initial configurations, these CA either tend to homogeneous states or generate self-similar patterns (in a two-dimensional space-time plot) with fractal dimensions ≈ 1.59 or ≈ 1.69 (Wolfram, 1983). For random initial configurations, complex self-organization phenomena are observed. CA dynamics is *locally irreversible*: different initial configurations can lead to the same final configuration (dynamical attractor). This phenomenon is also found in discrete

dynamical networks, because of their complex basins of attraction. Irreversibility in CA state evolution can, e.g., be characterized by the time evolution of the *average state entropy*

$$\langle S \rangle = \frac{1}{M} \sum_{i=1}^M \sum_{j=1}^n p_j \log_2 p_j, \quad (4.7)$$

taken over a statistical ensemble starting from M different initial states, where p_j is the probability for state j . Whereas for reversible systems, time evolution almost always leads to an increase of entropy (which reflects *Liouville's Theorem*, that the total number of possible configurations must remain constant with time), in irreversible systems entropy may decrease with time. Indeed, in CA typically a *decrease* in entropy with time is found, which reflects the fact that the number of accessible state configurations decreases with time (Wolfram, 1983). Consequently, CA in principal have the capability to transform spatio-temporal “disorder” into “order”, which predestinates them as simple, yet non-trivial tools for the study of spatial pattern formation.

Wolfram proposed a qualitative classification of (asymptotic) CA behavior into four classes, intending to identify “universality classes” of all possible CA behavior (Wolfram, 1984a,b):

1. A (fixed) homogeneous state
2. (Stable) periodic attractors (periodic in time and/or space)
3. “Chaotic” dynamics (in time and space)
4. “Complex” dynamics.

An example of a class 4 CA is the two-dimensional “Game of Life” introduced by Conway (cf. Gardner, 1983).

Langton (1986) introduced the λ parameter as an attempt to make Wolfram’s scheme more quantitative. The λ parameter is obtained as a result of a statistic of the output states in the CA lookup table. Strictly speaking, it is defined as the fraction of *non-quiescent states* in this table, where the *quiescent state* of a CA is an arbitrarily chosen (but fixed) state $\sigma \in \Sigma$. For $\Sigma = \{0, 1\}$, e.g., if state 0 is chosen as quiescent state, λ is simply the fraction of 1s among the output states in the rule table. λ can also be interpreted in terms of update dynamics: if an (infinite length) CA is started with a random initial configuration, such that all possible neighborhood configurations occur with equal probability, then λ is the fraction of 1s in the CA lattice at the next time step.

Langton created CA rule tables at random, but with increasing λ values, and observed that CA dynamics typically changes in a way, that goes through Wolfram’s

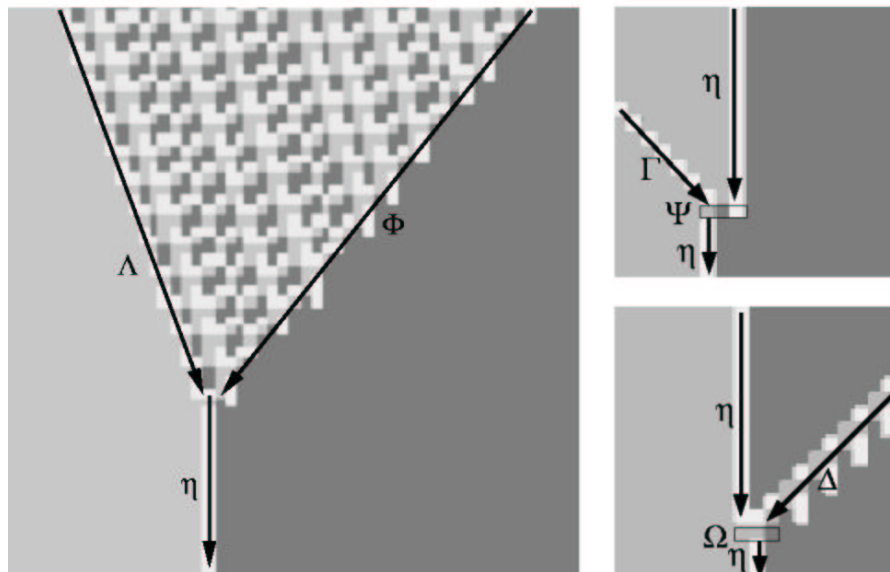


Figure 4.2: Quasi-particles in the CA model of pattern formation introduced by Rohlf and Bornholdt (2003) (cf. chapters 8 and 9). The Ψ and the Ω particle are transient excitations that “decay” immediately into a stable η particle.

classes in the order $1 \rightarrow 2 \rightarrow 4 \rightarrow 3$, analogous to a bifurcation sequence in a dynamical system when some control parameter is increased. Hence, Langton argued, “complex” dynamical behavior of CA is typically found at intermediate or “critical” λ values (metaphorically called the “edge of chaos”). The “critical” λ values, Langton claimed, are associated with a phase transition between periodic and chaotic CA dynamics (Langton, 1986, 1990, 1991).

The (parallel) *computing capacities* of CA have attracted much interest. It could be shown, for example, that the “Game of Life” is capable of *universal computation* (in the sense of a Turing machine): some basic dynamical configurations in this two-dimensional CA can perform basic logical operations, as, for example, AND, or and NOT, others in principle can serve as registers for information storage. The outlines of a proof that the “Game of Life” CA is universal were given by Berlekamp (1982) and independently by Gosper (cf. Gardner, 1983), and Chapman (2002) constructed a “life” pattern that implements the actions of a universal register machine, thus explicitly proving that this CA is a universal computer. Interestingly, there also exist one-dimensional CA with universal computing capacity (Smith, 1971; Lindgren and Nordahl, 1990; Wolfram, 2002).

Computational Mechanics (CM) of CA was developed as a kind of “high level description” of dynamical computing processes in CA (Hanson, 1993; Hanson and Crutch-

Particles				Interactions
Label	P	d	v	
Γ	1	1	1	$\Gamma + \eta \rightarrow \Psi$
Δ	2	-2	-1	$\Delta + \eta \rightarrow \Omega$
Λ	3	1	1/3	$\Lambda + \Phi \rightarrow \eta$
η	1	0	0	$\Psi \rightarrow \eta$
Φ	≈ 1	≈ -0.85	≈ -0.85	$\Omega \rightarrow \eta$
Ψ	0	0	0	
Ω	0	0	0	

Table 4.1: Particle catalog of the CA model introduced in Rohlf and Bornholdt (2003) (cf. chapters 8 and 9). The sign of v is defined by the direction of movement of the particle (positive sign for movement to the right side of the 1D-lattice, negative sign for movement to the left side).

field, 1997; Hordijk, 1998, 1999). CM is based on the observation, that in many CA, there are only a few important patterns, i.e. regularities in space-time dynamics, that dominate the global behavior of the system. Important in this context is the concept of *regular domains*, which are defined as a set of spatial configurations in the CA with the following two properties (Hordijk, 1999):

- Temporal invariance or periodicity
- Spatial homogeneity,

where the latter means a form of spatial translation invariance, i.e. the regular domain, or pattern, is position independent and can occur anywhere in the lattice. Many of the CA computing capacities, however, are rather based on *violations* of domain regularities. Interestingly, many of these violations are regular, repeating structures themselves.

This observation is taken as the starting point of a higher level description of CA computation called “particle computation” (Das *et al.*, 1994; Crutchfield and Mitchell, 1995). In this context, a *particle* α is defined as a spatially localized (i.e., bounded width), temporally periodic boundary between two adjacent domains. In terms of a more physical language, a particle may be viewed as a regular, soliton-like excitation moving over the CA lattice. Hence, particles have a characteristic period P (first repeat time of a characteristic pattern) and velocity $v = d/P$, where d is the displacement of the particle on the lattice during the period. CA computation arises from *particle interactions*: similar to particles in elementary particle physics, CA particles can, e.g.,

annihilate each other or can be transformed into a set of new particles, when they meet each other. Examples of these processes are shown in Fig. 4.2 (adapted from Rohlf and Bornholdt (2003, 2004b), cf. chapters 8 and 9). Particles and their interactions can be summed up in a *particle catalog*. The particle catalog for the model introduced in chapter 8 of this work is shown in Table 4.1. Note that the Φ particle in this model is not a particle in the strict sense, as it shows some stochastic irregularities in its “trajectory” due to randomness in the initial condition, however, on average, it has a characteristic period and velocity.

As outlined above, state space becomes large even for quite simple CA. If one wishes to solve particular problems with CA, such as a pre-defined computational task or self-organization of spatial pattern with certain properties, one has to define an appropriate *search strategy*. In most cases, classical optimization routines fail in this case, however, it has been shown that *genetic algorithms* (GA) (Holland, 1975) are very efficient in finding candidate CA rule tables (Crutchfield and Mitchell, 1995; Hordijk, 1999); hence, in this thesis genetic algorithms have been applied for evolution of CA solutions in the context of spatial pattern formation (cf. chapters 8 and 9, and the appendix for details about the GA that has been applied).

Chapter 5

Mathematical models of Morphogenesis

In this chapter, a review of some concepts and mathematical modeling approaches commonly used in the theory of morphogenesis is provided. The main focus is set on two classes of systems, that - to some degree - can be regarded as autonomous strains of research:

- reaction-diffusion systems (Sec. 5.1)
- coupled regulatory networks (Sec. 5.3).

Coupled chemical reaction networks, in some sense a in-between of both classes, are discussed in Sec. 5.2.

5.1 Reaction-diffusion models and morphogenetic gradients

5.1.1 Turing-type models as a paradigm of pattern formation

First systematic observations how diffusive chemicals could influence animal development were made in the late nineteenth century. In the first half of the twentieth century, for certain model organisms, as e.g. the Sea Urchin, first hypotheses about physico-chemical mechanisms of regulation were formulated, which served as phenomenological models for explanation of experimental observations. A prominent example is the *double gradient hypothesis* of Sea Urchin development (Hörstadius, 1935, 1973). However, the mechanism leading to generation and maintenance of such “morphogen

gradients” as well as their chemical nature remained obscure. In 1952, Turing proposed the first mathematical theory of morphogenesis including a plausible physico-chemical mechanism for generation of spatial patterns.

Turing-type models of morphogenesis are based on reaction-diffusion approximations involving a small number of molecular species (morphogens) diffusing on a continuous domain (Meinhardt, 1982). Turing patterns arise as a consequence of spatial instabilities in models of the form:

$$\frac{\partial \vec{\phi}}{\partial t} = \Delta \vec{\phi} + F(\vec{\phi}), \quad (5.1)$$

where Δ is the Laplace operator and $\vec{\phi}(\vec{r}, t) = (\phi_1, \dots, \phi_n)$ is the vector involving the concentration ϕ_i of each morphogen. The two terms on the right-hand side introduce the diffusion and the reaction parts, respectively (\vec{r} and t are spatial and temporal coordinates, respectively). For $n = 2$ the model can be solved analytically and has, for example, spatially periodic solutions with a characteristic wavelength λ . In particular, most models are based on two morphogens with activatory and inhibitory properties and different diffusion constants. Gierer and Meinhardt (1972) showed that two features are prerequisite for stable morphogen gradients in this type of model: local self-enhancement and *long-range inhibition*, thus the activator a is *auto-catalytic* and has a small diffusion constant compared to the inhibitor h , which counteracts the auto-catalytic feedback loop by *lateral inhibition*. This is described by the following set of partial differential equations for the concentrations of a and h (Gierer and Meinhardt, 1972; Koch and Meinhardt, 1994):

$$\frac{\partial a}{\partial t} = D_a \Delta a + \rho_a \frac{a^2}{(1 + \kappa_a a^2)h} - \mu_a a + \sigma_a \quad (5.2)$$

$$\frac{\partial h}{\partial t} = D_h \Delta h + \rho_h a^2 - \mu_h h + \sigma_h \quad (5.3)$$

D_a and D_h are the diffusion constants, μ_a and μ_h the removal rates, and ρ_a, ρ_h the cross-reaction coefficients; σ_a, σ_h are basic production terms; κ_a is a saturation constant. For stationary patterns the condition $\mu_h > \mu_a$ is necessary. The saturation constant κ_a determines the form of the pattern: for $\kappa_a = 0$ irregular patterns arise, whereas $\kappa_a > 0$ leads to stripe-like patterns. Lateral inhibition can also be achieved by the depletion of a substance s required for the auto-catalysis of a . In its simplest form the system then can be formulated as

$$\frac{\partial a}{\partial t} = D_a \Delta a + \rho_a (a^2 s - a) \quad (5.4)$$

$$\frac{\partial s}{\partial t} = D_h \Delta s + \rho_s (1 - a^2 s) \quad (5.5)$$

Based solely on the mechanisms of either Eqn. 5.3 or 5.5 a monotonic gradient can be maintained only if the tissue is small (less than 1mm or 100 cells in diameter, Wolpert (1969)). If an organism grows beyond this, cells must respond to the gradient by activating particular genes. Koch and Meinhardt (1994) argue that the gene activity, once triggered, should be independent of the evoking signal, thus such a biochemical switch requires either direct or indirect auto-catalytic activation Meinhardt (1978); this genetic switch can be modeled by (Koch and Meinhardt, 1994):

$$\frac{\partial y}{\partial t} = \rho_y \frac{y^2}{1 + \kappa_y y^2} - \mu_y y + \sigma_{ext}, \quad (5.6)$$

where σ_{ext} is the external stimulus. In the absence of such a signal, the system has two stable steady states: a low one at $y = 0$ and a high one at $y = (\rho_y + \sqrt{\rho_y^2 - 4\kappa_y \mu_y^2}) / 2\kappa_y \mu_y$. If σ_{ext} exceeds a certain threshold, the system switches from the low to the high state.

Combinations of several systems 5.3 or 5.5 with 5.6 lead to complex patterns. Mostly, the systems are combined in a *hierarchical* way: A first system A establishes a primary pattern which is used to modify and trigger a second system B , e.g. a genetic switch.

5.1.2 Example: Gierer-Meinhardt model of Hydra pattern formation

Gierer and Meinhardt applied their activator-inhibitor model to pattern formation in the fresh water polyp *Hydra*, and showed that - under certain assumptions - this model could explain a number of experimental observations (see section ... for details on the developmental biology of *Hydra*).

The model is based on reaction-diffusion equations, including several subsystems with activators and inhibitors for the distinct pattern forming systems in *Hydra* (head, foot, hypostome). In addition to their original Turing-like model the authors had to include a “source density” to account for the regeneration properties of *Hydra*: local “sources” of morphogens stabilize the animal-wide morphogen gradients. The concentration change of the primary head activator a_H and its inhibitor h_H is modeled by two coupled equations,

$$\frac{\partial a_H}{\partial t} = \frac{\mu_H \rho (a_H^2 - \rho_{0_H})}{h_H} - \mu_H a_H + D_{a_H} \frac{\partial^2 a_H}{\partial x^2} \quad (5.7)$$

$$\frac{\partial h_H}{\partial t} = \mu_H \rho a_H^2 - \nu_H h_H + D_{h_H} \frac{\partial^2 h_H}{\partial x^2} + \rho_{1_H}, \quad (5.8)$$

where $\rho(x)$ is the source density and μ and ν are the rates of activator and inhibitor degradation, respectively. The RD-equations used for foot and tentacle formation

are very similar, and $\partial\rho/\partial t$ follows a diffusion equation with additional terms for the feedback effects of head and foot activators.

Head activation is assumed to take place at regions of the highest source density, foot activation has the opposite behavior, hence - under this assumption - the model correctly reproduces the polarity of *Hydra*, regeneration properties also are reproduced quite well (e.g. cutting experiments). Budding is regarded as a trigger of second head activator maximum and a source density minimum in this model; the head system has an inhibitory, the foot system an activating effect, both effects together might regulate the position of the budding point at the head-foot axis. The periodic spacing of successive buds at nearly 180 degree angles arises as a direct consequence of lateral inhibition in this model.

Despite its success in explaining numerous experimental observations, this model of *Hydra* pattern formation shares limitations inherent to all Turing-type models, which shall be discussed in the next section.

5.1.3 Limits of reaction-diffusion models and gradient-based pattern control

Activator-inhibitor models of biological pattern formation have been studied in detail and were applied to model morphogenesis of numerous organisms (Koch and Meinhardt, 1994). Although this type of model provides a very intuitive phenomenological framework and apparently convincing explanations for numerous experimental observations, it has inherent limitations that leave several questions unresolved.

The first major point is that morphogen gradients originating from a Turing-type reaction scheme have not been observed experimentally so far. Axis formation of *Drosophila* is often quoted as a striking example of gradient-controlled development (Alberts *et al.*, 2002), however, this example does not provide an argument in favor of activator-inhibitor models. Morphogen gradients that control position dependent gene activation in the *Drosophila* embryo are imposed externally (by follicle cells of the mother, (Alberts *et al.*, 2002)), and *Drosophila* has a very special mode of development involving a *syncytium* (i.e., in early embryogenesis, cell nuclei are not separated by cell membranes, but lie all in the same cytoplasm), which drastically facilitates establishment of diffusive gradients. However, in later stages of *Drosophila* development, where cellular organization is already present, gradients of proteins (namely, *DPP* and *Wingless* during development of the imaginal disk) also have been observed (Entchev *et al.*, 2000; Strigini and Cohen, 2000). For this case, Meinhardt (1983) proposed a “boundary model” which envisioned three long-range gradients, whereas in fact there are only two. Hence, this example does not fit really convincingly with a reaction-

diffusion model either. For Vulva development of *Caenorhabditis elegans*, there is some indirect evidence for pattern control by a morphogene gradient (Kenyon, 1995), but evidently this gradient is produced and maintained by a single cell (the “anchor cell”) and extends only over a few cells, which makes involvement of long range inhibition (the second essential ingredient of the Gierer-Meinhardt model) quite improbable.

Here, we are pointed at an inherent shortcoming of reaction-diffusion models of pattern formation: their dependence on *parameter tuning*, namely of diffusion constants, to work at all. The diffusion constants of activator and inhibitor have to be separated by several orders of magnitude to achieve short-range activation and long-range inhibition at the same time (Lander, Nie and Wan (2002)). If both activator and inhibitor are proteins as typically found in biological cells, this is by no means a trivial assumption. Partially, this problem may be relaxed by introducing additional parameters (e.g. protein degradation), but of course this makes the models more complex and thus even more prone to parameter tuning.

Doubts about the plausibility of RD-models have been issued at a even more basic level: in a cellular environment, diffusion itself as a driving force of pattern formation is a highly non-trivial assumption. Gap-junctions between cells allow only for diffusion of small molecules and ions (Kumar and Gilula, 1996), and hence cannot contribute significantly to establishment of morphogen gradients: proteins are simply “too large”. Consequently, one has to assume that systems of membrane-associated *receptors* control the gradient. Kerszberg and Wolpert (1998) asserted that capture of morphogens by receptors so impedes diffusion that useful stable gradients can never arise by that mechanism, and propose that active transport mechanisms (as, e.g., transcytosis or “bucket brigade” transfer of molecules) establish the gradient. However, Lander, Nie and Wan (2002) showed that certain cell biological processes would have to occur at implausibly fast rates for these mechanisms to work, at that receptor-based diffusion in principle could work given that receptor numbers, kinetics and internalization and degradation meet certain conditions. They point out that these conditions seem to fit with observations, however, doubts about robustness and biological plausibility of the necessary “tuning” of several critical parameters remain.

Last, in the reaction-diffusion picture, *cell internal capacities of information processing*, e.g. by chemical or gene regulatory networks, are neglected completely. In the next section, we will review an approach of Kaneko et al. that still relies on diffusion, however, represents a significant step forward by inclusion of cellular organization and cell-internal information processing into a theory of biological pattern formation.

5.2 Chemical networks as models for differentiation and pattern formation

Several approaches have been suggested to overcome the inherent limitations of reaction-diffusion models, namely their oversimplification of metabolism and - even more significant - the neglect of cellular organization. Cellularity leads to new problems and questions beyond the scope of simple RD-models, as, for example, cell *differentiation*, occurrence of more complex processes of local signal-transfer and -processing (e.g. involving membrane-associated receptors or active transport) and complex spatial rearrangements of cell populations during development, including cell divisions, cell movements and cell death (apoptosis).

Cellular differentiation due to a coupling of intra-cellular chemical dynamics and cell-cell signalling by diffusing morphogens was modeled by Furusawa and Kaneko (2000, 2001). Their model consists of a one-dimensional string of cells, surrounded by a medium. Cellular dynamics is modeled by inclusion of biochemical reactions with enzymes, simple cell-cell interactions through diffusion and cell division as a result of biochemical reactions within each cell. Core of the model is a catalytic network among k chemicals, represented by a matrix $Con(\ell, j, m)$ which takes on unity when the reaction from the chemical l to m is catalyzed by j , zero otherwise. Each cell takes penetrable chemicals from the medium as nutrient, and the reaction network transforms them to unpenetrable chemicals. As a result the volume of the cell increases, which triggers cell division (when the volume is doubled).¹ In the model it is assumed that there are two types of chemicals, one which can penetrate the membrane and one which can not. The notation σ_m is used, which takes the value 1 if the chemical $c_i^{(m)}$ is penetrable, and 0 otherwise. The concentration of the i th chemical in cell ℓ then is given by

$$\frac{dc_i^{(\ell)}(t)}{dt} = \Delta c_i^{(\ell)}(t) - c_i^{(\ell)}(t) \sum_{l=1}^k \Delta c_i^{(\ell)}(t), \quad (5.9)$$

with

$$\begin{aligned} \Delta c_i^{(\ell)}(t) = & \sum_{m,j} Con(m, j, \ell) e_1 c_i^{(m)}(t) (c_i^{(j)}(t))^\alpha - \sum_{m',j'} Con(\ell, j', m') e_1 c_i^{(\ell)}(t) (c_i^{(j')}(t))^\alpha \\ & + \sigma_\ell D(C^{(\ell)}(p_i^x, t) - c_i^{(\ell)}(t)). \end{aligned} \quad (5.10)$$

¹Hogeweg (2000) uses a similar approach of volume triggered cell division, which - in some instances - was indeed observed experimentally (Ruoslahti, 1997; Chen and Mrksich, 1997). However, for setting up a more general model scenario of pattern formation and morphogenesis, *genetic control* of the cell cycle certainly cannot be neglected any more. Indeed, this approach was taken, e.g., by Jackson, Johnson and Nash (1986).

The variable p_i^x denotes the location of the i -th cell, and α is the degree of catalyzation. The second term in Eqn. 5.10 represents the dilution effect by changing the volume of the cell. On the other hand, the diffusion of chemicals in the medium is governed by the following equation for $C^{(\ell)}$:

$$\frac{\partial C^{(\ell)}(x, t)}{\partial t} = -\tilde{D}\nabla^2 C^{(\ell)}(x, t) + \sum_i \delta(x - p_i^x) \sigma_\ell D(C^{(\ell)}(x, t) - c_i^{(\ell)}(t)), \quad (5.11)$$

where the boundary condition is chosen to be $C(0, t) = C(x_{max}, t) = const.$, \tilde{D} is the diffusion constant of the environment, and x_{max} denotes the extent of the medium. In this model, by an extensive statistical analysis two types of reaction networks were identified: the first class shows fast growth (cell number increases exponentially in time); during subsequent divisions, starting from a homogeneous “stem” cell population, cells become differentiated. In the differentiated cells, chemical dynamics is simpler (periodic orbits or point attractors) than in the stem cells, which show chaotic dynamics. As a result of differentiation, the cells make a cooperative use of the chemicals which leads to the observed fast growth. The second class of networks shows no differentiation and a slower ensemble growth compared to “differentiated” cell ensembles of the first class. The developmental process of cells of the first class is irreversible; the loss of multipotency, known to exist in the developmental process of real organisms), is explained in terms of a transition from complex (chaotic) dynamics to simple dynamics in differentiated cells.

In the most recent version of this model (Furusawa and Kaneko, 2003), in addition to chemical dynamics and volume-triggered cell divisions, also cell death and cell adhesion are included. In this extended model, self-organization of *position information*, i.e. patterns of cell-differentiation that depend on the position of the cell in the cell ensemble, is observed. In fact, this is an almost complete *multi level model* of morphogenesis, that integrates three levels of complexity (cell internal dynamics, cell-to-cell communication and “high level” morphogenetic processes as, for example, cell movement); in this sense, the model can be compared well with a *multi-level model of morphogenesis* introduced by Hogeweg (2000), however, this model is based on Boolean networks and local cell-cell communication by membrane-bound receptors.

The model introduced by Furusawa and Kaneko reproduces several features observed in animal development with striking accuracy, e.g., irreversibility of cell differentiation (loss of multipotency), determination of cell fates in *cell lineages* during development by means of successive redifferentiations, and, as was mentioned last, to some extent it also provides a possible mechanism for self-organization of ordered spatial patterns. From the viewpoint of Statistical Physics, however, the insights into the

dynamics of morphogenetic processes that can be gained from this model are limited, mainly due to its high dimensionality. Chemical dynamics itself is already quite complex (coupled, non-linear partial differential equations), and dimensionality is again increased significantly by diffusion-mediated cell-cell communication and higher-level morphogenetic processes; the number of free parameters is much higher than in older reaction-diffusion models. Hence, it seems almost impossible to “disentangle” the complex interplay of subprocesses in this model and, e.g., to distinguish the dominant factors leading to the emergence of the two “classes” of cell ensembles (e.g., structure and topology of the reaction network versus confinement to certain regions in dynamical parameter space).

5.3 Coupled regulatory networks and local information transfer

The models of pattern formation and morphogenesis discussed in the previous sections, despite their different perspectives and complexity, had two main points in common: a *non-local mechanism of information transfer* (diffusion), combined with *quite complex chemical dynamics*. The following questions arise:

- Can patterns of similar (or even higher) complexity and robustness as in diffusion-based models self-organize from local dynamics (i.e. a combination of local information transfer between cells and cell-internal information processing)?
- What does most significantly contribute to pattern complexity and robustness: chemical/regulatory dynamics or the structure/topology of interactions (networks)?

The first point is well motivated by experimental observations in developmental biology. *Direct contact induction*, as discussed in Sec. 2.2.3, is such a local (or, at least short-range) mechanism of information transfer. Surprisingly, the pattern formation potential of induction has been neglected by theoreticians until recently. The second point sets the focus on the “network perspective”: diffusion-based models of morphogenesis rely on appropriate choices of reaction kinetics and parameters, however, if there exist robust modules or “elementary networks” capable to generate basic patterns, this may not be so important any more. In this case, robustness of development would result considerably from topology and information processing in regulatory networks rather than from precise control of kinetic parameters. Interestingly, such intrinsic robustness of pattern formation against parameter variation was found by von Dassow

et al. (2000) in a model of the *Drosophila* segment polarity network (cf. also Dearden and Akam, 2000). In the following, we shall give a short review of pattern formation models that are based on local signaling between cells and cell-internal information processing by regulatory networks.

Jackson, Johnson and Nash (1986) modeled the dynamics of gene networks capable of pattern generation by discrete, Boolean Networks as introduced by Kauffman (1969, 1993). In this model small numbers of cells arranged in a one-dimensional, growing array are able to communicate with nearest neighbors through inductive interactions, that are modeled in terms of Boolean (logical) functions similar to those of the internal networks. The model is completed by means of a schematic cell cycle which governs the conditions under which the cell will divide or cease division. Eventually, the system settles down into a final state; in other words, the tissue falls into a attractor (i.e. some recurrent pattern of gene expression). One of the interesting results of the study was that cell-cell interactions allowed generation of new cell types, i.e. of dynamical attractors not present in an isolated network, and such a capacity was shown to be maximal when about 20% of the genes were involved in such interactions. This kind of neighborhood-dependent differentiation or cell memory, respectively, is also found in the more complex Boolean network based morphogenetic model studied by Hogeweg (2000).

An alternative theoretical approach to interacting gene networks, based on continuous dynamics and a connectionist architecture similar to neural networks, was considered by Mjolsness, Sharp and Reinitz (1991) in their *connectionist model*. The basic model is defined by a set of differential equations with the following structure:

$$\frac{dg_{ij}}{dt} = -\lambda_j g_{ij} + \Theta \left[\sum_{l=1}^{N_g} W^{jl} g_{il} + \theta_j \right] + D_j \nabla^2 g_{ij}, \quad (5.12)$$

where g_{ij} denotes the concentration of the j th gene product in the i th cell, W^{ij} is a random matrix describing interactions among genes with both positive and negative values, θ_j is the threshold in gene response and D_j the diffusion constant of gene product j . The function defining the interactions among genes, $\Theta(z)$, is assumed to be a sigmoidal one. Specifically, the MSR model uses $\Theta(z) = 1/(1 + \exp(-\mu z))$, but other choices lead to similar results. Using this model and searching over parameter space by means of optimization methods (simulated annealing), these authors have been able to discover correlations in experimental data on gene expression in developmental processes (Mjolsness, Sharp and Reinitz, 1991) and to reproduce the gap gene expression patterns along the antero-posterior axis of *Drosophila* (Reinitz and Mjolsness, 1995). Note that the diffusion constant D_j was chosen in a way that allowed only for diffu-

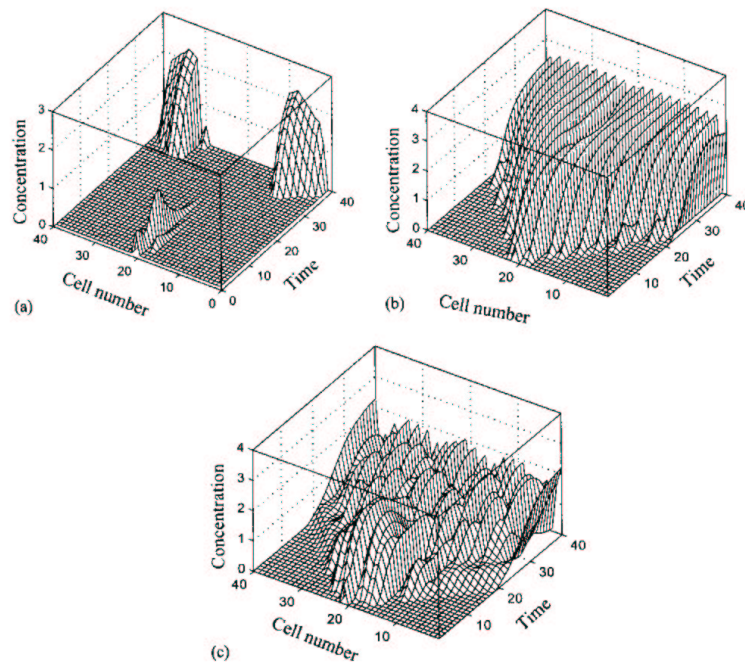


Figure 5.1: Typical spatio-temporal patterns in the DCI model studied by Salazar *et al.*: (a) transient activation of gene expression with decay and further recruitment at later times; (b) stripe pattern and (c) low-dimensional chaos (after Solé, Salazar-Ciudad and Garcia-Fernandez (2002)).

sion between nearest neighbors, hence this model can be considered “local” in spite of diffusion based information transfer (as described by the Laplacian in Eqn. 5.12).

The connectionist model was extended to a *direct contact induction model* (DCIM) by Salazar *et al.* (Salazar-Ciudad, Garcia and Solé, 2000; Salazar-Ciudad, Newman and Solé, 2001; Salazar-Ciudad, Solé, and Newman, 2001). In this model, N_c cells are arranged in a line or $N_c \times N_c$ cells are arranged on a square lattice. Each cell has the same network of N_g genes; N_h of the genes code for hormones (meaning molecules that affect gene expression in other cells), N_r genes code for cellular or nuclear receptors of the hormones and the rest are transcription factors or proteins involved in signal transduction inside the cell. The model assumes that the change in the activity or chance of transcription induced by interaction follows a saturating function of Hill shape. Enhancers and protein-binding sites are characterized by weights $W_{jk} \in [-1, +1]$, binding of hormones and receptors by $W_{hl}^* \in [0, 1]$. In the DCIM, it is assumed that the hormones are membrane-bound and then can only effect the immediate neighbors.

The authors now study randomly generated networks, which are modeled by a set of coupled, nonlinear partial differential equations. The change of activity of the j th

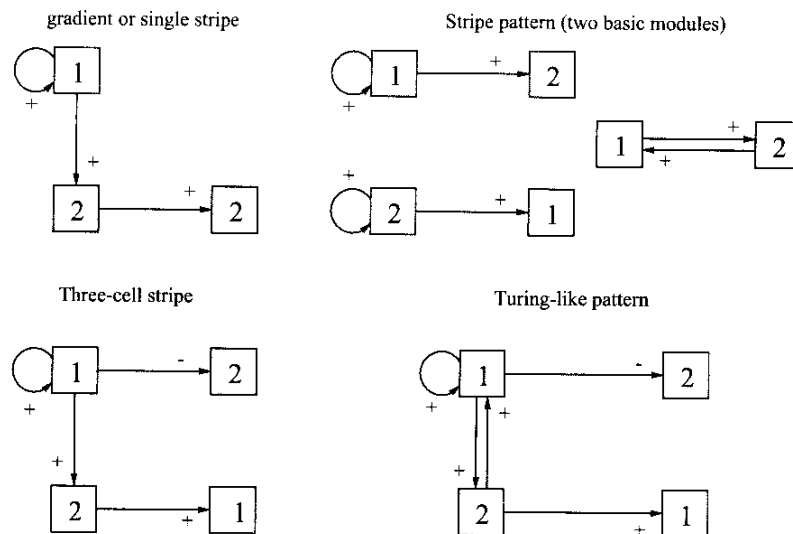


Figure 5.2: Minimal gene networks (“modules”) in the DCI model studied by Salazar *et al.*, intra-cellular interactions plotted only at the left side, cell-to-cell inputs the left to right (after Solé, Salazar-Ciudad and Garcia-Fernandez (2002)).

gene in the i th cell in the DM is modeled by:

$$\frac{\partial g_{ij}}{\partial t} = \frac{\Theta \left[\sum_{k=1}^{N_g} W_{jk} g_{ik} \right]}{K_m + \Theta \left[\sum_{k=1}^{N_g} W_{jk} g_{ik} \right]} - \mu g_{ij} \quad (5.13)$$

for the non-receptor gene products, where $i = 1, \dots, N_c$ and $j = N_h + 1, \dots, N_g$, and by

$$\frac{\partial g_{il}}{\partial t} = \frac{h_{il}}{K_m + h_{il}} - \mu g_{il} \quad (5.14)$$

with

$$h_{il} = \Theta \left[\sum_{k=1}^{N_g} W_{lk} g_{ik} + \sum_{k=1}^{N_h} W_{hl}^* (g_{i-1,l} + g_{i+1,l}) \right] \quad (5.15)$$

for the receptors, where $i = 1, \dots, N_c$ and $l = N_g - N_r + 1, \dots, N_g$.

μ is the rate of protein degradation, K_m is the binding rate constant, $D_l \in [0, 1]$ is the diffusion rate constant of the i -th hormone (randomly generated).

Salazar-Ciudad, Garcia and Solé studied ensemble statistics of these randomly generated networks, starting dynamics from random initial conditions, and explored the space of possible spatio-temporal patterns and the structure of the networks which generated them. One main result of the studies is that inductive interactions lead

to a larger number of stable patterns than diffusive models. Typical spatio-temporal patterns displayed by the DCIM are shown in Fig. 5.1. The authors also identified a number of minimal gene networks underlying the observed patterns (Solé, Salazar-Ciudad and Garcia-Fernandez (1999, 2002), cf. Fig. 5.2). The structure of networks that generate spatially inhomogeneous patterns can be assigned to one of two different classes: networks with *random* (“emergent”) topology or networks with *hierarchical* organization. Emergent networks generally produce the more complex patterns, e.g. when the number of stripes in the pattern is larger than four, the network tends to be emergent, whereas when the number of stripes is smaller, it is usually hierarchic. The range of phenotypic variation in hierarchical networks is much more restricted than in emergent ones. Furthermore, the authors find that hierarchical networks are *more robust* with respect to mutations, i.e. they are more likely to produce the same target pattern under change of some of the W_{ij} than emergent networks, and they produce more finely tuned patterns. Thus, the authors argue, in the course of evolution emergent networks are likely to be gradually replaced by hierarchical networks. The authors finally use this approach to model the evolution of segmentation mechanisms as in the *Drosophila* embryo (Salazar-Ciudad, Solé, and Newman, 2001; Szathmary, 2001), and hypothesize that an original gene network of emergent type with a clock-like behavior has been replaced by the genetic hierarchy seen in modern-day *Drosophila* in the course of evolution.

To summarize this part, different theoretical approaches have proven that morphogenesis based on local cell-cell communication and cell-internal information processing by gene regulatory networks is at least as efficient as pattern formation by reaction-diffusion systems. Furthermore, dynamics of regulatory networks with appropriate topologies contributes significantly to *robustness* of pattern formation (von Dassow *et al.*, 2000; Salazar-Ciudad, Garcia and Solé, 2000; Solé, Salazar-Ciudad and Garcia-Fernandez, 2002).

Part II

Dynamics and evolution of
regulatory networks and
morphogenesis: numerical and
analytical studies

Chapter 6

Criticality in Random Threshold Networks: Annealed Approximation and Beyond

6.1 Introduction

Recently, research on complex networks has become increasingly popular in the statistical physicist's community (Strogatz, 2001; Dorogovtsev and Mendes, 2001; Albert and Barabási, 2002). The variety of fields where networks of many interacting units are found, e.g. gene regulation, neural systems, food webs, species relationships in biological evolution, economic interactions and the organization of the internet, gives rise to the question of common, underlying dynamical and structural principles, as well as to the study of simple model systems suitable as a theoretical framework for a wide class of complex systems. Such a theoretical approach is provided, for example, by Random Boolean Networks (RBN), originally introduced by Kauffman to model the dynamics of genetic regulatory networks in biological organisms (Kauffman, 1969, 1993). In this class of models the state $\sigma_i \in \{0, 1\}$ of a network node i is a logical function of the states of k_i other nodes chosen at random. The logical functions are also chosen at random, either with equal probability or with a suitably defined bias. A phase transition with respect to the average number \bar{K} of inputs per node is observed at a critical connectivity $K_c = 2^{-1}$: for $\bar{K} < K_c$ a “frozen” phase is found with short limit cycles and isolated islands of activity, whereas for $\bar{K} \geq K_c$ one finds a “chaotic”

¹This is the critical value for unbiased Boolean functions; a bias p in the choice of the Boolean functions, where p denotes the mean fraction of 1's in the sites outputs, shifts the critical connectivity to $K_c = 1/[2p(1-p)]$.

phase with limit cycles diverging exponentially with system size, and small perturbations (“damage”) propagating through the whole system. While a more detailed understanding of this transition by means of percolation theory (Albert and Barabási, 2000) is still in its infancy, theoretical insight primarily was gained by application of the so-called *annealed approximation* introduced by Derrida and Pomeau (Derrida and Pomeau, 1986) and successive extensions (Bastolla and Parisi, 1996; Solé and Luque, 1995; Luque and Solé, 1996); the analytical study of damage spreading by means of this approximation allows for the exact determination of critical points in the limit of large system sizes N .

Closely related to RBN are Random Threshold Networks (RTN), first studied as diluted, non-symmetric spin glasses (Derrida and Flyvbjerg, 1987) and diluted, asymmetric neural networks (Derrida and Gardner, 1987; Kree and Zippelius, 1987). Networks of this kind also show complex dynamical behavior similar to RBN at a critical connectivity K_c . For the case where the number of inputs per node K is constant, theoretical insight could be obtained in the framework of the annealed approximation (Kürten, 1988a,b), but the extension of these results to more realistic szenarios, where the number of inputs per node is allowed to vary, fails because of the complexity of the analytical expressions involved.

In the following, an alternative approach is followed which allows us to calculate the critical value K_c for RTN with discrete weights and zero threshold, but with non-constant number of inputs per node. Starting from a combinatorical investigation, an additional degree of complexity in this type of RTN is identified, which is not found in RBN: damage propagation depends on the *in-degree* of signal-receiving sites. The introduction of the *average probability of damage propagation* allows to apply the annealed approximation and leads to the surprising result, that this class of random networks has a critical connectivity $K_c < 2$, contrary to RBN where one always has $K_c \geq 2$. Numerical evidence is presented supporting the analytical results. Finally, a short outlook on the complexity of phenomena in RTN near K_c is given, which are beyond the scope of the annealed approximation, and it will be discussed how the results of this study could relate to properties of real-world networks, such as robustness (Barkai and Leibler, 1997; Alon *et al.*, 1999; Albert, Jeong, and Barabási, 2000; Bornholdt and Sneppen, 2000) and non-trivial degree-distributions (Barabási and Albert, 1999; Strogatz, 2001).

6.2 Random Threshold Networks

A Random Threshold Network (RTN) consists of N randomly interconnected binary sites (spins) with states $\sigma_i = \pm 1$. For each site i , its state at time $t + 1$ is a function

of the inputs it receives from other spins at time t :

$$\sigma_i(t+1) = \text{sgn}(f_i(t)) \quad (6.1)$$

with

$$f_i(t) = \sum_{j=1}^N c_{ij} \sigma_j(t) + h. \quad (6.2)$$

The N network sites are updated *in parallel*. In the following discussion the threshold parameter h is set to zero. The interaction weights c_{ij} take discrete values $c_{ij} = \pm 1$, $+1$ or -1 with equal probability. If i does not receive signals from j , one has $c_{ij} = 0$. The in-degree k_i of site i thus is defined as the number of weights c_{ij} with $c_{ij} \neq 0$. If \bar{K} denotes the average connectivity of the network, for large N the statistical distribution of in- and out-degrees follows a Poissonian:

$$\text{Prob}(k_i = k) = \frac{\bar{K}^k}{k!} e^{-\bar{K}} \quad (6.3)$$

This corresponds to the case where each weight has equal probability $p = \bar{K}/N$ to take a non-zero value.

6.3 Damage Spreading in RTN: dependence on the in-degree k

The most convenient way to distinguish the ordered (frozen) and the chaotic phase of a discrete dynamical network is to study the so-called *damage spreading*: for $\bar{K} < K_c$, a small local perturbation, e.g. changing the state of a single site $\sigma_i \rightarrow -\sigma_i$, vanishes, whereas above K_c it percolates through the network. In RBN, the probability p_s that a site i propagates the damage when a single input j is changed ($1 \rightarrow 0$ or $0 \rightarrow 1$) does not depend on the in-degree k of site i : if the possible 2^{2^k} Boolean functions of the k inputs are chosen with equal probability, one has $p_s = 1/2$; if the Boolean functions are chosen with a bias p , where p denotes the mean percentage of 1's in the output of i , one has $p_s = 2p(1-p)$ (Derrida and Pomeau, 1986). Hence, in RBN damage spreading only depends on the *out-degree* of the perturbed site. In RTN, the situation turns out to be more complex: here we find that damage spreading strongly depends on the *in-degree* of the sites. In the following, we will use combinatorial considerations to derive the exact distribution $p_s(k)$ for RTN.

Consider a site i having k arbitrary input spins, $k \in \{0, 1, 2, \dots, N\}$. Let k_+ denote the number of spins equal to $+1$, k_- the number of spins equal to -1 , hence $k = k_+ + k_-$.

input spins $c_{ij}\sigma_j(t)$	$\sigma_i(t+1)$	$\sigma'_i(t+1)$	# input config.
$\uparrow \uparrow \uparrow$	\uparrow	\uparrow	$\binom{3}{0}$
$\uparrow \uparrow \downarrow$	\uparrow	\downarrow	$\binom{3}{1}$
$\uparrow \downarrow \downarrow$	\downarrow	\uparrow	$\binom{3}{2}$
$\downarrow \downarrow \downarrow$	\downarrow	\downarrow	$\binom{3}{3}$

Figure 6.1: Combinatorial derivation of $p_s(k)$, here for $k = 3$. Reversal of the thick-lined inputs $c_{ij}\sigma_j(t)$ leads to a spin-flip of site i , i.e. $\sigma'_i(t+1) = -\sigma_i(t+1)$ (thick-lined vectors). The total number of input configurations is 8 (see the right column showing the corresponding numbers of spin configurations), hence the number of possible spin-flips is $3 \cdot 8 = 24$. The total number of input-spin reversals leading to damage propagation is $2 \cdot (3 + 3) = 12$. Thus we find $p_s(3) = 1/2$. The generalization for $k = 5, 7, 9, \dots$ is straight-forward.

The state $\sigma_i(t+1)$ of site i at time $t+1$ is given by eqn. (6.1) and (6.2). Let us first calculate the probability $p_s(k)$ that a change of the sign of one arbitrary input spin at time t reverses the sign of i 's output at $t+1$, i.e. that it leads to $\sigma'_i(t+1) = -\sigma_i(t+1)$. The number of possible configurations of k spins is 2^k ; as in each of these configurations k spins can be flipped (reversed in sign), the total number of possible spin-flips is

$$Z_{total} = k \cdot 2^k. \quad (6.4)$$

Thus p_s is defined as the number of spin reversals leading to $\sigma'_i(t+1) = -\sigma_i(t+1)$, divided by Z_{total} .

For $k = 1$ it is easy to see that $p_s = 1$, whereas for $k = 2$ one gets $p_s = 1/2$. For $k \geq 3$ we have to analyze even and odd k separately:

Odd k : By the definition of the transition function for site i (eqn. (6.2)) one recognizes that there are only two configurations in which a flip of a single input spin at time t can lead to a spin-flip of site i at time $t+1$; these configurations are given by the condition:

$$|k_+ - k_-| = 1. \quad (6.5)$$

If $k_+ = k_- + 1$ (i.e. $\sigma_i(t+1) = 1$), the reversal of a positive input spin always leads to $\sigma'_i(t+1) = -1$, whereas flips of negative input spins do not change the state of site i (for $k_+ = k_- - 1$ it is vice versa). Thus in both configurations $(k+1)/2$ spin flips of

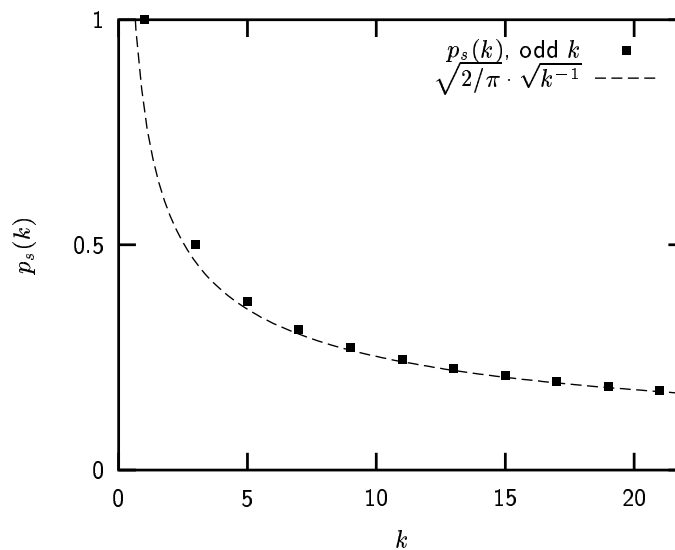


Figure 6.2: Local probability for damage propagation, $p_s(k)$, as a function of the in-degree k of the signal-receiving site, for odd k . The squares correspond to the exact result of eqn. (6.7), the dashed curve shows the Stirling approximation of eqn. (6.13).

k possible lead to $\sigma'_i(t+1) = -\sigma_i(t+1)$. The total number of configurations fulfilling eqn. (6.5) is

$$Z_{\text{oddflip}} = \binom{k}{(k-1)/2} + \binom{k}{(k+1)/2} = 2 \cdot \binom{k}{(k+1)/2}. \quad (6.6)$$

Hence the number of spin-flips leading to damage spreading at site i is $Z_{\text{oddflip}} \cdot (k+1)/2$ and we obtain for $k = 3, 5, 7, \dots$

$$p_s(k) = \frac{Z_{\text{oddflip}} \cdot (k+1)}{2 \cdot Z_{\text{total}}} = \frac{(k+1) \cdot \binom{k}{(k+1)/2}}{k \cdot 2^k} \quad (6.7)$$

Fig. 6.1 demonstrates the above derivation at the example $k = 3$. $p_s(k)$ for odd k is shown in Fig. 6.2.

Even k : For even k , there are also only two configurations in which changing a single input of site i at time t leads to $\sigma'_i(t+1) = -\sigma_i(t+1)$, given by the conditions

$$k_+ = k_- \quad (*) \quad \text{or} \quad k_- = k_+ + 2 \quad (**). \quad (6.8)$$

In the case $k_+ = k_- + 2$ neither the reversal of a positive nor the reversal of a negative input-spin changes the output of i , due to the signum function in eqn. (2). In the case (*), $k/2$ spin-flips ($+1 \rightarrow -1$) change the output of site i ($+1 \rightarrow -1$), in the case (**)

$(k+2)/2$ spin-flips ($-1 \rightarrow +1$) do so. The number of configurations fulfilling eqn. (8) is given by

$$Z_{\text{evenflip}} = Z(*) + Z(**) = \binom{k}{k/2} + \binom{k}{(k+2)/2} \quad (6.9)$$

So we finally obtain for $k = 2, 4, 6, \dots$

$$p_s(k) = \frac{k \cdot Z(*) + (k+2) \cdot Z(**)}{2 \cdot Z_{\text{total}}} \quad (6.10)$$

$$= \frac{k \cdot \binom{k}{k/2} + (k+2) \cdot \binom{k}{(k+2)/2}}{k \cdot 2^{k+1}} \quad (6.11)$$

$$= p_s(k+1), \quad (6.12)$$

as can be seen by some simple algebra.

Using the Stirling formula for $k!$, one recognizes that the asymptotic behavior of $p_s(k)$ for odd k is given by

$$p_s(k) = \frac{1}{k+1} \cdot \sqrt{\frac{2k}{\pi}} \cdot \left(\frac{k}{\sqrt{k^2-1}} \right)^k \approx \sqrt{\frac{2}{\pi}} \cdot \frac{1}{\sqrt{k}}. \quad (6.13)$$

Thus it turns out that

$$\lim_{N \rightarrow \infty} \lim_{k \rightarrow N} p_s(k) = 0. \quad (6.14)$$

However, this does *not* mean that there is no damage spreading for $k \rightarrow N$: in each timestep, damage increases $\propto k \cdot p_s(k) \sim \sqrt{k} \sim \sqrt{N}$ in this limit, which ensures the existence of a “chaotic” regime.

6.4 Average probability for damage spreading

Choosing an arbitrary network site i with at least one input and changing the sign of one input-spin will lead to a spin-flip of site i with probability $p_s(k)$, dependent on the number of inputs k . Now we are interested in the *expectation value* of this probability, i.e. the average probability for damage spreading $\langle p_s \rangle(\bar{K})$ for a given average network connectivity \bar{K} . For large N , this problem is equivalent to connecting a new site j with state σ_j to an arbitrary site i of a network of size $N-1$ with average connectivity \bar{K} ; thus k_i is increased by one. Hence, changing the sign of σ_j at time t will lead to a different output of site i at time $t+1$ with probability $p_s(k_i+1)$. Using the Poisson approximation and taking the thermodynamic limit $N \rightarrow \infty$, this leads to

$$\langle p_s \rangle(\bar{K}) = e^{-\bar{K}} \sum_{k=0}^{\infty} \frac{\bar{K}^k}{k!} \cdot p_s(k+1) \quad (6.15)$$

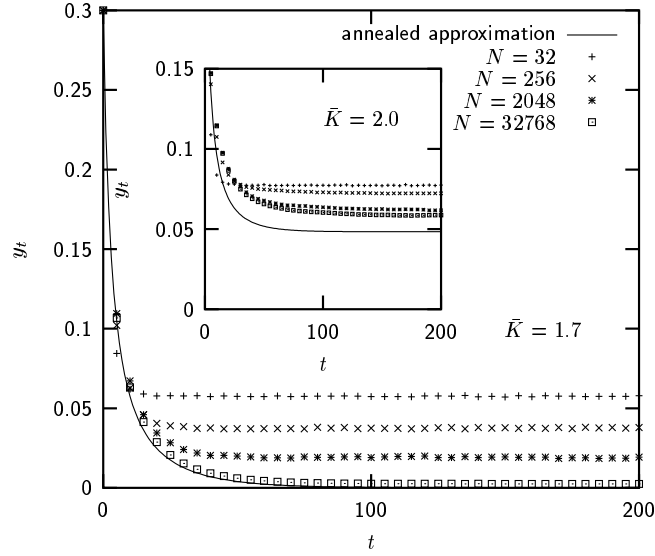


Figure 6.3: Time evolution of the relative Hamming distance y_t of two spin configurations with $y_0 = 0.3$ for different network sizes N with $\bar{K} = 1.7$ (numerical simulations, ensemble statistics over 10000 networks / 400 networks ($N = 32768$)). For large N , one finds convergence against the prediction of the annealed approximation (solid curve). The inset shows the same for overcritical networks ($\bar{K} = 2.0$).

Splitting the sum for even and odd k yields

$$\langle p_s \rangle(\bar{K}) = e^{-\bar{K}} \left\{ 1 + \sum_{i=1}^{\infty} \frac{\bar{K}^{2i-1}}{(2i-1)!} \cdot p_s(2i) + \sum_{i=1}^{\infty} \frac{\bar{K}^{2i}}{(2i)!} \cdot p_s(2i+1) \right\} \quad (6.16)$$

Using the relation $p_s(k) = p_s(k+1)$ for even k one obtains

$$\langle p_s \rangle(\bar{K}) = e^{-\bar{K}} \cdot \left\{ 1 + \sum_{i=1}^{\infty} \bar{K}^{2i-1} p_s(2i+1) \cdot \left(\frac{1}{(2i-1)!} + \frac{\bar{K}}{(2i)!} \right) \right\} \quad (6.17)$$

Inserting (7) into this equation finally yields

$$\langle p_s \rangle(\bar{K}) = e^{-\bar{K}} \left\{ 1 + \frac{1}{2} \sum_{i=1}^{\infty} \frac{1}{(i!)^2} \left(\frac{1}{2i} + \bar{K} \right) \cdot \left(\frac{\bar{K}}{2} \right)^{2i-1} \right\}. \quad (6.18)$$

Formula (6.18) is the central result of this chapter, as it allows for an analytic calculation of the critical connectivity K_c of RTN. For practical use it is worth to notice that the sum in eqn. (6.18) converges very fast, setting the upper limit to ten is sufficient for a relative error $\Delta \langle p_s \rangle / \langle p_s \rangle < \mathcal{O}(10^{-4})$.

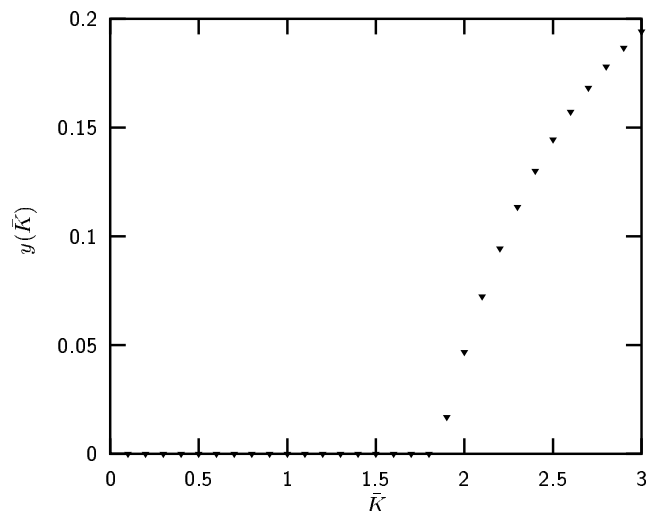


Figure 6.4: The normalized distance y for $t \rightarrow \infty$ as a function of \bar{K} , calculated by numerical solution of eqn. (6.23). One finds $y(\bar{K}) > 0$ for $\bar{K} \geq K_c = 1.849$.

6.5 Calculation of K_c by Derrida's Annealed Approximation

The still most powerful analytical approach to describe damage spreading in random networks of automata - and thus to calculate the critical value K_c of \bar{K} - is the so-called *annealed approximation* introduced by Derrida and Pomeau (1986). This approximation, originally applied to Kauffman networks (Boolean networks with constant number K of inputs per node), neglects the fact that the (Boolean) functions and the interactions between the spins are quenched, i.e. constant over time, and instead randomly reassigns inputs and functions to all spins at each time step. As a central result of this approximation Derrida and Pomeau derived a recursion formula which describes the time evolution of the Hamming distance $d(\vec{\sigma}, \vec{\sigma}')$ of two spin configurations $\vec{\sigma}$ and $\vec{\sigma}'$. The normalized distance $y = d(\vec{\sigma}, \vec{\sigma}')/N$ at time $t + 1$ for $p_s = 1/2$ and in-degree k is given by the recursion

$$y_{t+1} = \frac{1 - (1 - y_t)^k}{2}. \quad (6.19)$$

They also give a straight-forward generalization to networks where k is not constant:

$$y_{t+1} = \sum_k \rho_k \frac{1 - (1 - y_t)^k}{2}, \quad (6.20)$$

where ρ_k is the probability that a network site has k inputs. In the case of an average probability for damage propagation $\langle p_s \rangle(\bar{K})$ depending on \bar{K} , as for the RTN discussed

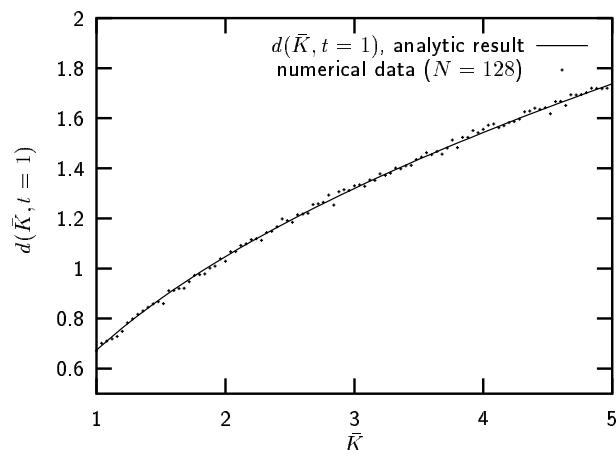


Figure 6.5: Average Hamming distance d_t for two configurations $\vec{\sigma}$ and $\vec{\sigma}'$ differing in one bit at time $t - 1$ as a function of the average connectivity (ensemble average over 10000 random networks with $N = 128$ for each data point). One finds excellent agreement with the prediction of the annealed approximation (solid curve, $d_{t,annealed} = \langle p_s \rangle(\bar{K}) \cdot \bar{K}$).

in this article, this generalizes to

$$y_{t+1} = \langle p_s \rangle(\bar{K}) \sum_k \rho_k \cdot [1 - (1 - y_t)^k]. \quad (6.21)$$

In the following, we focus our discussion on random networks, where each link has equal probability to take a non-zero value (as first introduced by Erdős and Rényi, 1960). Thus, for average connectivities $\bar{K} \ll N$, ρ_k is given by a Poissonian, leading to

$$\begin{aligned} y_{t+1} &= \langle p_s \rangle(\bar{K}) e^{-\bar{K}} \sum_k \frac{\bar{K}^k}{k!} \cdot [1 - (1 - y_t)^k] \\ &= \langle p_s \rangle(\bar{K}) e^{-\bar{K}} \left(\sum_k \frac{\bar{K}^k}{k!} - \sum_k \frac{\bar{K}^k}{k!} \cdot (1 - y_t)^k \right) \\ &= \langle p_s \rangle(\bar{K}) \left(1 - e^{-\bar{K}} \sum_k \frac{[\bar{K} (1 - y_t)]^k}{k!} \right) \\ &= \langle p_s \rangle(\bar{K}) (1 - e^{-\bar{K} y_t}). \end{aligned} \quad (6.22)$$

Fig. 6.3 compares the time evolution of the Hamming distance, measured in numerical simulations of RTN of different size N , to the prediction of eqn. (22).

As well for subcritical as for supercritical \bar{K} one finds convergence against the annealed approximation, but for $\bar{K} \geq K_c$ the convergence is quite slow. In the limit $t \rightarrow \infty$ we expect that the normalized distance $y(\bar{K})$ evolves to a constant

value, i.e. $y_{t+1} = y_t = y$. Inserting this condition in eqn. (6.22) leads to the fixed point equation

$$f(y) \equiv y - \langle p_s \rangle(\bar{K}) (1 - \exp[-\bar{K} \cdot y]) = 0. \quad (6.23)$$

Solutions y^* of eqn. (6.23) for some values of the average connectivity \bar{K} are shown in Fig. 4; evidently, there exists a critical connectivity K_c above which the Hamming distance of two initially nearby trajectories increases to a non-zero value, indicating damage spreading through the network. The exact value K_c can be obtained easily by a linear stability analysis of eqn. (6.23). $y_0^* = 0$ is always a solution of (6.23), but is attractive (stable) only for $\bar{K} < K_c$.

This fixed point is stable only if

$$\lim_{\varepsilon \rightarrow 0} \left| \frac{df(y)}{dy} \right| (y_0^* + \varepsilon) < 1. \quad (6.24)$$

One has

$$\left| \frac{df(y)}{dy} \right| (y_0^* + \varepsilon) = \langle p_s \rangle(\bar{K}) \cdot \bar{K} \cdot \exp[-\bar{K} \cdot \varepsilon], \quad (6.25)$$

i.e.

$$\lim_{\varepsilon \rightarrow 0} \left| \frac{df(y)}{dy} \right| (y_0^* + \varepsilon) = \langle p_s \rangle(\bar{K}) \cdot \bar{K}. \quad (6.26)$$

Inserting this result into (6.24) we find that $y_0^* = 0$ is attractive if

$$\langle p_s \rangle(\bar{K}) \cdot \bar{K} < 1. \quad (6.27)$$

Thus the critical connectivity K_c can be obtained by solving the equation

$$\langle p_s \rangle(K_c) \cdot K_c = 1, \quad (6.28)$$

$\langle p_s \rangle(\bar{K})$ given by eqn. (6.18).

Eqn. (6.27) and (6.28) have a simple interpretation in terms of damage spreading: $\langle p_s \rangle(\bar{K}) \cdot \bar{K}$ is the expectation value $\langle d_t \rangle$ of the Hamming distance - i.e. of the damage - at time t , if at time $t - 1$ a single spin is reversed, as this spin on average has \bar{K} outputs to other spins, which all will propagate damage with probability $\langle p_s \rangle(\bar{K})$. Fig. 6.5 shows that the prediction $\langle d_{t,annealed} \rangle$ of our annealed approximation agrees perfectly well with the average d_t measured between two spin configurations in RTN with $d_{t-1} = 1$ (ensemble statistics).

A numerical solution of eqn. (6.28) for the RTN discussed here yields the critical value

$$K_c = 1.849 \pm 0.001. \quad (6.29)$$

Thus we get the remarkable result that in RTN with $h = 0$ and discrete interaction weights, marginal damage spreading - i.e. the percolation transition from frozen to chaotic dynamics - is found *below* $\bar{K} = 2$, in contrast to RBN where one always has $K_c \geq 2$.

6.6 Beyond the Annealed Approximation: Complexity in RTN

So far we presented numerical evidence that the annealed approximation correctly predicts the *average* damage spreading behavior in RTN and thus the critical connectivity K_c in the limit of large system sizes, however, this coarse-grained approach of course does not capture the whole complexity of the network dynamics near criticality. Concerning the statistics of damage spreading, even for quite large N scale-free distributions are found in a certain range around K_c ; skewed, super-critical distributions, with an increasing maximum moving towards $N/2$ with increasing \bar{K} , are found for $\bar{K} \geq 2.1$ ($N = 8192$, Fig. 6.6). In the ordered regime, the distributions decay exponentially. Presumably, the good convergence of the average Hamming distance measured in numerical simulations for $\bar{K} < K_c$ against the annealed approximation directly reflects averaging over these exponential distributions with well-defined characteristic scale, whereas averaging over the the scale-free damage-distributions around K_c leads to the observed weak finite-size scaling $\propto 1/\log(N)$ against the analytical result (Fig. 6.3) at the order-chaos transition. Furthermore, for finite N , even the transition from critical (scale-free) to supercritical distributions (Gaussian distributions in the limit $\bar{K} \rightarrow N$) is rather smooth: even deep in the “chaotic” regime (e.g. $\bar{K} = 2.5$, Fig. 6.6) damage becomes zero for more than 60% of the initial conditions/ the networks tested, and the damage distributions are clearly bimodal.

Attractor periods of RTN are found to be power-law distributed in a certain range of connectivities around K_c , due to dynamical correlations between network sites, which are neglected completely in the framework of Derrida’s annealed approximation. We shall briefly discuss this: The *average correlation* $\text{Corr}(i, j)$ of a pair (i, j) of sites is defined as the average over the products $\sigma_i(t)\sigma_j(t)$ in dynamical network evolution between two distinct points of time T_1 and T_2 :

$$\text{Corr}(i, j) = \frac{1}{T_2 - T_1} \sum_{t=T_1}^{T_2} \sigma_i(t)\sigma_j(t) \quad (6.30)$$

The *global average correlation* can be defined as

$$\langle \text{Corr} \rangle_g = \frac{1}{N(N-1)} \sum_{i=1}^N \sum_{j=1}^N |\text{Corr}(i, j)| \quad \text{with } i \neq j, \quad (6.31)$$

neglecting trivial auto-correlations.

In the ordered (frozen) regime, there are few, isolated (non-frozen) clusters with dynamical activity, but their dynamics is *not correlated*. A good measure here is

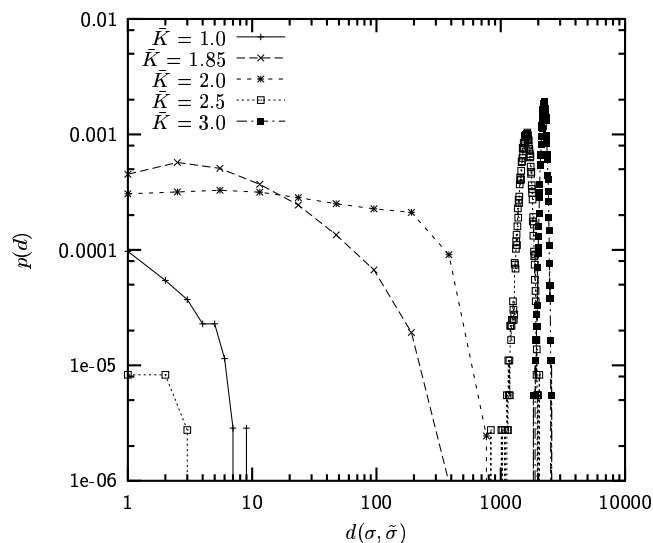


Figure 6.6: Statistical distributions $p(d)$ of the Hamming distances $d(\sigma, \tilde{\sigma})$ for two spin configurations $\sigma, \tilde{\sigma}$ with $d(t=0) = 0.3 \cdot N$. Only the tails of the distributions are shown in this log-log-plot. Ensemble statistics was taken after 100 dynamical updates over 5000 RTN with size $N = 8192$ and Poissonian distributed in- and out-degree, testing 20 different initial conditions for each network. In the ordered regime ($\bar{K} = 1$) damage is suppressed exponentially. Near criticality and slightly above ($\bar{K} = 1.85, \bar{K} = 2.0$) the tails approximately obey power-laws (log-binned data). For $2.1 < \bar{K} < 2.6$, the distributions are bimodal (shown here for $\bar{K} = 2.5$), with an increasing maximum at large damage values. For larger \bar{K} , the tails of the distributions become Gaussian, with a maximum approaching $d = N/2$ with increasing \bar{K} . Not visible in this plot is the pronounced maximum of all distributions at $d = 0$: even for $\bar{K} = 3.0$ damage finally becomes zero for about 56% of the networks/ the initial conditions tested; this maximum vanishes for $\bar{K} \rightarrow N$.

the *average correlation of non-correlated sites* $\langle \text{Corr} \rangle_{nc}$, i.e. the average correlation of pairs of sites with $0 \leq \text{Corr}(i, j) < 1$ (Fig. 6.7). $\langle \text{Corr} \rangle_{nc}$ is almost zero in the ordered regime. Due to this uncorrelated dynamics of a few “active islands” in a frozen network the distributions of attractor periods P decay $\propto \exp(-P)$. Near K_c , however, $\langle \text{Corr} \rangle_{nc}$ shows a pronounced maximum, i.e. the dynamics of active clusters becomes strongly correlated; thus, at this percolation transition, attractor periods show scale-free distributions. In the chaotic regime, damage spreading destroys most of the correlations, consequently, $\langle \text{Corr} \rangle_{nc}$ decays once again.

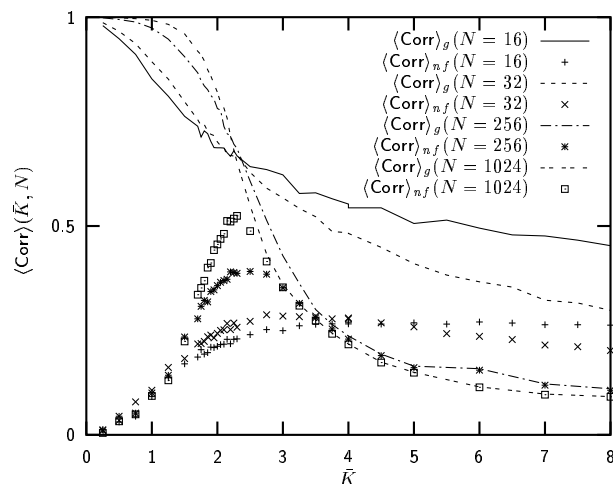


Figure 6.7: Global average correlation $\langle \text{Corr} \rangle_g$ (lines) and average correlation of non-correlated pairs $\langle \text{Corr} \rangle_{nc}$ (crosses or points) as a function of the average connectivity \bar{K} for four different system sizes N . For each value of \bar{K} sampled ensemble-averages were taken over 1000 RTN. For each network the average was taken over a subset of 5000 randomly chosen pairs of sites. Individual runs were limited to $T_{max} = 20000$.

6.7 Discussion

In this chapter, damage spreading in Random Threshold Networks (RTN) with zero threshold and discrete weights was investigated. This kind of discrete dynamical network shows complex dynamics similar to Boolean networks, with a transition from ordered to chaotic dynamics at a critical average connectivity K_c . Using combinatorial methods, the exact distribution $p_s(k)$ for local damage propagation was derived. It was shown that Derrida’s annealed approximation can be applied to this class of models; this theoretical analysis yielded the surprising result $K_c = 1.849$, in contrast to RBN, where $K_c \geq 2$. The ansatz proposed in this study could offer a road-map for an analytical treatment of similar systems with additional complexity (non-discrete weights, $h \neq 0$).

An interesting result of our studies is that in dynamical networks with sparse asymmetric interactions and some kind of threshold characteristics governing dynamics, a high number k of inputs per node *can stabilize dynamics*, as the effect of single, local errors is reduced like $1/\sqrt{k}$; this, however, is counteracted by the overall topological randomness in the network, which allows for ordered dynamics only at small *average* connectivities \bar{K} . This effect of “local damage suppression” at nodes with high in-degree is demonstrated in Fig. 6.8, directly comparing damage distributions in RTN and RBN: damage spreading was investigated for networks with “flat”, scale-free in-degree dis-

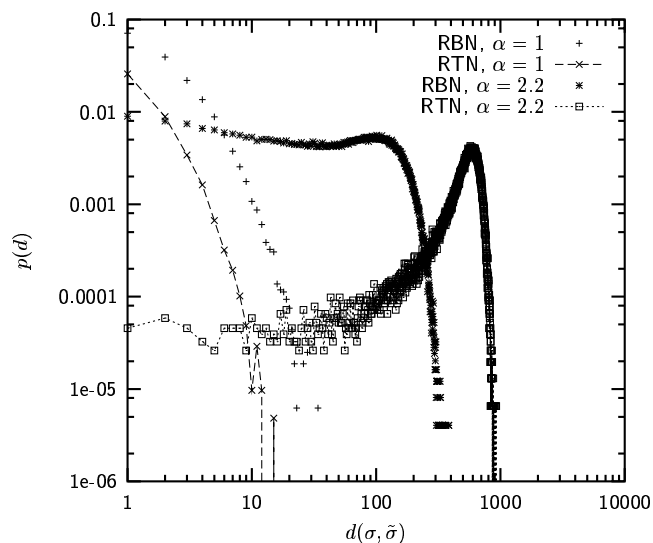


Figure 6.8: Statistical distributions $p(d)$ of the Hamming distances $d(\sigma, \tilde{\sigma})$ for two spin configurations $\sigma, \tilde{\sigma}$ with $d(t=0) = 0.3 \cdot N$. Only the tails of the distributions are shown in this log-log-plot. Ensemble statistics was taken after 100 dynamical updates over 10000 RTN or RBN, respectively, with size $N = 2048$ and average connectivity $\bar{K} = 2.5$, Poissonian distributed out-degree and in-degree-distribution $p(k_{in}) \propto k_{in}^{-\alpha}$. The choice $\alpha = 1$ leads to clustered, hierarchical networks with ordered dynamics, but in RTN damage spreading is stronger suppressed: $p(d)$ shows a faster exponential decay than in RBN. This is a direct effect of the local damage suppression due to highly connected nodes in RTN. For $\alpha = 2.2$, the opposite is observed: Whereas RBN show an only slightly overcritical $p(d)$ with a power-law regime prevailing over almost two decades, RTN show a strongly skewed, overcritical $p(d)$. In this regime the local damage suppression effect $\propto \sqrt{k_{in}}$ is not strong enough to counteract chaotic dynamics increasing $\propto k_{out}$.

tributions $\propto k_{in}^{-\alpha}$, whereas the out-degree follows a Poissonian. For very flat in-degree distributions ($\alpha = 1$), both RTN and RBN have a hierarchical, clustered structure and show ordered dynamics, but in RTN damage is stronger suppressed than in RBN, due to local damage suppression at highly connected nodes. For steeper in-degree distributions ($\alpha = 2.2$ in Fig. 6.8) this effect is not strong enough, and the dynamics in RTN is even “more chaotic” in RTN than in RBN. Concerning models including *self-organization of network topology* this could have important consequences. Topological evolution of RTN by a local coupling of control- and order parameters (connectivity and magnetization/correlations of network sites) was introduced in Bornholdt and Rohlf (2000). Does “damage suppression” at highly connected nodes affect self-organization of network structure in such models, e.g. leading to broader degree-distributions that deviate from a Poissonian? To answer this question, in the next chapter a simplified

version of this model will be studied, which can be regarded as a kind of “toy-model” for evolution of gene regulatory networks based on local “rewiring” of regulatory interactions. The results are compared to an analysis of statistical data available for the gene regulatory networks of *E. Coli* and yeast.

Chapter 7

Gene Regulatory Networks: A Discrete Model of Dynamics and Topological Evolution

7.1 Introduction

In the last chapter, the dynamics of damage propagation in Random Threshold Networks was studied with respect to two different types of degree-distributions, following either a Poissonian or a power-law with a given exponent. It was shown that local damage suppression at network nodes with high in-degree can have a global effect on network dynamics, given an adequate network topology (e.g. a scale-free network topology with an exponent $\gamma < 1.5$). Clearly, suppression of update errors, i.e. *robustness* of dynamics is an essential feature of regulatory networks – hence, what types of evolutionary processes can lead to network topologies with this property? Interestingly, statistical data available for gene regulatory networks in several organisms indeed indicate distinct deviations from random graphs of the Erdős-Rényi type.

For example, Thieffry *et al.* (1998) analyzed the gene regulatory network of *E. Coli* at transcription level. The observed statistical distribution of the number of regulatory *outputs*, i.e. of genes controlled by a given regulatory gene, is shown in Fig. 7.1, the inset shows the distribution of the number of regulatory *inputs* a given gene receives. For both distributions the best fit to the data is an exponential. Interpreting genes as nodes and regulatory interactions as edges (links) of a directed graph, these distributions of in- and out-links deviate clearly from the Poissonian expected for a random graph with equal probability for each node to receive a new link. Another interesting feature is the low connectivity (average number of interactions per gene) of

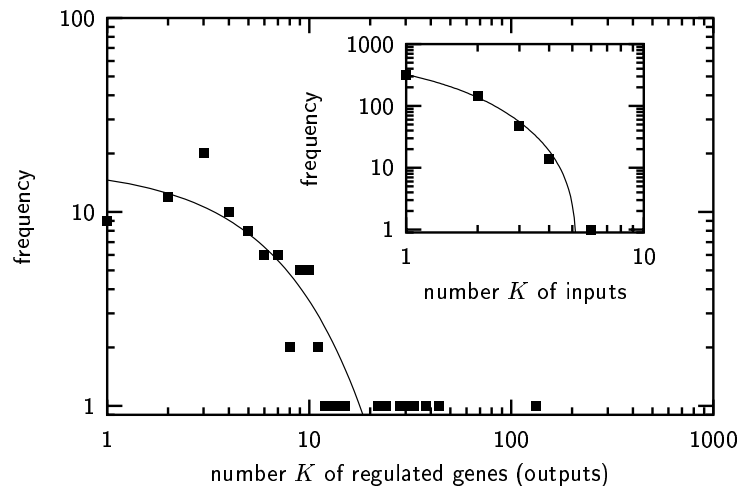


Figure 7.1: Statistical distribution of regulatory outputs, i.e. the number K of genes controlled by a given regulatory gene (*E. Coli*, transcription level), log-log-plot. The inset shows the distribution of regulatory inputs. Data (squares) were taken from (Thieffry *et al.*, 1998), fits proportional to $\exp[-K]$ were added by the authors.

the network, which is between 2 and 3. Similar results have been obtained for other organisms; for the transcriptional regulatory network of yeast, Guelzim *et al.* (2002) found an in-degree distribution following an exponential, and an out-degree distribution following a power-law.

In this chapter we address the question how the observed network topology, here for the example of *E. Coli*, might be related to general dynamical properties of gene regulatory networks. We study a simple evolutionary algorithm in the framework of discrete dynamical networks which mimics functional change and selection at the single-gene level by a local coupling between an order parameter (dynamical activity of a network node) and a control parameter (connectivity). The model investigated here is a simplified version of the very general model of topological self-organization in dynamical networks introduced by Bornholdt and Rohlf (2000). As a result we observe the self-organization of a network topology with an average connectivity between 2 and 3, and with a non-random degree-distribution that shows clear deviations from a Poissonian. The network dynamics is stabilized in a non-chaotic regime by a feedback between control parameter (connectivity) and order parameter (activity of nodes), that allows for robust self-regulation over a wide parameter range.

7.2 Evolution of network topologies based on local dynamical rules

7.2.1 The modeling framework: discrete dynamical networks

Discrete dynamical networks are a theoretical idealization of the dynamics found in gene regulatory networks: the networks nodes (genes) can take only two states, “on” or “off”, depending on regulatory interactions with other genes (at the level of transcription) modeled by weighted connections, which are either activating or inhibitory. Kauffman (1969, 1993) introduced Random Boolean Networks, where the state of a gene is a logical function of its regulatory inputs, whereas more recent studies (Wagner, 1994; Bornholdt and Sneppen, 2000) favor networks, where the state of a gene is determined by a weighted sum of its inputs plus a threshold (Random Threshold Networks, RTN). In our study we use a threshold network defined in the following paragraph.

7.2.2 Definition of the threshold network

Let us consider a network of N randomly interconnected binary elements with states $\sigma_i = \pm 1$. For each site i , its state at time $t + 1$ is a function of the inputs it receives from other elements at time t :

$$\sigma_i(t + 1) = \text{sgn}(f_i(t)) \quad (7.1)$$

with

$$f_i(t) = \sum_{j=1}^N c_{ij} \sigma_j(t) + h. \quad (7.2)$$

The interaction weights c_{ij} take discrete values $c_{ij} = \pm 1$, with $c_{ij} = 0$ if site i does not receive any input from element j . In the following, the threshold parameter h is set to zero. The dynamics of the network states is generated by iterating this rule starting from a random initial condition, using parallel updates of the N network sites.

7.2.3 Network dynamics

RTN show complex dynamics, similar to Random Boolean Networks (Kauffman, 1969). The number of possible system states $\vec{\sigma} = (\sigma_1, \sigma_2, \dots, \sigma_N)$ is 2^N . Hence, applying parallel updates as discussed here, the system dynamics inevitably settles down to a periodic attractor of P repeating states ($1 \leq P \leq 2^N$) after a number T of transient states. The scaling properties of T and P as a function of system size N are highly non-trivial.

In particular, a phase transition is observed at a critical average connectivity $K_c \approx 2$ (Rohlf and Bornholdt, 2002, cf. chapter 6) with lengths of transients and attractors diverging exponentially with system size for an average connectivity larger than K_c . In this dynamical regime, small perturbations spread rapidly through the network, whereas near K_c network dynamics is considerably robust. Furthermore, at K_c the range of possible dynamic behavior is broad, reflected by scale-free distributions of attractor periods and transient lengths with a median scaling proportional to \sqrt{N} . Kauffman postulated that gene regulatory networks may exhibit properties of dynamical networks near criticality (Kauffman, 1969). This would confine network evolution to considerable small average connectivities close to K_c , a hypothesis now becoming testable by modern experimental techniques (e.g. microarrays).

7.2.4 An algorithm for evolving network topology

As a dynamical order parameter we introduce the average activity $A(i)$ of a site i

$$A(i) = \frac{1}{T_2 - T_1} \sum_{t=T_1}^{T_2} \sigma_i(t) \quad (7.3)$$

taken between two arbitrary points of time T_1 and T_2 in the course of network dynamics. Roughly speaking, $A(i)$ characterizes how often (on average) a gene changes between the state of being expressed (active) and being inactive. A value of $A(i)$ close to one indicates an inflexible expression pattern almost constant in time, whereas $A(i) \approx 0$ means that gene i is turned on and off very often in a rather chaotic manner. The first case indicates that the gene is regulated insufficiently, due to missing or bad integration into the regulatory network, whereas the latter case could result from conflicting signals of (too many) regulatory inputs. Both extremes are often hazardous for biological organisms, thus selective pressure would tend to favor mutations which lead to additional regulatory inputs in the first case, and vice versa in the latter case. These simple assumptions are integrated in the following algorithm evolving the topology of our model network (a simplified version of the algorithm introduced in Bornholdt and Rohlf (2000)):

1. Choose a random network with an average connectivity K_{ini} .
2. Choose a random initial state vector $\vec{\sigma}(t=0) = (\sigma_1, \sigma_2, \dots, \sigma_N)$ for the N binary network sites.
3. Calculate the new system states $\vec{\sigma}(t)$ according to eqn. (7.2), using parallel update of the N sites.

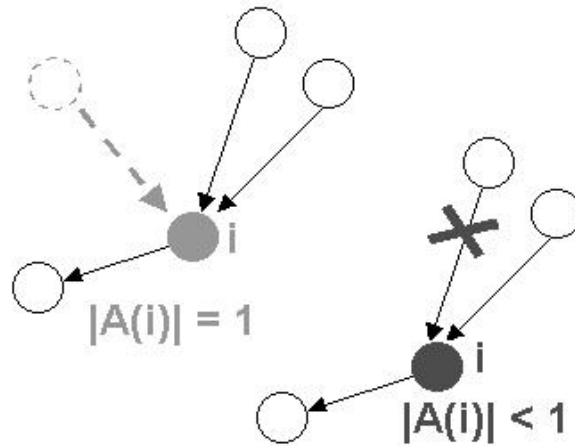


Figure 7.2: The selective criterion leading to topological self-organization: a gene which is regulated insufficiently ($|A(i)| = 1$) receives an additional regulatory input, a gene which tends to show chaotic expression patterns ($|A(i)| < 1$), due to conflicting regulatory signals, loses one of its inputs.

4. After T_u updates the network topology is changed according to the following local rewiring rules:
5. A site i is chosen at random and its average activity $A(i)$ over the last T updates is determined ($T < T_u$).
6. If $|A(i)| = 1$, i receives a new link c_{ij} from a site j selected at random, choosing $c_{ij} = +1$ or -1 with equal probability. If $|A(i)| < 1$, one of the existing non-zero links of site i is set to zero.
7. Go to step number 3. and iterate.

Selective pressure for moderate gene activity between the extremes of either static or chaotic dynamics enters into the model via local rewiring (step 6.): genes with very static expression patterns ($|A(i)| = 1$) receive a new (random) input - this adjusts their dynamics to the overall expression pattern in the network, whereas genes with fluctuating dynamics ($|A(i)| < 1$) reduce their connectivity, which favors simplicity of input patterns.

In step 5. and 6. the *average correlation* $\text{Corr}(i, j)$ between two sites i and j , defined as the average over the products $\sigma_i(t)\sigma_j(t)$ in dynamical network evolution between two distinct points of time T_1 and T_2

$$\text{Corr}(i, j) = \frac{1}{T_2 - T_1} \sum_{t=T_1}^{T_2} \sigma_i(t)\sigma_j(t) \quad (7.4)$$

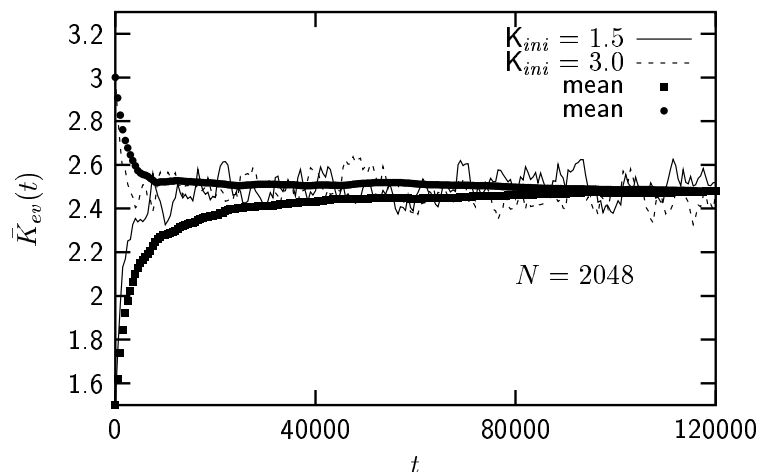


Figure 7.3: Independent from the initial conditions the proposed model leads to a robust evolution of a network topology with average connectivity $\bar{K}_{ev}(t)$ between 2 and 3, shown here for two networks of size $N = 2048$ with different initial connectivities K_{ini} . The mean value over the evolutionary process is $\bar{K} = 2.49$.

can be applied instead of $A(i)$ (results not shown here). If $|\text{Corr}(i, j)| = 1$, i receives a new link c_{ij} from site j , choosing $c_{ij} = +1$ or -1 with equal probability, if $|\text{Corr}(i, j)| < 1$, one of the existing non-zero links of site i or of site j is set to zero.

7.2.5 Results

Independent from the initial conditions (initial state vector and connectivity) we observe a robust evolution of a network topology with an average connectivity \bar{K} between 2 and 3; this is shown in Fig. 7.3 for a network of size $N = 2048$. One finds small, limited fluctuations around the mean value due to the evolutionary dynamics. For large networks \bar{K} approaches the critical value K_c , where a percolation transition from ordered to chaotic dynamics is found (Bornholdt and Rohlf, 2000, cf. also chapter 6 of this thesis.)

For the averaged statistical distribution $p(K)$ of in- and out-links (Fig. 7.4) of the evolved networks one observes a flat exponential decay

$$p(K) \approx p_0 \cdot \exp[-\alpha K], \quad (7.5)$$

with p_0 and α constant, similar to the findings for the transcriptional network of *E. Coli*. This distribution clearly deviates from the Poissonian

$$p_{\text{random}}(K) = \frac{\bar{K}^K}{K!} \exp[-\bar{K}] \quad (7.6)$$

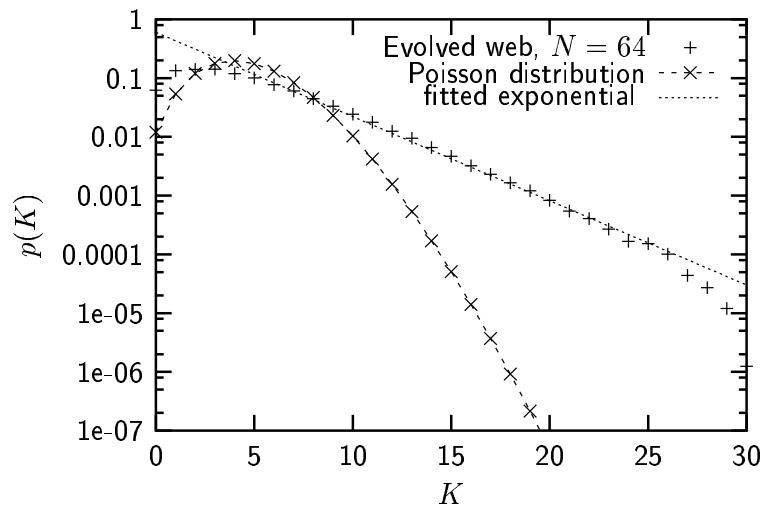


Figure 7.4: Statistical distribution $p(K)$ of the number of regulatory interactions K per gene in the proposed model for a network of size $N = 64$. Compared to the Poisson distribution for random networks with $\bar{K} = 4.46$, it shows a flatter decay proportional to $\exp[-K]$, similar to observations for the transcriptional network of *E. Coli*.

expected for a random network with the same average connectivity \bar{K} .

Since the central step of our topology-evolving algorithm - step 6. - involves a local coupling between a dynamical order parameter (average activity) and a topological control parameter, one would expect a self-organization of global dynamics showing deviations from random networks, corresponding to the evolved non-random structure. Indeed one observes self-organization processes stabilizing the dynamics of the evolving networks in a regime of moderate, non-chaotic dynamics: the ratio between inactive or “frozen” genes and active genes, which occasionally change their expression state, is at an intermediate level, contrary to the step decay of the fraction of frozen genes found for randomly generated networks with increasing \bar{K} (Fig. 7.5).

7.3 Discussion

Discrete dynamical networks have been discussed in theoretical biology as models of general dynamical properties of genetic regulatory networks since about 30 years. In the framework of such models we studied possible mechanisms of structural self-organization of genetic regulatory networks as a consequence of functional change at the single-gene level and selection. We started from the assumption, that for a gene with an inflexible, “frozen” expression pattern biological selection might favor mutations leading to new regulatory inputs (new functions), whereas genes with high number

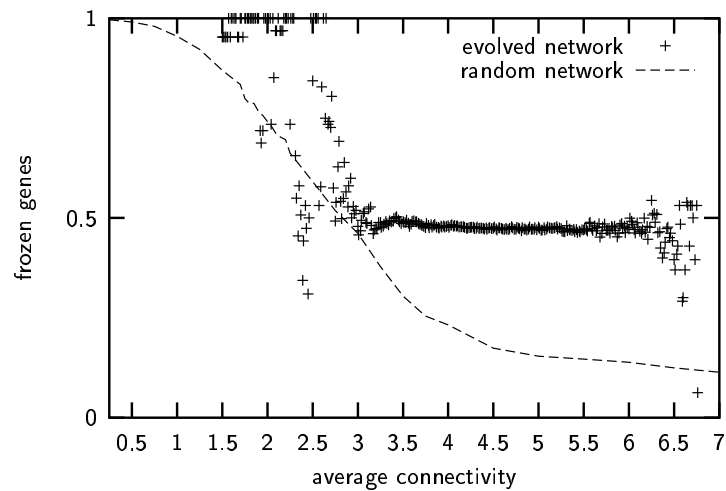


Figure 7.5: The fraction of “frozen” genes as a function of the average connectivity for an evolved network of size $N = 64$ (crosses). The dashed line shows the corresponding curve for random networks.

of (conflicting) regulatory inputs tend to push towards chaotic patterns of activity, thus mutations reducing the number of inputs would be favored by selection. Our model, based on a local coupling between a dynamical order parameter (activity or correlation of genes) and rewiring of regulatory interactions, predicts an average connectivity and non-random network structure well compatible with current experimental data, as it was shown by an analysis of statistical data available for the transcriptional network of *E. Coli* (Thieffry *et al.*, 1998). Hence, one might speculate that similar mechanisms involving an interplay between network dynamics, mutations and selection may play a role in the evolution of regulatory networks of simple organisms.

The selective criterion discussed in this chapter, mimicing functional changes at the single-gene level, was formulated as an abstract, local rewiring rule. However, if regulatory networks have to solve a specific task at a higher organizational level, e.g. the regulation of a spatially extended pattern in a multicellular organism, such a local model of network evolution may not apply any more, and network structure and -organization as a whole may become important; hence, *morphogenesis by coupled regulatory networks* will require a different methodological approach. This problem will be studied in more detail in the following chapter.

Chapter 8

Morphogenesis by coupled regulatory networks

8.1 Introduction

Understanding the molecular machinery that regulates development of multicellular organisms is among the most fascinating problems of modern science. Today, a growing experimental record about the regulatory mechanisms involved in development is accumulating, in particular in well-studied model-organisms as, e.g., *Drosophila* or *Hydra* (Technau *et al.*, 2000; Bosch, 1998). Still, the genomic details known today are not sufficient to derive dynamical models of developmental gene regulation processes in full detail. Phenomenological models of developmental processes, on the other hand, are well established today. Pioneering work in this field was done by Turing, who in his seminal paper (Turing, 1952) considered a purely physico-chemical origin of biological pattern formation. His theory is based on an instability in a system of coupled reaction-diffusion equations. In this type of model, for certain parameter choices, stochastic fluctuations in the initial conditions can lead to self-organization and maintenance of spatial patterns, e.g. concentration gradients or periodic patterns. This principle has been successfully applied to biological morphogenesis in numerous applications (Gierer and Meinhardt, 1972; Meinhardt and Gierer, 2000). However, as current experiments make us wonder about the astonishingly high complexity of single regulating genes in development (Bosch and Khalturin, 2002), they also seem to suggest that diffusion models will not be able to capture all details of developmental regulation, and point at a complex network of regulating interactions instead.

The role of information processing in gene regulatory networks during development has entered the focus of theoretical research only recently. One pioneering study was

published by Jackson, Johnson and Nash (1986), who investigated the dynamics of spatial pattern formation in a system of locally coupled, identical dynamical networks. In this model, gene regulatory dynamics is approximated by Boolean networks with a subset of nodes communicating not only with nodes in the (intracellular) network, but also with some nodes in the neighboring cells. Boolean networks are minimal models of information processing in network structures and have been discussed as models of gene regulation since the end of the 1960s (Kauffman, 1969, 1993). The model of Jackson et al. demonstrated the enormous pattern forming potential of local information processing, similar to *direct contact induction* (Slack, 1993) as known in developmental biology. More recently, Salazar-Ciudad et al. introduced a gene network model based on continuous dynamics (Salazar-Ciudad, Garcia and Solé, 2000; Solé, Salazar-Ciudad and Garcia-Fernandez, 2002) and coupling their networks by direct contact induction. Interestingly, they observe a larger variety of spatial patterns than Turing-type models with diffusive morphogens, and find that patterns are less sensitive to initial conditions, with more time-independent (stationary) patterns. This matches well the intuition that networks of regulators have the potential for more general dynamical mechanisms than diffusion driven models. Let us here therefore study interacting networks in pattern formation and in particular consider information-transfer-based processes. In (Rohlf and Bornholdt, 2003), we introduced a simple cellular automata (CA) model of spatial pattern formation based on local information transfer, that performs de novo pattern formation by generating and regulating a domain boundary. In section 8.2.2, this model is defined and in section 8.3.1 the resulting (deterministic) dynamics is analyzed. A study of the dynamics under noise and cell flow is delayed to chapter 9. Here, we will focus on the question how this very general mechanism could emerge as a result of interacting nodes in coupled identical networks, similar to gene regulation networks in interacting cells. We will show that the model can be mapped on a system of coupled regulatory networks by means of two steps: first, the CA model is translated into a Boolean network model. The basic logic of this model is analyzed and then, in a second step, mapped on a threshold network. The resulting network has a hierarchical structure of information processing and size and complexity well compatible with experimental observations in biological organisms (von Dassow *et al.*, 2000; Bosch and Khalturin, 2002).

The genetic algorithm applied for identification of model solutions is explained in appendix A of this thesis, and in appendix B a comparison of the statistical properties of different solutions is provided.

8.2 Problem and Model

Let us first motivate and define the morphogenetic problem that we will use as a test case. Then we undertake a three-step approach to find a genetic network model that solves this pattern formation problem. In the first step, we summarize the properties of the cellular automata model introduced in Rohlf and Bornholdt (2003). Cellular automata as dynamical systems discrete in time and state space are known to display a wide variety of complex patterns (Wolfram, 1983, 1984a,b) and are capable of solving complex computational tasks, including universal computation. We searched for solutions (i.e. rule tables which solve the problem) by aid of a genetic algorithm (for details, see Appendix). Candidate solutions have to fulfill four demands: Their update dynamics has to generate a spatial pattern which 1) obeys a predefined scaling ratio $\alpha/(1-\alpha)$, 2) is independent of the initial condition chosen at random, 3) is independent of the system size (i.e. the number of cells N_C) and 4) is stationary (a fixed point). In the second step, this cellular automata rule table is “translated” into (spatially coupled) Boolean networks, using binary coding of the cellular automata states. The logical structure of the obtained network is reduced to a minimal form, and then, in step three, translated into a threshold network.

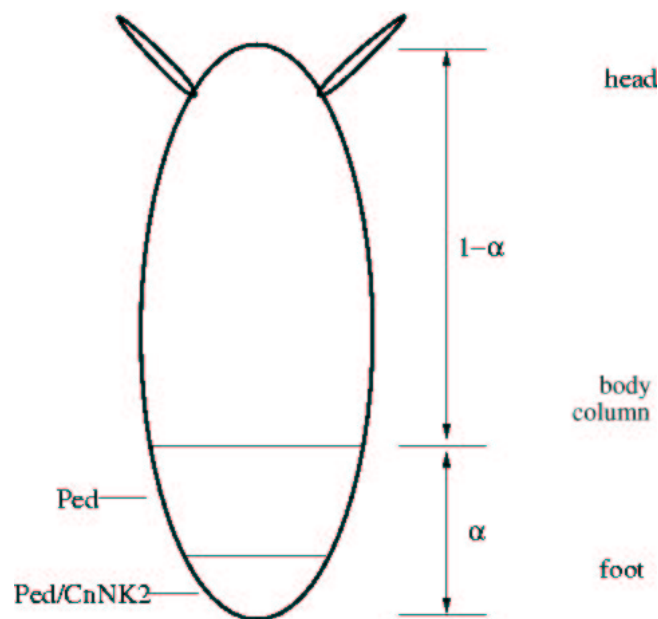


Figure 8.1: Gene expression domains in *Hydra*, here for the example of the “foot” genes Pedibin (Ped) and Cn-NK2. The Pedibin domain shows a quite sharp boundary towards the body region of the animal. The relative position of the boundary, given by the ratio $\alpha/(1-\alpha)$, is independent of the absolute size of the animal (proportion regulation).

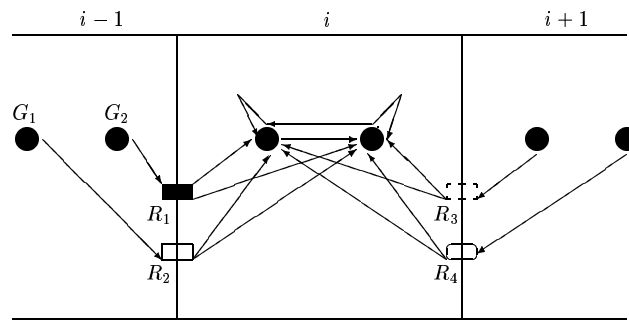


Figure 8.2: Diagram showing the interaction structure of the minimal network needed to solve the asymmetric expression task. For the sake of clarity, intracellular interactions between the two genes G_1 and G_2 are shown only for cell i , and likewise outgoing intercellular signals from the two genes two the neighbor cells $i - 1$ and $i + 1$ were left out. The transcription factors produced by gene G_1 and G_2 in cell $i - 1$ couple to the receptor systems R_1 and R_2 , respectively, whereas in cell $i + 1$ the transcription factors produced by these genes couple to the receptor systems R_3 and R_4 (biased signaling). In cell i , the receptors release factors which regulate the activity of G_1 and G_2 .

8.2.1 A test scenario: Hydra foot formation

A classical model organism for studies of position dependent gene activation is the fresh water polyp *Hydra*, which has three distinct body regions - a head with mouth and tentacles, a body column and a foot region. The positions of these regions are accurately regulated along the body axis.

The problem we shall focus on is sketched schematically in Fig. 10.6 for the example of the “foot” genes *Pedibin* and *CN-NK2* (Thomsen, 2001; Thomsen *et al.*, 2003). *CN-NK2* is active only in the lowest foot region, the expression domain of *Pedibin* extends a bit more upwards on the body column, where it is turned off at some point, i.e. the domain shows a sharp boundary. The relative position of the boundary is almost independent of the animal’s size, i.e. the ratio $\alpha/(1 - \alpha)$ (as denoted in Fig. 10.6) is almost invariant under changes of body size. As the *Pedibin/CN-NK2* system presumably plays an important role in determining the foot region, this invariance appears to be an essential prerequisite for maintaining the correct body proportions (proportion regulation) and to establish the head-foot polarity. Such regulation of position information is a quite general problem in biological development (Wolpert, 1969). An interesting problem is how the specific properties of this regulation can be achieved by a small network of regulatory genes and if so, whether local communication between the cells (networks) is sufficient. This basic question is the central motivation for the present study. In particular, we consider the simplified problem of regulating

one domain as, for example, the foot region versus the rest of the body. We consider this as a one dimensional problem as first approximation to the well-defined head-foot-axis in *Hydra*.

Developmental processes exhibit an astonishing robustness. This often includes the ability of de novo pattern formation, e.g., to regenerate a *Hydra* even after complete dissociation of the cell ensemble in a centrifuge (Müller, 1993). Further, they are robust in the face of a steady cell flux: *Hydra* cells constantly move from the central body region along the body axis towards the top and bottom, where they differentiate into the respective cell types according to their position on the head-foot axis. The global pattern of gene activity is maintained in this dynamic environment. Let us take these observations as a starting point for a detailed study how the interplay of noise-induced regulatory dynamics and cell flow may stabilize a developmental system.

8.2.2 One dimensional cellular automata: Definitions

To define a model system that performs the pattern formation task of domain self-organization (Rohlf and Bornholdt, 2003), consider a one-dimensional cellular automaton with parallel update (Wolfram, 1983). N_C cells are arranged on a one-dimensional lattice, and each cell is labeled uniquely with an index $i \in \{0, 1, \dots, N_C - 1\}$. Each cell can take n possible states $\sigma_i \in \{0, 1, \dots, n\}$. The state $\sigma_i(t)$ of cell i is a function of its own state $\sigma_i(t-1)$ and of its neighbor's states $\sigma_{i-1}(t-1)$ and $\sigma_{i+1}(t-1)$ at time $t-1$, i.e.

$$\sigma_i(t) = f[\sigma_{i-1}(t-1), \sigma_i(t-1), \sigma_{i+1}(t-1)] \quad (8.1)$$

with $f : \{0, 1, \dots, n\}^3 \mapsto \{0, 1, \dots, n\}$ (a cellular automaton with *neighborhood 3*). At the system boundaries, for simplicity we choose a discrete analogue of zero flux boundary conditions, i.e. we set $\sigma_{-1} = \sigma_{N_C+1} = \text{const.} = 0$. Other choices, e.g. asymmetric boundaries with cell update depending only on the inner neighbor cell, lead to similar results. The state evolution of course strongly depends on the choice of f : for a three-state cellular automaton ($n = 3$), there are $3^{27} \approx 7.626 \cdot 10^{12}$ possible update rules, each of which has a unique set of dynamical attractors. As we will show in the results section, $n = 3$ is the minimal number of states necessary to solve the pattern formation problem formulated above.

Now we can formulate the problem we intend to solve as follows: Find a set \mathcal{F} of functions (update rules) which, given a random initial vector $\vec{\sigma} = (\sigma_0, \dots, \sigma_{N_C-1})$, within T update steps evolves the system's dynamics to a fixed point attractor with the property:

$$\vec{\sigma}^* := \begin{cases} \sigma_i = 2 & \text{if } i < [\alpha \cdot N_C] \\ \sigma_i \neq 2 & \text{if } i \geq [\alpha \cdot N_C] \end{cases} \quad (8.2)$$

where $[\cdot]$ is the Gauss bracket. The scaling parameter α may take any value $0 < \alpha < 1$. For simplicity, it is fixed here to $\alpha = 0.3$. Notice that α does not depend on N_C , i.e. we are looking for a set of solutions where the ratio of the domain sizes $r := \alpha/(1 - \alpha)$ is *invariant under changes of the system size*. This clearly is a non-trivial task when only local information transfer is allowed. The ratio r is a *global property* of the system, which has to emerge from purely local (next neighbor) interactions between the cell's states.

8.2.3 Translation into spatially coupled Boolean networks

One can now take a step further towards biological systems, by transferring the dynamics we found for a cellular automata chain onto cells in a line that communicate with each other, similar to biological cells. Identifying states with cell types and assuming that the model cells have a network of regulators inside, each of them capable to reproduce the rules of a cellular automaton, we obtain a model mimicking basic properties of a biological genetic network in development.

Cellular automata rule tables can easily be translated into logical (Boolean) networks, e.g., for $n = 3$, two internal nodes can be used for binary coding of the cell states. One then has

$$(\sigma_1^i(t), \sigma_2^i(t)) = f_1(\sigma_{1,2}^{i-1}(t-1), \sigma_{1,2}^i(t-1), \sigma_{1,2}^{i+1}(t-1)) \quad (8.3)$$

with $f_1 : \{0, 1\}^6 \mapsto \{0, 1\}^2$. The so obtained rule tables are, by application of Boolean logic, transformed into a minimized *conjunctive normal form*, which only makes use of the the three logical operators NOT, AND and OR with a minimal number of AND operations. This is a rather realistic assumption for gene regulatory networks, as the AND operation is more difficult to realize on the basis of interactions between transcription factors. Other logical functions as, e.g., XOR, are even harder to realize biochemically Davidson (2001). The structure of the constructed network and its biological interpretation is shown in Fig. 8.2. Note that the up-down symmetry of the body axis is broken locally by a spatially asymmetric receptor distribution (Marciniak, 2003).

8.2.4 Translation into spatially coupled threshold networks

Perhaps the simplest model for transcriptional regulation networks are threshold networks, a subset of Boolean networks, where logical functions are modeled by weighted sums of the nodes' input states plus a threshold h (Kürten, 1988a; ?). They have proven to be valuable tools to address questions associated to the dynamics and evolution of

gene regulatory networks (Wagner, 1994; Bornholdt and Sneppen, 2000; Bornholdt and Rohlf, 2000; Rohlf and Bornholdt, 2002, 2004a).

Any Boolean network which has been reduced to its minimized form (here, the conjunctive normal form), can be coded as a dynamical threshold network, with minimally three hierarchies of information processing (“input layer”: signals from the neighbor cells at time $t - 1$, “hidden layer”: logical processing of the signals, “output layer”: states of the two “pattern genes” in cell i at time t). The genes’ states now may take values $\sigma_i = \pm 1$, and likewise for the interaction weights one has $c_{ij}^l = \pm 1$ for activating and inhibiting regulation, respectively, and $c_{ij}^l = 0$ if gene i does not receive an input from gene j in cell l . The dynamics then is defined as

$$\sigma_j^i(t) = \text{sign}(f_j(t-1)) \quad (8.4)$$

with

$$f_j(t) = \sum_{k=1}^2 \sum_{l=i-1}^{i+1} c_{kj}^l \sigma_k^l + h_j \quad (8.5)$$

for the “hidden” genes, where $\sigma_k^l, k \in \{1, 2\}$ is the state of the k th *output* gene in cell l (there are no couplings between the genes in the “hidden” layer). The threshold h_j is given by

$$h_j = \sum_{k=1}^2 \sum_{l=i-1}^{i+1} |c_{kj}^l| - 2, \quad (8.6)$$

which implements a logical OR operation. For the “output” (pattern forming) genes, one simply has

$$f_k(t) = \sum_{l=1}^{k_{in}^k} \sigma_l - k_{in}^k, \quad (8.7)$$

i.e. the weights are all set to one, and the (negative) threshold equals the number of inputs k_{in}^k that gene k receives from the “hidden” genes (logical AND).

8.3 Results

8.3.1 Cellular Automata Model

The first major outcome of the cellular automata model is that a number of $n = 3$ different states is necessary and sufficient for this class of systems to solve the given pattern formation task. The update table of the fittest solution found during optimization runs, which solves the problem independent of system size for about 98 percent of (randomly chosen) initial conditions (i.e. has fitness $\Phi = 0.98$), is shown in Table I.

index	σ_{i-1}	σ_i	σ_{i+1}	σ_i	index	σ_{i-1}	σ_i	σ_{i+1}	σ_i
0	0	0	0	0	14	1	1	2	2
1	0	0	1	2	15	1	2	0	0
2	0	0	2	1	16	1	2	1	0
3	0	1	0	0	17	1	2	2	1
4	0	1	1	2	18	2	0	0	0
5	0	1	2	2	19	2	0	1	0
6	0	2	0	1	20	2	0	2	0
7	0	2	1	2	21	2	1	0	1
8	0	2	2	2	22	2	1	1	2
9	1	0	0	0	23	2	1	2	2
10	1	0	1	1	24	2	2	0	1
11	1	0	2	1	25	2	2	1	2
12	1	1	0	0	26	2	2	2	2
13	1	1	1	1					

Table 8.1: Rule table of the best cellular automata solution found during genetic algorithm search. In the left column, the rule table index is shown, running from 0 to 26, in the middle column the three input states at time t are shown, the right column shows the corresponding output states at time $t + 1$.

Fig. 8.3 shows the typical update dynamics of this solution. The finite size scaling of the self-organized relative domain size α as a function of the the number of cells N_C is shown in Fig. 8.4. In the limit of large system sizes, α converges towards

$$\alpha_\infty = 0.281 \pm 0.001. \quad (8.8)$$

The variance of α vanishes with a power of N_C , i.e. the relative size of fluctuations induced by different initial conditions becomes arbitrarily small with increasing system size. Hence, the pattern self-organization in this system exhibits considerable robustness against fluctuations in the initial conditions. The main mechanism leading to stabilization at $\alpha_\infty = 0.281$ is a modulation of the traveling velocity of the right phase boundary in Fig. 2 such that the boundary on average moves slightly more than one cell to the left per update step, whereas the left boundary moves one cell to the right exactly every third update step. the modulation of the right boundary can be seen as the result of interacting phase boundaries reminiscent of particle interactions, that were introduced and defined in section 4.3.

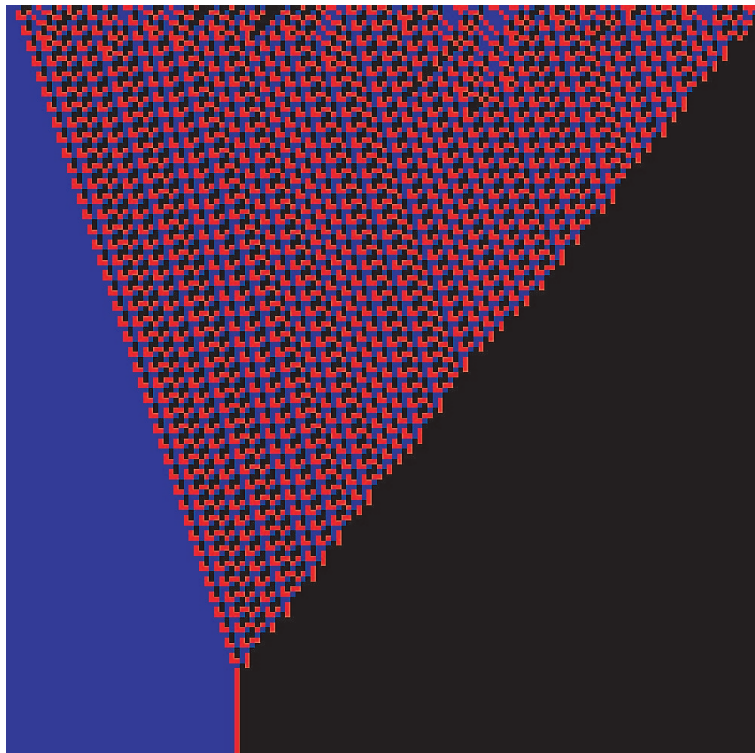


Figure 8.3: A typical dynamical run for the Automata as defined in Table 8.1, here for a system of size $N_C = 250$ cells (deterministic dynamics, no noise), starting from a random initial configuration. Time is running on the y -axis from top to bottom. Cells with state $\sigma_i = 0$ are depicted in black color, cells with $\sigma_i = 1$ in red and cells with $\sigma_i = 2$ in blue.

8.3.2 Interaction topology of the minimal network

In this section let us translate the rule table found in the previous section into a (gene) regulatory network. We will see that this network has biologically realistic properties regarding the number of genes necessary for information processing and the complexity of interaction structure, making it well conceivable that similar “developmental modules” exist in biological systems.

8.3.2.1 Boolean representation

The rule table of Table 1 first is translated into binary coding, i.e $0 \rightarrow 00$, $1 \rightarrow 01$ and $2 \rightarrow 10$, this corresponds to two “genes” G_1 and G_2 one of which (G_1) is active only in a domain at the left side of the cell chain. The so obtained Boolean update table is reduced to its minimized conjunctive normal form, using a Quine-McCluskey algorithm (McCluskey, 1956). For the construction of the network topology we use

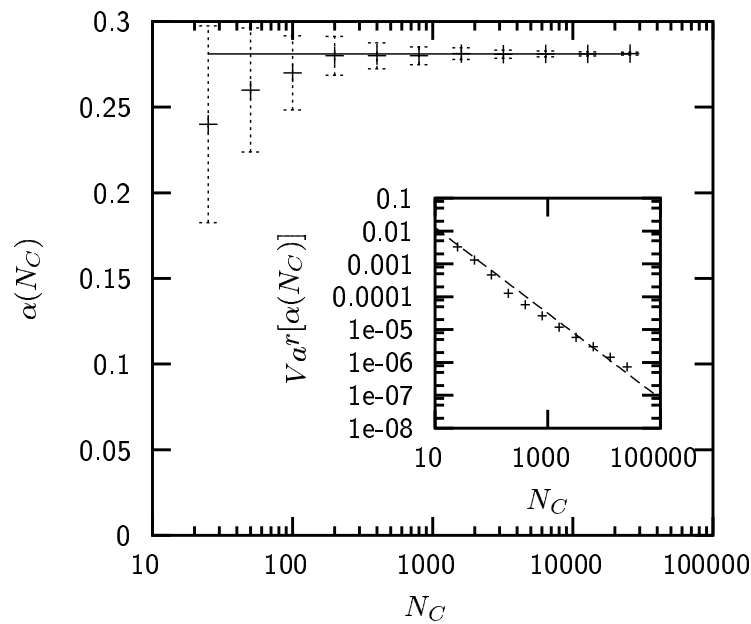


Figure 8.4: Finite size scaling of the self-organized relative domain size α as a function of the total number of cells N_C . In the limit of large system sizes, α converges towards a fixed value $\alpha_\infty = 0.281 \pm 0.001$ (as denoted by the straight line fit). The inset shows the finite size scaling of the variance $Var[\alpha(N_C)]$; the straight line in this log-log plot has slope -1.3 and indicates that fluctuations vanish with a power of the system size.

the conjunctive normal form, as it is a somewhat biologically plausible solution with a minimal number of logical AND operations. In principle, other network topologies, e.g. with more levels of hierarchy, are possible and biologically plausible, however, they involve a higher number of logical sub-processing steps, i.e. a higher number of genes, hence we will not discuss them here.

Considering the huge number of possible input configurations which the outputs theoretically could depend on, the complexity of the resulting network is surprisingly low. As shown in Fig. 8.5, the output state of gene G_1 only depends on five different input configurations of at maximum four different inputs, gene number two on six different input configurations of at maximum four different inputs. This indicates that the spatial information flowing into that network is strongly reduced by internal information processing (only a small number of input states leads to output “1”), as expected for the simple stationary target pattern. Nevertheless, this information processing is sufficient to solve the non-trivial task of domain size scaling.

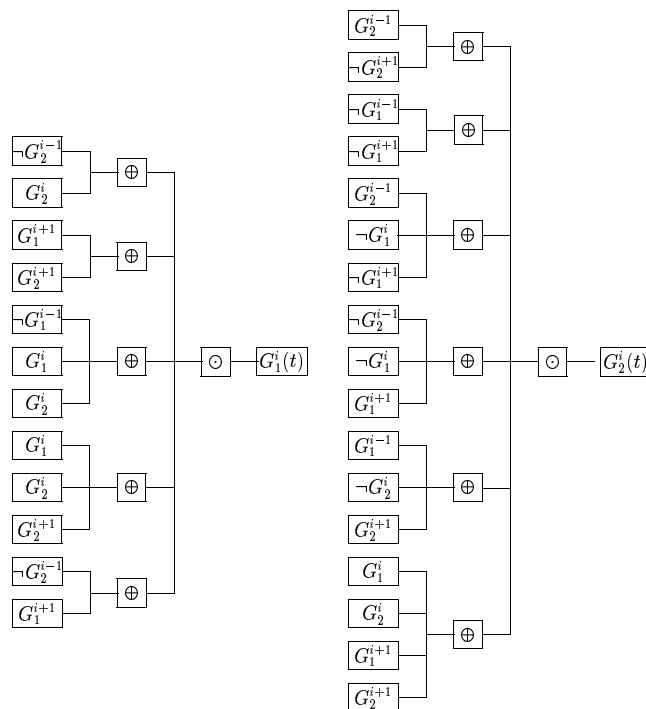


Figure 8.5: Boolean representation of the minimal network, minimized conjunctive normal form. G_a^b with $a \in \{1, 2\}$ and $b \in \{i-1, i, i+1\}$ denotes gene a in cell number b . The inputs in the left branches of the trees are given by the genes' states at time $t-1$. \neg denotes NOT, \odot denotes logical AND and \oplus logical OR.

8.3.2.2 Threshold network implementation

Alternatively, the Boolean representation of the pattern formation system can be translated into a three-layered threshold network (Fig. 8.6). The states of the genes G_1 and G_2 at time t in a cell i and its two neighbor cells $i-1$ and $i+1$ serve as inputs of 11 information-processing genes (“hidden” layer). The state of these genes then defines the state of G_1 and G_2 in cell i at time $t+2$ (output layer). Additionally, there is some feedback from G_1 and G_2 to the information processing layer, as expected for the dependence on cell-internal dynamics already present in the cellular automata implementation of the model. The resulting stationary spatial patterns of gene G_1 (the “domain gene”), G_2 (active only at the domain boundary) and the “hidden” genes after system relaxation are shown in Table 2, for a system size of $N_C = 30$ cells. From this seemingly redundant pattern in equilibrium one hardly guesses the higher level of genetic information processing during the self-organizing phase. The network we construct here, regarded as a “developmental module” defining the head-foot polarity through spatially asymmetric gene expression, has a size similar to biological modules

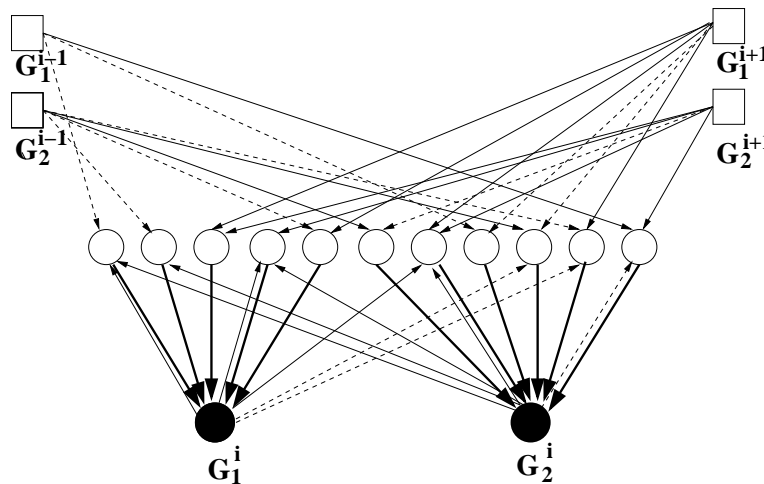


Figure 8.6: Threshold network realization of the pattern formation system. Solid line arrows denote links with $w_{ij} = +1$, dashed arrows denote links with $w_{ij} = -1$. The inputs from the genes in the neighbor cells ($G_1^{i-1}, G_2^{i-1}, G_1^{i+1}$ and G_2^{i+1}) are processed by a layer of “hidden genes” (hollow circles) with different thresholds h (as these “genes” perform a logical OR operation, one has $h = k_{in} - 2$). The threshold of gene G_1 is $h_1 = -5$, for gene G_2 one has $h_2 = -6$ (logical AND).

(compare, for example, the segment polarity network in *Drosophila*, von Dassow *et al.*, 2000)) as well as similar complexity (average connectivity $\bar{K} \approx 3$).

8.4 Discussion

In this chapter, the dynamics of pattern formation motivated by animal morphogenesis and the largely observed participation of complex gene regulation networks in their coordination and control was considered. A simple developmental problem was chosen to study toy models of interacting networks that control pattern formation and morphogenesis in this setting. In particular, the goal was to explore how networks can offer additional mechanisms beyond the standard diffusion based process of the Turing instability. The obtained results suggest that main functions of morphogenesis can be performed by dynamical networks without relying on diffusive biochemical signals, but using local signaling between neighboring cells. This includes solving the problem of generating global position information from purely local interactions. But also it goes beyond diffusion based models as it offers solutions to developmental problems that are difficult for such models and avoids their inherent problem of fine tuned model parameters. It may be not too far fetched to speculate that some developmental processes

gene number	spatial pattern
1	11111111111111111111111111111111
2	111111111110111111111111111111
3	111111111000000000000000000000
4	111111111100000000000000000000
5	111111111110111111111111111111
6	111111110111111111111111111111
7	111111111100000000000000000000
8	100000000111111111111111111111
9	000000000111111111111111111111
10	111111111111111111111111111111
11	111111111111111111111111111111
G_1	111111111000000000000000000000
G_2	000000000100000000000000000000

Table 8.2: Spatial gene activity patterns of the network shown in Fig. 6. (stationary patterns after the system has reached equilibrium), $N_C = 30$. The numbers identify the “hidden” genes in Fig. 8.6 from the left hand side to the right. States were transformed by $\sigma \rightarrow (\sigma + 1)/2$.

as, e.g., the establishment of position information, may rely on this type of internal information processing rather than on interpretation of, e.g., chemical gradients. As only ingredient to the present model, symmetry has to be broken locally by a spatially asymmetric receptor distribution, e.g., as proposed by Marciniak (2003), to provide the necessary input information. Also hybrid approaches are conceivable. For example such broken symmetry could be provided by a gradient, even without fine-tuning (contrary to standard gradient-based models) since only the direction of the gradient has to be read out.

The network model derived here performs accurate regulation of position information and robust *de novo* pattern formation from random conditions, with a mechanism based on local information transfer rather than the Turing instability. Non-local information is transmitted through soliton-like quasi-particles instead of long-range gradients. Two realizations as discrete dynamical networks, Boolean networks and threshold networks, have been developed. The resulting networks have size and complexity comparable to developmental gene regulation modules as observed in animals, e.g., *Drosophila* (von Dassow *et al.*, 2000) or *Hydra* (Bosch, 1998; Bosch and Khalturin, 2002). The threshold networks (as models for transcriptional regulation networks) process position information in a hierarchical manner; in the present study, hierarchy levels were limited to three, but realizations with more levels of hierarchy, i.e. more

“pre-processing” of information are also possible. Similar hierarchical and modular organization are typical signatures of gene regulatory networks in organisms (Davidson, 2001).

Chapter 9

Noise-induced control, phase transitions, and cell flow in pattern formation

9.1 Introduction

Morphogenesis and pattern formation work with great precision and reliability in the presence of various sources of noise, acting on a large range of scales. Gene regulation itself is intrinsically a stochastic process, as transcription factors bind stochastically to the respective binding sites (Lewin, 1997; Alberts *et al.*, 2002). The question how coordinated (for example, periodic) gene expression is achieved in spite of stochasticity in switching events (Gardner and Collins, 2000) and although there is no such thing as a central “update clock”, is fascinating in itself and is a current research topic in Statistical Physics (Kepler and Elston, 2001; Sasai and Wolynes, 2003; Klemm and Bornholdt, 2003). However, in development fluctuations and perturbations can occur on even more coarse-grained scales. There is evidence, for example, that proliferation and differentiation of stem cells in some cases is a stochastic process (Abkowitz, Catlin and Gutter, 1996), a feature which is also predicted by some theoretical approaches (Kaneko and Furusawa, 1999; Furusawa and Kaneko, 2001). Cell movements and errors in cell differentiation lead to additional sources of “developmental noise”. However, overall development proceeds in spite of all these perturbations, given they do not exceed a certain tolerance level – morphogenesis is an *intrinsically robust* process (Edelman, 1988; Goodwin, Kauffman and Murray, 1993).

In the following we will study dynamics and robustness of the model introduced in chapter 8 with respect to noise. Two kinds of perturbations frequently occur: Stochas-

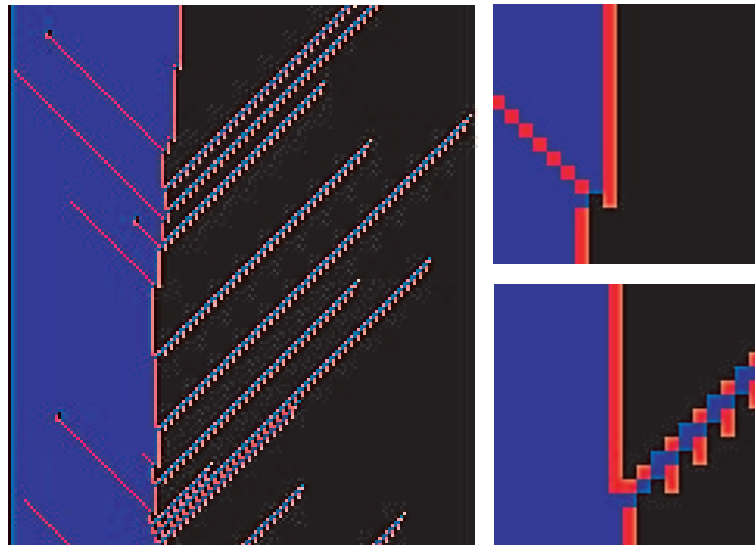


Figure 9.1: Quasi-particles, started by stochastic update errors, lead to control of the boundary position under noise (left panel). The Γ particle (top right) leads to readjustment of the boundary two cells to the left, the Δ particle (bottom right) leads to readjustment of the boundary one cell to the right.

tic update errors and external forces caused by a directed *cell flow* due to cell proliferations. Both types of perturbations are very common during animal development, e.g., in *Hydra* cells continuously move from the central body region along the body axis towards the top and bottom, and differentiate into the respective cell types along the way according to their position on the head-foot axis.

We show that the model is robust with respect to stochastic update errors. Noise-induced excitations (quasi-particles) contribute to pattern stabilization by stochastic control of the boundary position that separates the two activity domains in the stationary state, even in presence of cell flow. Two phase transitions are found, a first order phase transition at vanishing noise, and a second order phase transition at increased cell flow.

9.2 Dynamics under noise

Let us define stochastic update errors with probability p per cell, leading to an average error rate $r_e = p N_C$. Interestingly, this stochastic noise starts moving “particle” excitations in the cellular automaton which, as a result indeed stabilize the developmental structure of the system. To prepare for the details of these effects, define first how we measure the boundary position properly in the presence of noise. Let us use a

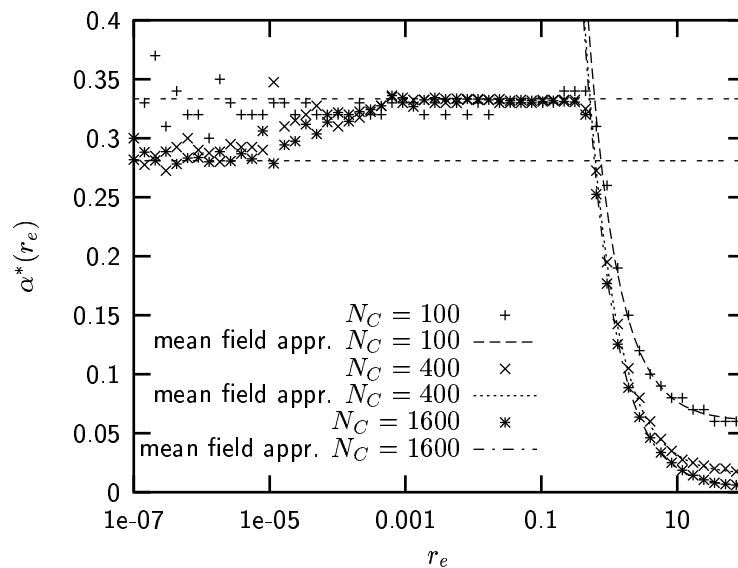


Figure 9.2: Average boundary position α^* as a function of the error rate r_e for system sizes $N_C = 100$, $N_C = 400$, and $N_C = 1600$. The abscissa is logarithmic. Numerical data are averaged over 200 different initial conditions with $2 \cdot 10^6$ updates each. The dashed curves show the mean field approximation given by Eqn. 9.1, the straight dashed lines mark the unperturbed solution $\alpha^* = 0.281$ and the solution under noise, $\alpha^* = 1/3$.

statistical method to measure the boundary position in order to get conclusive results also for high p : Starting at $i = 0$, we put a “measuring frame” of size w over cell i and the next $w - 1$ cells, move this frame to the right and, for each i , measure the fraction z of cells with state $\sigma = 2$ within the frame. The algorithm stops when z drops below $1/2$ and the boundary position is defined to be $i + w/2$. It is easy to see that, for not too high p , there are only two different quasi-particles (i.e. state perturbations moving through the homogeneous phases), as shown in Fig. 9.1. In the following, these particles are called Γ and Δ . The Γ particle is started in the σ_2 phase by a stochastic error $\sigma_i = 2 \rightarrow \sigma_i \neq 2$ at some $i < \alpha N_C$, moves to the right and, when reaching the domain boundary, readjusts it two cells to the left of its original position. The Δ particle is started in the σ_0 phase by a stochastic error $\sigma_i = 0 \rightarrow \sigma_i \neq 0$ at some $i > \alpha N_C$ and moves to the left. Interaction with the domain boundary readjusts it one cell to the right. Thus we find that the average position α^* of the boundary is given by the rate equation

$$2\alpha^* r_e = (1 - \alpha^*) r_e, \quad (9.1)$$

i.e. $\alpha^* = 1/3$. Interestingly, for not too high error rates r_e , α^* is independent from r_e and thus from p . If we consider the average boundary position α^* as a system-specific

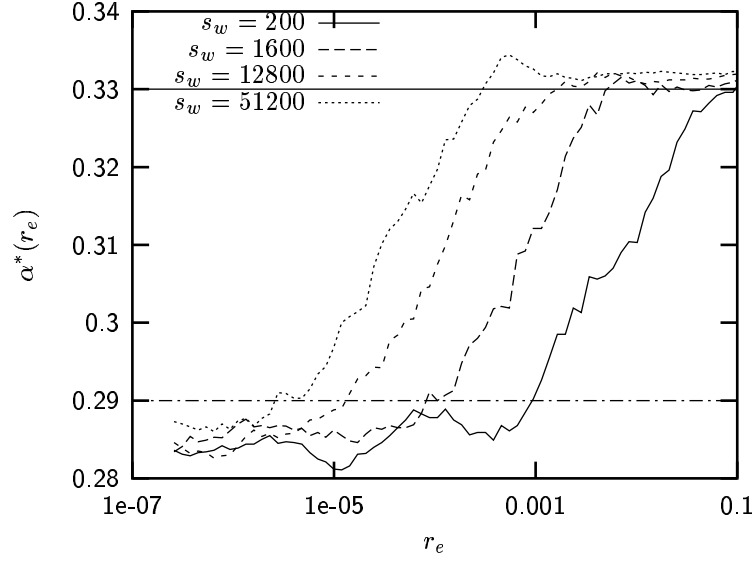


Figure 9.3: Average domain boundary position α^* as a function of the error rate r_e , sampled over update windows of different lengths s_w (ensemble statistics, 400 different initial conditions for each data point). The abscissa is logarithmic. With increasing s_w , the transition from the solution $\alpha_{det}^* = 0.281$ under deterministic dynamics to $\alpha^* = 1/3$ under noise is shifted towards $r_e = 0$. The two straight lines define a lower boundary α_{low}^* and an upper boundary α_{up}^* , as explained in the text.

order parameter which is controlled by the two quasi-particles, then comparing the solution of Eqn. (9.1) to the equilibrium position in the noiseless case indicates that the system undergoes a first order phase transition with respect to α^* at $p = r_e = 0$.

Figs. 9.3 and 9.4 show noise dependence and finite size scaling of the transition. In case of a first order phase transition at $r_e = 0$, we would expect a shift of the transition point $r_e^{trans}(s_w)$ towards $r_e = 0$ which is proportional to s_w^{-1} as well as a divergence of the slope at the transition point when s_w is increased, i.e. $d\alpha^*/dr_e(r_e^{trans}) \rightarrow \infty$ when $s_w \rightarrow \infty$.

The shift of $r_e^{trans}(s_w)$ is most easily measured by defining a lower and an upper boundary α_{low}^* and α_{up}^* , respectively (Fig. 9.3); when α^* crosses these boundaries, two transition points r_e^{up} and r_e^{low} are obtained. We find that $r_e^{up} \approx c_{up}s_w^{-1}$ and $r_e^{low} \approx c_{low}s_w^{-1}$ with $c_{up} > c_{low}$ as expected (Fig. 9.4), which implies that the difference $\Delta r_e^{trans}(s_w) := r_e^{up} - r_e^{low}$ scales as

$$\Delta r_e^{trans}(s_w) = (c_{up} - c_{low}) s_w^{-1}, \quad (9.2)$$

hence, because $\Delta\alpha^*(r_e^{trans}) = const. = \alpha_{up}^* - \alpha_{low}^*$, indeed $d\alpha^*/dr_e(r_e^{trans})$ diverges when the sampling window size goes to infinity.

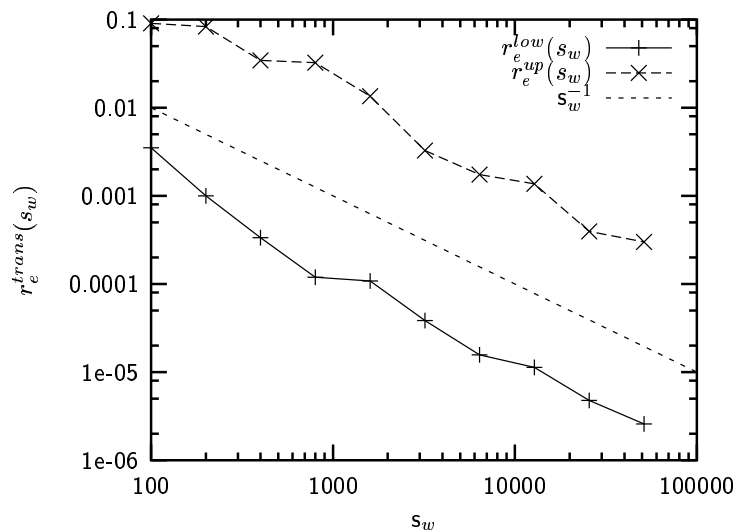


Figure 9.4: Finite size scaling of the upper and lower transition points r_e^{up} and r_e^{low} , i.e. the points where α^* crosses α_{low}^* and α_{up}^* , respectively (Fig. 7), as a function of the sampling window length s_w . Both r_e^{up} and r_e^{low} vanish $\propto s_w^{-1}$, as indicated by the line with slope -1 in this log-log-plot.

The solution $\alpha^* = 1/3$ is stable only for $0 < r_e \leq 1/2$. As shown in Fig. 9.1, the interaction of a Γ particle with the boundary needs only one update time step, whereas the boundary readjustment following a Δ particle interaction takes three update time steps. Hence, we conclude that the term on the right hand side of Eqn. (10), which gives the flow rate of Δ particles at the boundary, for large r_e will saturate at $1/3$, leading to

$$2\alpha^* r_e = \frac{1}{3} \quad (9.3)$$

with the solution

$$\alpha^* = \frac{1}{6} r_e^{-1} + \Theta(N_C) \quad (9.4)$$

for $r_e > 1/2$. Hence, there is a crossover from the solution $\alpha^* = 1/3$ to another solution vanishing with r_e^{-1} around $r_e = 1/2$. The finite size scaling term $\Theta(N_C)$ can be estimated from the following consideration: for $p \rightarrow 1$, the average domain size created by “pure chance” is given by $\alpha^* = N_C^{-1} \sum_{n=0}^{N_C} (1/3)^n \cdot n \approx (3/4) N_C^{-1}$. If the measuring window has size w , we obtain $\Theta(N_C) \approx (3/4) w N_C^{-1}$. To summarize, we find that the self-organized boundary position is given by

$$\alpha^* = \begin{cases} 0.281 \pm 0.001 & \text{if } r_e = 0 \\ 1/3 & \text{if } 0 < r_e \leq 1/2 \\ (1/6)r_e^{-1} + \Theta(N_C) & \text{if } r_e > 1/2 \end{cases} \quad (9.5)$$

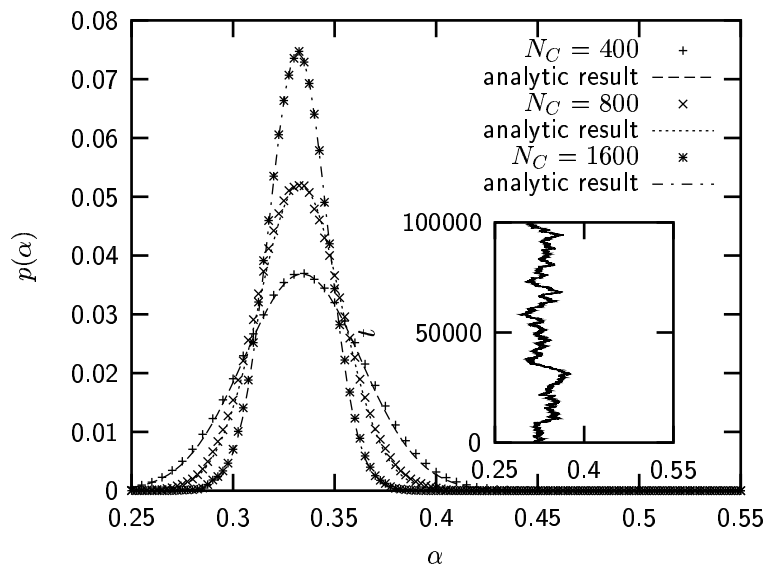


Figure 9.5: For the system with stochastic update errors, fluctuations of the boundary position α around the average position $\alpha^* = 1/3$ are Gaussian distributed. The figure compares the numerically obtained stationary probability distribution with the analytic result of Eqn. 9.10 for three different system sizes. All data are gained for $r_e = 0.1$ and averaged over 100 different initial conditions with $2 \cdot 10^6$ updates each. The inset shows a typical timeseries of the boundary position.

with a first-order phase transition at $r_e = 0$ and a crossover around $r_e = 1/2$.

Now let us consider the fluctuations of α around α^* given by the master equation

$$\begin{aligned}
 p^\tau(\alpha) &= 2\alpha r_e p^{\tau-1}(\alpha + 2\delta) + (1 - \alpha) r_e p^{\tau-1}(\alpha - \delta) \\
 &+ (N_C - r_e) p^{\tau-1}(\alpha) - 2\alpha r_e p^{\tau-1}(\alpha) \\
 &- (1 - \alpha) r_e p^{\tau-1}(\alpha)
 \end{aligned} \tag{9.6}$$

with $\delta = 1/N_C$. Eqn. (9.6) determines the probability $p^\tau(\alpha)$ to find the boundary at position α at update time step τ , given its position at time $\tau - 1$. This equation can be simplified as we are interested only in the *stationary probability distribution* of α . It is easy to see that the error rate r_e just provides a time scale for relaxation towards the stationary distribution and has no effect on the stationary distribution itself. Therefore, we may consider the limit $r_e \rightarrow r_e^{max} := N_C$, divide through r_e and neglect the last three terms on the right handside of Eqn. (9.6) (which become zero in this limit). We obtain

$$p^\tau(\alpha) = 2\alpha p^{\tau-1}(\alpha + 2\delta) + (1 - \alpha) p^{\tau-1}(\alpha - \delta). \tag{9.7}$$

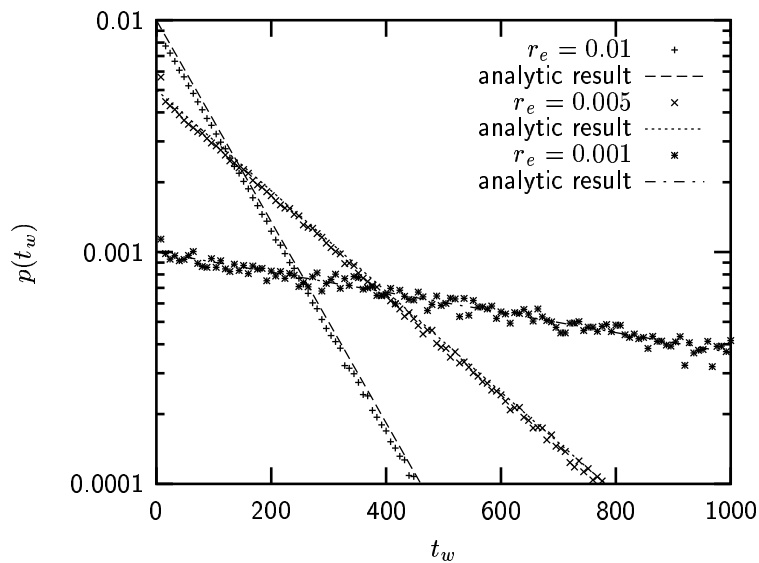


Figure 9.6: Probability distribution $p(t)$ of waiting times for boundary readjustments in the model with stochastic update errors for three different values of r_e , semi-log plot. As expected for a Poisson process, $p(t)$ is an exponential.

To study this equation, we consider the continuum limit $N_C \rightarrow \infty$. Let us introduce the scaling variables $x = (\alpha - \alpha^*)\sqrt{N_C}$, $t = \tau/N_C$ and the probability density $f(x, t) = N_C p^\tau(\alpha N_C)$. Inserting these definitions into Eqn. (9.7) and ignoring all subdominant powers $\mathcal{O}(1/N_C)$, we obtain a Fokker-Planck equation (Gardiner, 1983):

$$\frac{\partial f(x, t)}{\partial t} = \left(\frac{\partial^2}{\partial x^2} + 3 \frac{\partial}{\partial x} x \right) f(x, t). \quad (9.8)$$

The stationary solution of this equation is given by

$$f(x) = \sqrt{\frac{3}{2\pi}} \exp \left[-\frac{3}{2} x^2 \right], \quad (9.9)$$

i.e. in the long time limit $t \rightarrow \infty$, the probability density for the boundary position α is a Gaussian with mean α^* :

$$p(\alpha, N_C) = \sqrt{\frac{3 N_C}{2\pi}} \exp \left[-\frac{3 N_C}{2} (\alpha - \alpha^*)^2 \right]. \quad (9.10)$$

From Eqn. (9.10) we see that the variance of α vanishes $\sim 1/N_C$ and the relative boundary position becomes sharp in the limit of large system sizes. Fig. 9.5 shows that this continuum approximation for $N_C \geq 400$ provides very good correspondence with the numerically obtained probability distributions.

The stochastic nature of boundary stabilization under noise is also reflected by the probability distribution of *waiting times* t for boundary readjustments due to particle interactions: the particle production is a Poisson process with the parameter $\lambda = r_e$ and the waiting time distribution is given by

$$p_{wait}(t) = r_e \exp(-r_e t) \quad (9.11)$$

with an average waiting time $\langle t \rangle = r_e^{-2}$. Fig. 9.6 shows the waiting time distributions for different error rates r_e .

9.3 Noise-induced pattern control under cell flow

In a biological organism, a pattern has to be robust not only with respect to dynamical noise, but also with respect, e.g., to “mechanical” perturbations. In *Hydra*, e.g., there is a steady flow of cells directed towards the animal’s head and foot, due to continued proliferation of stem cells (David and Campbell, 1972); the stationary pattern of gene activity is maintained in spite of this cell flow. Let us now study the robustness of the model with respect to this type of perturbation. Let us consider a constant cell flow with rate r_f , which is directed towards the left or the right system boundary. In Eqn. (9.1), we now get an additional drift term r_f on the left hand side:

$$2\alpha^* r_e \pm r_f = r_e(1 - \alpha^*), \quad (9.12)$$

with the solution

$$\alpha^* = \begin{cases} \frac{1}{3} \left(1 - \frac{r_f}{r_e}\right) & \text{if } r_e \geq r_f \\ 0 & \text{if } r_e < r_f \end{cases} \quad (9.13)$$

for the case of cell flow directed towards the left system boundary (plus sign in eqn. (9.12)). One observes that α^* undergoes a second order phase transition at the critical value $r_e^{crit} = r_f$. Below r_e^{crit} , the domain size α^* vanishes, and above r_e^{crit} it grows until it reaches the value $\alpha_{max}^* = 1/3$ of the system without cell flow. For cell flow directed towards the right system boundary (minus sign in eqn. (9.12)), we obtain

$$\alpha^* = \begin{cases} \frac{1}{3} \left(1 + \frac{r_f}{r_e}\right) & \text{if } r_f \leq 2r_e \\ 1 & \text{if } r_f > 2r_e. \end{cases} \quad (9.14)$$

In this case, the critical cell flow rate is given by $r_f = 2r_e$, for cell flow rates larger than this value the σ_2 -domain extends over the whole system, i.e. $\alpha^* = 1$.

Fig. 9.7 compares the results of numerical simulations with the mean field approximation of Eqn. (9.12). In numerical simulations, cell flow is realized by application of

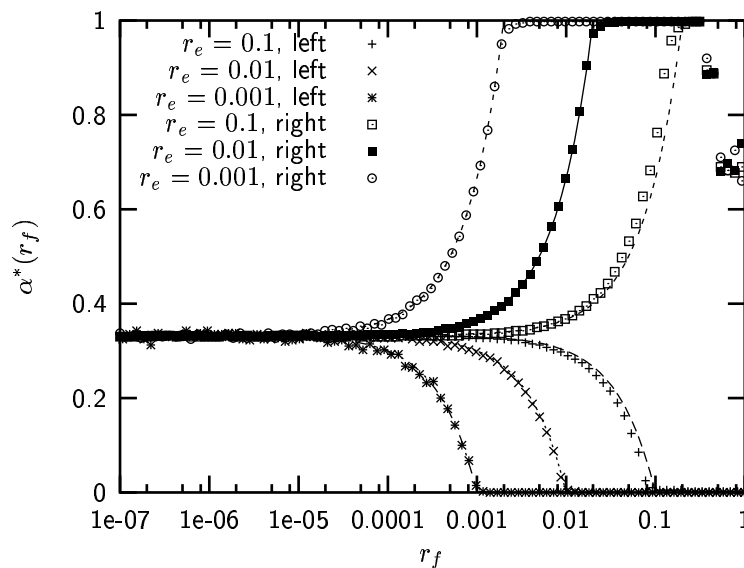


Figure 9.7: Average domain size α^* as a function of the the cell flow rate r_f for three different error rates r_e ; numerical data (crosses and points) were sampled over 10 different initial conditions and $1e6$ updates for each data point. “Left” indicates cell flow directed to the left system boundary, “right” to the right system boundary, respectively. The dashed lines are the corresponding solutions of Eqn. (9.12).

the translation operator $\Theta \sigma_i := \sigma_{i+1}$ to all cells with $0 \leq i < N_C - 1$ every r_f^{-1} time steps and leaving σ_{N_C-1} unchanged. In case of cell flow directed to the right system boundary, in the limit $r_f \rightarrow 1$ the boundary position α^* detected in numerical simulations deviates from the mean field prediction, due to a boundary effect at the left system boundary (stochastic production of finite lifetime stationary oscillators, leading to intermittent flows of Γ particles through the system).

To summarize this part, we see that in the model stochastic errors in dynamical updates for $r_e > r_f$ indeed *stabilize* the global pattern against the mechanical stress of directed cell flow.

9.4 Summary

Robustness of the model of pattern formation introduced in chapter 8 was studied in detail for two types of perturbations, stochastic update errors (noise) and directed cell flow. A first order phase transition is observed for vanishing noise and a second order phase transition at increasing cell flow. Fluctuations of the noise-controlled boundary position were studied numerically for finite size systems and analytically in

the continuum limit. We find that the relative size of fluctuations vanishes with $1/N_C$, which means that the boundary position becomes sharp in the limit of large system sizes. Dynamics under cell flow is studied in detail numerically and analytically by a mean field approximation. A basic observation is that noise-induced perturbations act as quasi-particles that stabilize the pattern against the directed force of cell flow. At a critical cell flow rate, there is a second order phase transition towards a vanishing domain size or a domain extending over the whole system, depending on the direction of cell flow, respectively.

Several extensions of this model are conceivable. In the present model, the cell flow rate r_f is considered as a free parameter, the global pattern, however, can be controlled easily by an appropriate choice of the error rate r_e . This may suggest to extend the model by introduction of some kind of dynamical coupling between r_e and r_f , treating r_f as a function of r_e . Interestingly, similar approaches have been studied by Hogeweg (2000) and Furusawa and Kaneko (2000, 2003): In both models of morphogenesis, the rate of cell divisions is controlled by cell differentiation and cell-to-cell signaling. Dynamics in both models, however, is deterministic. An extension of our model as outlined above may open up for interesting studies how *stochastic* signaling events could control and stabilize a global expression pattern and cell flow as an integrated system. Other possible extensions of the model concern the dimensionality: In two or three dimensions other mechanisms of symmetry breaking might be present, possibly leading to new, interesting dynamical effects.

Chapter 10

A two-dimensional model of pattern formation

10.1 Introduction

How shall one describe morphogenesis, concerning the dimensionality of its underlying space? Organisms are complex three-dimensional structures, and development needs time. Indeed, even after the embryonic phase development does not stop, hence it has been argued that multicellular organisms rather should be envisioned as four-dimensional structures similar to an “inverted cone” (Arthur, 1997, the cone reflects the fact that - in most cases - development starts from a single cell). However, in many cases, the dimensionality of a given pattern formation problem may be reduced without losing the essential information. In chapters 8 and 9, we approximated self-organization of *position information* as a one-dimensional system, inspired by the well-defined head-foot axis observed for the polyp *Hydra*. In this chapter, we shall proceed in the other direction and extend this model to two (spatial) dimensions. However, if pattern formation is based on local signals between direct neighbor cells, in two dimensions new problems arise. The boundaries of gene expression domains in *Hydra*, e.g., are very sharp in many cases – examples are the genes *Budhead* (Martinez *et al.*, 1997) and *Cn-gsc* (Broun *et al.*, 1999) in the *Hydra* head region – and do not show much position variation. The pattern formation mechanism proposed in chapter 8 alone, however, would lead to self-organization of quite “rugged” boundaries in 2D, with spatial fluctuations distributed according to a Gaussian.

We overcome this problem by introduction of a second pattern generating mechanism - neighborhood-dependent state changes (“redifferentiation” of cells), that competes with symmetry breaking during initial pattern formation. We show that, for small

probabilities of “redifferentiation”, the self-organized boundary position maps onto the one-dimensional case, however, with the additional feature of a smooth boundary that cannot be obtained with a trivial extension of the model. Besides this *stochastic time scale separation* of the two pattern-generating processes, an explicit time dependence of neighborhood-dependent redifferentiation is discussed, also leading to robust self-organization of smooth domain boundaries.

10.2 Problem and model

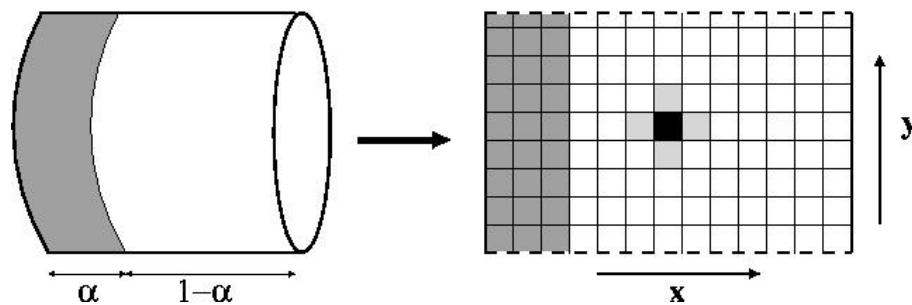


Figure 10.1: Basic geometry of the two-dimensional model. Pattern formation is assumed to take place on the surface of a cylinder (left diagram), αN cells on the left side of the surface take state $\sigma = 2$ (dark grey). The geometry is mapped to a regular grid with fixed boundaries in x -direction and periodic boundaries in y -direction (right diagram). Every single cell updates its state depending on a von-Neumann-neighborhood (black cell, with neighbors shaded in light grey).

So far, self-organization of position information and spatial pattern formation, inspired by experimental observations made for the fresh water polyp *Hydra*, was approximated as a one-dimensional problem. However, the *Hydra* body rather has a tube-like structure, with two layers of tissue (epithelia) surrounding a hollow interior (the gastrovascular space) - hence pattern control in the adult polyp may better be considered as a two-dimensional problem, with pattern formation processes taking place on the surface of a cylinder. Consequently, the problem of spatial domain formation and -control as introduced in) generalizes in two dimensions as sketched in Fig. 10.1.

Expression domain boundaries observed in experiments, as already outlined in the introduction, are sharp and well-localized on the head-foot axis. This observation of “smooth boundaries” can not be explained by a trivial extension of the 1D-model as a set of N parallel strings of cells: in this case, initial pattern formation leads to a “rugged” boundary, with Gaussian distributed deviations from the mean position (compare Fig. 10.2, panels 1)b) and 2)b)). In addition, in *Hydra* there is a flux of

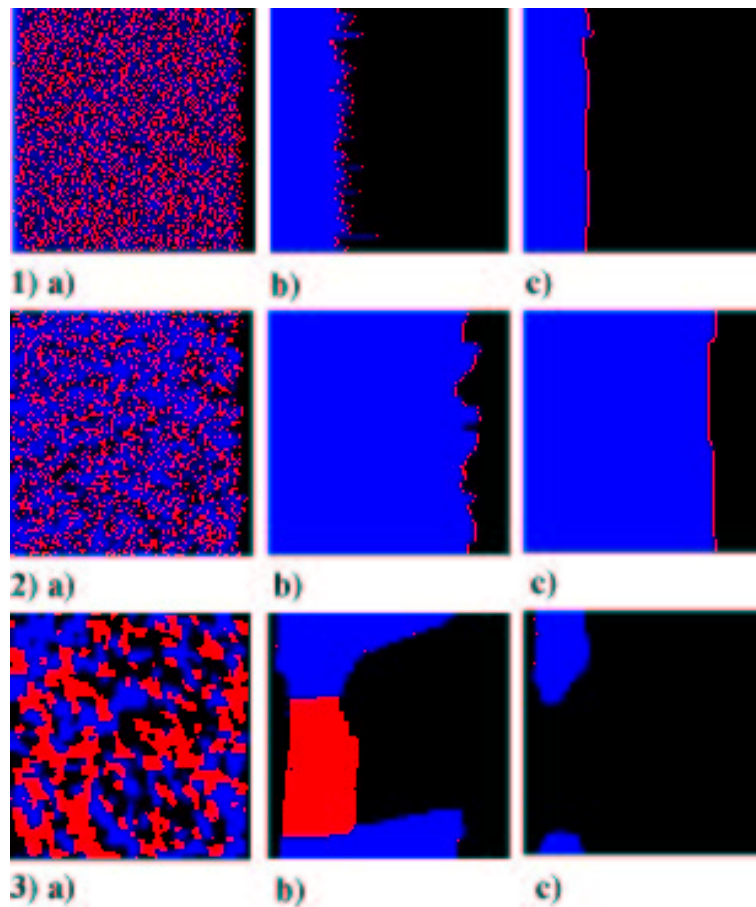


Figure 10.2: Snapshots of CA simulations for a system of size 100×100 cells, time is running from the left to the right. Parameters were chosen as follows: 1)a) - c) $p_h = 0.01$, 2)a) - c) $p_h = 0.2$, 3) a) - c) $p_h = 0.95$.

cells directed towards head and foot due to continued proliferation of stem cells, and cells redifferentiate when they cross expression boundaries. The question is, whether in 2D there might exist a single mechanism that could explain both observations, in addition to the mechanism of initial pattern formation as proposed in chapter 8. We will show that simple neighborhood-dependent redifferentiation of cell states is a candidate mechanism that may explain the observations.

In the following, we study a model implementation as a three-state two-dimensional cellular automaton with parallel update.

Let (i, j) denote the position of a cell on a two-dimensional grid of $N = N_1 \times N_2$ cells with von-Neuman-neighborhood and periodic boundary conditions in y -direction (in x -direction, boundary conditions are the same as in the 1D model). Then, at

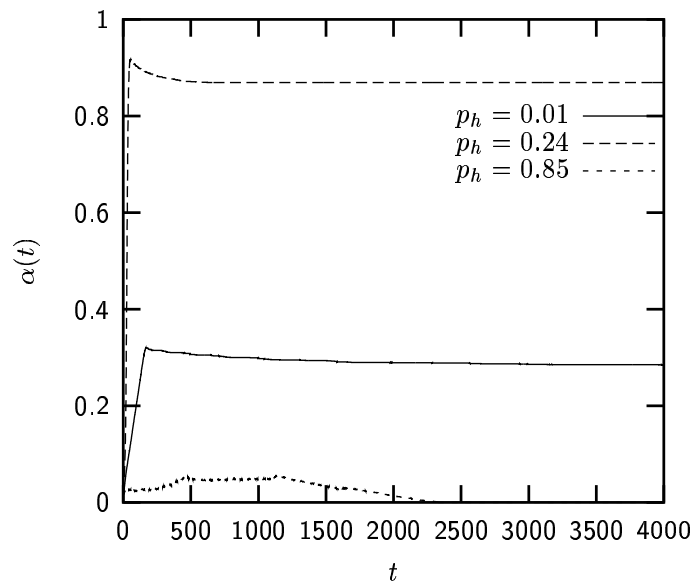


Figure 10.3: Time series of the average relative boundary position in x -direction, for three different values of the homogenization probability p_h . System size was 200×200 cells.

timestep t ,

- with probability p_h adjust the state $\sigma_{i,j}(t)$ of cell (i, j) to the state which is shared by the relative majority of the four neighbor cells at time $t - 1$ (if there is no such majority state, e.g. if two cells have $\sigma = 0$ and the other two have $\sigma = 1$, throw a coin)
- with probability $1 - p_h$, update $\sigma_{i,j}$ according to

$$\sigma_{ij}(t) = f(\sigma_{i-1,j}(t-1), \sigma_{i,j}(t-1), \sigma_{i+1,j}(t-1)), \quad (10.1)$$

where $f(t)$ is defined by the three-state rule table introduced in chapter 8.

Obviously, de-novo pattern formation will depend crucially on the “homogenization parameter” p_h : if p_h is too large, domain formation may fail completely, if it is too small, however, smoothing of domain boundaries will operate on a very slow time scale. An appropriate choice (i.e. an appropriate *time scale separation* between both pattern-organizing processes), however, may lead to “optimal” pattern formation. Notice that, for the sake of time scale separation, p_h does not necessarily have to be a stochastic parameter, a periodic process would work as well (in this case, initial pattern formation must occur with a short periodicity, interrupted by redifferentiation events which are separated by long time periods). However, stochasticity of cell differentiation, e.g. in

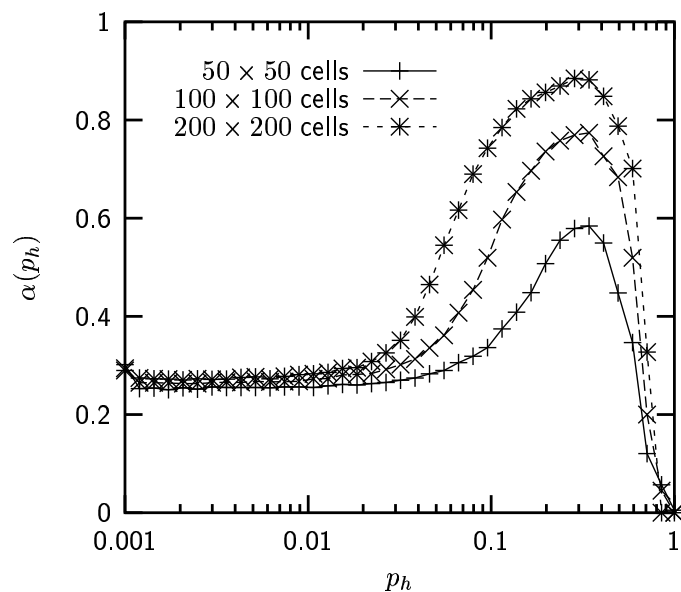


Figure 10.4: Average boundary position $\alpha(p_h)$ as a function of the homogenization probability p_h , three different system sizes. For small p_h , the value $\alpha_\infty \approx 0.281$ of the 1D-model is restored, a maximum is attained around $p_h \approx 0.3$; for $p_h \rightarrow 1$, α vanishes, i.e. no σ_2 -domain is formed any more (compare Fig. 2, 3 c).

stem cell production and -redifferentiation, indeed has been observed experimentally (Abkowitz, Catlin and Guttorp (1996)), hence we constrain our study to a stochastic parameter p_h . As a second szenario, we propose that p_h may be an explicit function of time, e.g. as a result of a “autocatalytic reaction” competing with the dynamics of initial pattern formation. For simplicity, we choose

$$p_h(t) = \begin{cases} f(t) & \text{if } f(t) < p_{max} \\ p_{max} & \text{else} \end{cases} \quad (10.2)$$

with $f(t) = p_0 \cdot \exp[t/\tau]$ and $0 < p_{max}, p_0 < 1$ in this case.

10.3 Results

10.3.1 Stochastic time scale separation

One finds in numerical simulations that the proposed two-stage pattern formation works over a wide range of p_h ; only for very large values $p_h \approx 1$ domain formation fails. Snapshots of simulations for three different parameter values are shown in Fig. 10.2. One finds, that for small values the boundary position $\alpha_\infty \approx 0.281$ is restored,

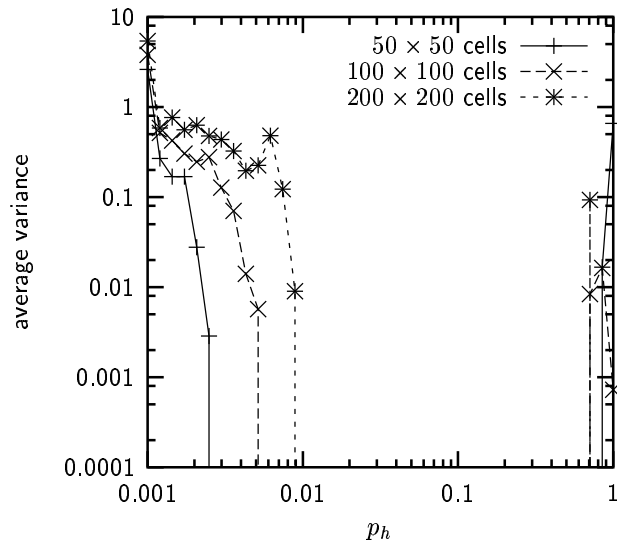


Figure 10.5: Average variance (number of cells squared) of the boundary position in x -direction, averaged over 50 different initial conditions for each data point, as a function of p_h . Data points were sampled after 10000 dynamical updates. For $p_h > 0.002$, the variance is smaller than one cell on average, i.e. practically zero for most initial conditions.

and neighborhood-dependent redifferentiation finally leads to smooth boundaries. For larger values of p_h , α is shifted to larger values (i.e. towards the right handside). In the following, we will confirm these observations with quantitative results.

Let $\alpha_i(t)$ denote the domain boundary position measured in the i th row of the grid; then

$$\langle \alpha \rangle(t) = \frac{1}{n_{rows}} \sum_{i=1}^{n_{rows}} \alpha_i(t) \quad (10.3)$$

defines the average boundary position at time t , where n_{rows} is the total number of rows (in simulations, we chose a square grid, i.e. $n_{rows} = \sqrt{N}$).

Typical time series for a system of 200×200 cells are shown in Fig. 10.3. Initial pattern formation takes place on a very fast time scale (fast increase of α), and the time needed for relaxation against the equilibrium position strongly depends on p_h ; it is fastest for $p_h \approx 0.3$. For $p_h \geq 0.85$, domain formation fails and $\alpha \rightarrow 0$ in equilibrium.

Interference between both pattern formation processes is captured quantitatively in Fig. 10.4, which shows $\langle \alpha \rangle(p_h)$ after 10000 updates as a function of p_h . One finds, that for small p_h there is no interference, i.e. the average boundary position of the 1D-model is restored. For $p_h \approx 0.3$, interference becomes maximal and the domain boundary is shifted strongly to the right system side. For even larger values of p_h , the

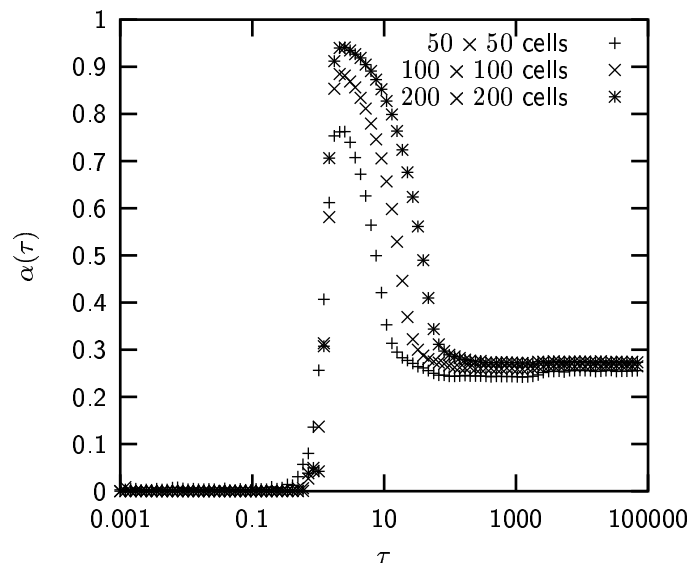


Figure 10.6: Average boundary position $\alpha(\tau)$ as a function of the time constant τ , for explicit time dependence of the homogenization probability p_h , as explained in the text. For small τ , $\alpha(\tau)$ vanishes, a maximum is reached for $\tau \approx 1$, for large τ the equilibrium value $\alpha_\infty \approx 0.281$ of the 1D-model is restored.

domain size rapidly goes to zero.

Effectiveness of boundary “smoothing” was analyzed by measuring the (spatial) variance

$$\text{Var}(\alpha) = \frac{1}{n_{rows}} \sum_{i=1}^{n_{rows}} (\alpha_i - \langle \alpha \rangle)^2 \quad (10.4)$$

of the boundary position after 10000 dynamical updates.

Results are shown in Fig. 10.5 for three different system sizes. One finds that the variance goes to zero already for very small values of p_h , i.e. boundaries indeed become smooth, even in a parameter region where no interference between initial symmetry-breaking and neighborhood-dependent redifferentiation is found.

10.3.2 Explicit time dependence of p_h

Here, simulations were made with the parameter choices $p_0 = 0.005$ and $p_{max} = 1.0$, and the boundary position was measured as a function of the time constant τ as defined above. One finds $\langle \alpha \rangle = 0$ for $\tau < 1$, then there is a rapid increase, followed by a decrease, and for $\tau > 100$ (with a slight shift towards larger values of τ with increasing system size) the equilibrium value of the 1D-model is restored. Boundary smoothness (i.e. zero variance) is obtained very fast in this szenario, as the probability

for neighborhood-dependent redifferentiation goes to $p_{max} = 1$ within $\tilde{t} = -\tau \ln p_0$ timesteps (results not shown).

10.4 Summary

We studied a two-dimensional extension of the pattern formation model introduced in chapter 8. A second pattern generating mechanism - neighborhood-dependent state changes (“redifferentiation” of cells) - was introduced, that competes with symmetry breaking during initial pattern formation. We showed that, for small probabilities of “redifferentiation”, the self-organized boundary position maps onto the one-dimensional case, however, with the additional feature of a smooth boundary that cannot be obtained with a trivial extension of the model. Besides this *stochastic time scale separation* of the two pattern-generating processes, an explicit time dependence of neighborhood-dependent redifferentiation was discussed, also leading to robust self-organization of smooth domain boundaries.

Chapter 11

Conclusions and Outlook

More than thirty years after the pioneering studies on Random Boolean Networks published by Kauffman (1969), statistical physics of complex dynamical networks has become a very active field of research again. These efforts take place in the face of more and more detailed data available for gene regulatory networks, that contribute significantly to the complexity of biological organisms by self-regulated, context-dependent, combinatorial retrieval of the information encoded in the genome. However, a detailed understanding of the global dynamics, structure and functions of these biological networks is still missing, increasing the need for new theoretical concepts and methods.

In this thesis, ideas from the statistical physics of discrete dynamical networks have been combined with concepts of cellular automata and emergent computation in order to understand basic dynamical properties of complex networks, to identify possible evolutionary processes leading to network topologies as observed in experiment, and to model the interplay of regulatory networks in morphogenesis of multicellular organisms.

The starting point was the statistical physics of sparsely connected **Random Threshold Networks (RTN)**. By a combinatorial derivation of the probability for propagation of dynamical perturbations as a function of the nodes' in-degree, combined with an annealed approximation, **we were able to calculate the critical connectivity analytically for the case of a degree-distribution following a Poissonian** (Rohlf and Bornholdt, 2002); more recently, it has been shown that this method indeed can be applied to RTN with arbitrary degree-distributions, e.g. following power-laws (Nakamura, 2003). **Local “damage suppression” at network hubs, as observed in RTN, can have global dynamical effects, as was demonstrated in this thesis for networks with scale-free topology.**

As a second step, evolution of dynamical networks based on local rewiring rules was studied for a simplified version of the model introduced by Bornholdt and Rohlf (2000) and discussed in the context of the evolution of gene regulatory networks. **It was**

found that this algorithm, based on a coupling between a local, dynamical order parameter (activity of nodes) and a topological control parameter (in-degree of nodes), **not only leads to self-organization of network topologies close to criticality, but also to a non-random distribution of dynamical activity and to a degree-distribution that deviates from a Poissonian.** These results compare well with data obtained from statistical analyses of the topology of real gene regulatory networks.

The third step of this thesis was motivated by the question, how well-defined functional tasks, e.g. the robust self-organization of a particular spatial pattern during **morphogenesis**, can be solved by network-based information processing, and how such processes are related to network complexity and structure. **Two experimental observations in the polyp *Hydra* have been taken as a starting point: “reboot-like” de-novo pattern formation from random cell aggregates, and scale-invariant regulation of the size gene expression domains.** First, a one-dimensional cellular automata model that solves this pattern formation task was derived by aid of genetic algorithms; second, this model was mapped on a system of locally coupled dynamical networks. **Contrary to models based on reaction-diffusion systems, in this new model no parameter tuning is needed, only a local bias in information transfer between cells**, e.g. due to a spatially asymmetric distribution of signal-transmitting receptor systems. Size and complexity of the network, as well as its hierarchical structure, compare well to observations made for “developmental modules” in multicellular organisms.

Morphogenesis operates with astonishing reliability in presence of various sources of noise, acting on a large range of scales. **Robustness of the proposed model of morphogenesis was studied with respect to two types of perturbations**, that frequently occur in development: random dynamical update errors and cell movements, e.g. due to continued proliferation of cells. By a comparison of a mean-field analysis with numerical simulations, it was shown that **noise-induced excitations (“particles”)** contribute to stabilization of spatial patterns, even in the presence of cell flow. Two phase transitions have been identified: **a first order phase transition at vanishing noise, and a second order phase transition at increased cell flow.**

Last, the model was extended into **two dimensions.** The mechanism of spatial symmetry breaking studied in the 1D-model was combined with a second, local dynamical rule: **neighborhood-dependent state changes** (“redifferentiation”) of cells, both mechanisms decoupled by stochastic time scale separation. **It was shown that this leads to robust self-organization of spatial activity domains with smooth and localized boundaries**, an essential prerequisite for robust regulation of position information in a two-dimensional tissue.

To conclude, numerical and analytical methods developed for discrete dynamical networks and spatially extended, discrete dynamical systems as, for example, cellular automata, have proven to be versatile tools to address questions ranging from network dynamics and -structure over network evolution to specific problems in morphogenesis and pattern formation. Our studies support the impression that information processing in regulatory networks can contribute significantly to the emergence of complexity and robustness in biological organisms; similar results have been found in other models of regulatory networks and morphogenesis that are, e.g., based on coupled differential equations (von Dassow *et al.*, 2000; Furusawa and Kaneko, 2001; Salazar-Ciudad, Solé, and Newman, 2001). Today, in most cases direct comparisons to molecular data are still not possible, however, from simple dynamical models as investigated in this thesis, one may well get an impression of network complexity and -structure necessary to generate specific dynamical features observed in experiments, or to solve certain functional tasks as they occur during morphogenesis. For example, studies of this kind may lead to identification of sub-networks or (functional) “modules” (Solé, Salazar-Ciudad and Garcia-Fernandez, 2002), and thereby support the identification of similar regulatory circuits in biological organisms. In any case, still much research remains to be done. Concerning the train of thoughts followed in this thesis, an integrated – possibly analytical – understanding of the dynamical properties of discrete dynamical networks is still lacking, thereby limiting the study of models of network evolution as well as possible applications to biological networks; in principle, the same holds for systems of coupled networks. Nevertheless, it is not too far-fetched to assume that in future research a combination of methods provided by Statistical Physics, dynamical systems theory and Bioinformatics, in addition to their experimental counterparts in molecular and developmental biology, will become more and more essential for a deeper understanding of the overwhelming complexity found in living organisms.

Results of this thesis have appeared in the following publications and preprints:

- Rohlf, T. and Bornholdt, S. (2002). *Criticality in Random Threshold Networks: Annealed Approximation and Beyond*. *Physica A* **310**, 245-259
- Rohlf, T. and Bornholdt, S. (2003). *Self-organized pattern formation and noise-induced control from particle computation*. Preprint: cond-mat/0312366, submitted
- Rohlf, T. and Bornholdt, S. (2004). *Gene Regulatory Networks: A Discrete Model of Dynamics and Topological Evolution*. In: *Function and regulation of cellular systems - Experiments and Models*, Eds. Deutsch, A., Howard, H., Falcke, M. and Zimmermann, W., Birkhäuser Verlag Basel

- Rohlf, T. and Bornholdt, S. (2004). *Morphogenesis by coupled regulatory networks*. Preprint: q-bio.MN/0401024, submitted

Appendix A

Evolution of Cellular Automata by Genetic Algorithms (GA)

In this chapter we will briefly recapitulate how the cellular automata model of spatial pattern formation, that has been introduced in chapter 8, has been found by application of a *genetic algorithm (GA)*.

A.1 Definition of the GA

In order to find a set \mathcal{F} of update rules that solve the problem as formulated chapter 8, cellular automata have been evolved using genetic algorithms (Mitchell, Hraber and Crutchfield, 1994). Genetic algorithms are population-based search algorithms, which are inspired by the interplay of random mutations and selection as observed in biological evolution (Holland, 1975). Starting from a randomly generated population of P rule tables f_n , the algorithm optimizes possible solutions by evaluating the fitness function

$$\Phi(f_n) = \frac{1}{(T_u - T) \cdot N_C} \sum_{t=T_u-T}^{T_u} \left(\sum_{i=0}^{[\alpha \cdot N_C]-1} \delta_{\sigma_i^n(t),2} + \sum_{i=[\alpha \cdot N_C]}^{N_C-1} \{1 - \delta_{\sigma_i^n(t),2}\} \right). \quad (\text{A.1})$$

The optimization algorithm then is defined as follows:

1. Generate a random initial population $\mathcal{F} = \{f_1, \dots, f_P\}$ of rule tables.
2. Randomly assign system sizes $N_C^{min} \leq N_c^n \leq N_C^{max}$ to all rule tables.

3. For each rule table, generate a random initial state vector.
4. Randomly mutate one entry of each rule table (generating a population \mathcal{F}^* of mutants).
5. Iterate dynamics over T_u time steps for \mathcal{F} and \mathcal{F}^* .
6. Evaluate $\Phi(f_n)$ and $\Phi(f_n^*)$ for all rule tables $0 < n \leq P$, averaging over the past $T_u - T$ update steps (with an additional penalty term if f_n does not converge to a fixed point).
7. For each n , replace f_n with f_n^* , if $\Phi(f_n) \leq \Phi(f_n^*)$.
8. Replace the least fit solution by a duplicate of the fittest one.
9. Go back to step 2 and iterate.

A.2 Evolution of cellular automata

The outcome of this search algorithm is a set of rule tables, which then can be “translated” into (spatially coupled) Boolean networks or threshold networks with suitable thresholds. This yields a set of (minimal) dynamical networks which solve the pattern formation task by means of internal information processing.

The genetic algorithm sketched above is run with the following parameter choices: $15 \leq N_C \leq 150$, i.e. during GA runs the system size is varied randomly between 15 and 150 cells, and the population size is set to $P = 100$. Fig. A.1 shows the fitness of the highest fitness mutant and the average fitness of the population as a function of the number of successive mutation steps during optimization. A useful solution is found rather quickly (after about 200 updates), with further optimization observed during further 10000 generations. At time step 10000 mutations are turned off, thus now the average fitness of the established population under random initial conditions and random fluctuations of the system size N_C is tested. The average fitness $\Phi \approx 0.98$ indicates a surprisingly high robustness against fluctuations in the initial start pattern, indicating that the system is capable of *de novo pattern formation*. In the “fitness picture”, Fig. A.2 confirms that the dynamically regulated domain size ratio $\alpha/(1 - \alpha)$ indeed is independent of system size (proportion regulation), there is only a weak decay of the fitness at small values of N_C . Interestingly, one also observes that for regeneration of Hydra polyps from random cell aggregates a minimum number of cells is required (Technau *et al.*, 2000). The model suggests that this observation might be explained by the dynamics of an underlying pattern generating mechanism, i.e. that

there has to be a minimum diversity in the initial condition for successful de novo pattern formation.

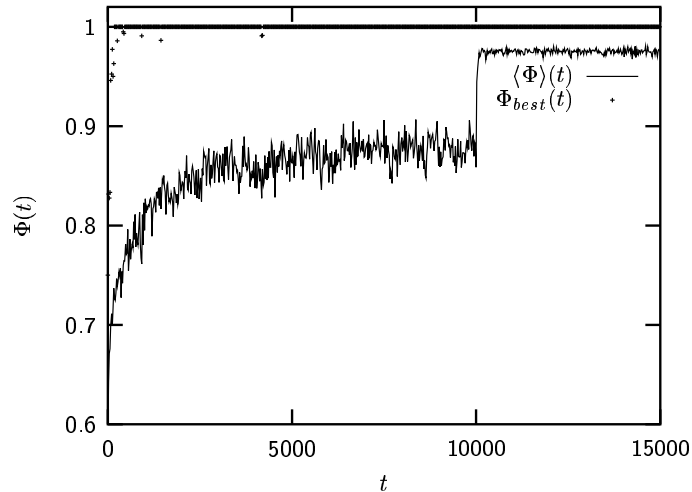


Figure A.1: Average fitness $\langle \Phi \rangle(t)$ of the mutant population \mathcal{F}^* and fitness of the highest fitness mutant $\Phi_{best}(t)$ as a function of simulation time during the genetic algorithm run that lead to the high fitness solution used in this paper. At time step 10000 mutations were turned off, in order to test the established population of optimized rule tables under different initial conditions (this corresponds to the sharp increase of $\langle \Phi \rangle(t)$ at time step 10000). The evolved population of rule tables has an average fitness of about 0.98, independent from the initial conditions and system size N_C (in the tested range, i.e. $15 \leq N_C \leq 150$).

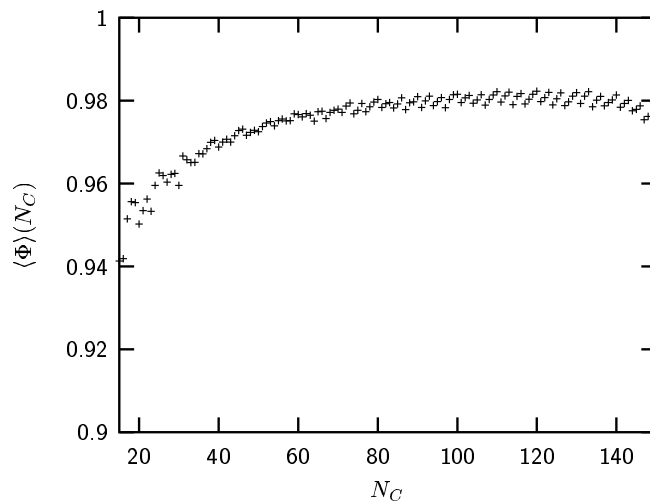


Figure A.2: Average fitness of the highest fitness rule table as a function of the system size N_C . For system sizes $N_C \geq 80$ the fitness is almost constant at about 0.98. Notice that the decrease of the fitness for small N_C is an effect of the *dynamics*, not of the genetic algorithm implementation (all N_C in the range $15 \leq N_C \leq 150$ were tested with equal probability), hence the dynamics of pattern formation may impose a lower boundary on the range of system (animal) sizes tolerated by natural selection.

Appendix B

Statistical analysis of CA solutions

An interesting question is how “difficult” it would be for an evolutionary process driven by random mutations and selection to find solutions for the pattern formation problem based on neighbor interactions between cells.

As we showed above, the genetic algorithm finds the correct solution fast, however, this does not necessarily mean that biological evolution could access the same solution as fast. If there is only one, singular solution, evolution may never succeed finding it, as the genotype which already exists cannot be modified in an arbitrary way without possibly destroying function of the organism (*developmental constrains*). To illustrate this point, we generated an ensemble of $N_E = 80$ different solutions with $\Phi \geq 0.96$ and performed a statistical analysis of the rule table structure.

As one can see in Fig. B.1, some positions in the rule table are quite fixed, i.e., there is not much variety in the outputs, whereas other positions are more variable. However, the output frequency distribution alone does not allow to really judge the “evolvability” of the solutions: if there are strong correlations between most of the rule table entries, evolutionary transitions from one solution to another would be almost impossible. To check this point, we studied statistical two point correlations between the rule table entries. The probability for finding state σ at position a and state σ' at rule table position b is given by

$$p^{ab}(\sigma, \sigma') = \frac{1}{N_E} \sum_{n=1}^{N_E} \delta_{\sigma^n(a), \sigma} \cdot \delta_{\sigma^n(b), \sigma'}, \quad (\text{B.1})$$

where δ is the Kronecker symbol and n runs over the statistical ensemble of size N_E . The two point correlation between a and b then is defined as

$$C^{ab} = c_1 \left(\max_{(\sigma, \sigma')} p^{ab}(\sigma, \sigma') - c_2 \right) \quad (\text{B.2})$$

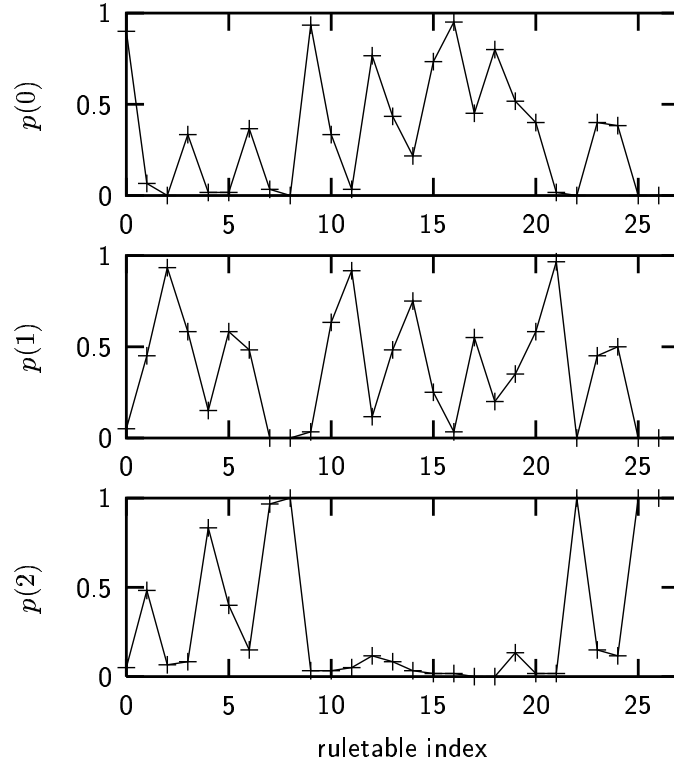


Figure B.1: Frequency distribution $p(\sigma_i)$ of outputs as a function of the rule table index as denoted in table 1. Ensemble statistics is taken over 80 different solutions with $\Phi \geq 0.96$. The upper panel shows the distribution for $\sigma_i = 0$, the middle panel the distribution for $\sigma_i = 1$ and the lower panel the distribution for $\sigma_i = 2$.

with $c_1 = 9/8$ and $c_2 = 1/9$ to obtain a proper normalization with respect to the two limiting cases of equal probabilities ($p^{ab}(\sigma, \sigma') = 1/9 \quad \forall(\sigma, \sigma')$) and $p^{ab}(\sigma, \sigma') = 1$ for $\sigma = \tilde{\sigma}, \quad \sigma' = \tilde{\sigma}'$ and $p^{ab}(\sigma, \sigma') = 0$ for all other (σ, σ')). Fig. B.2 shows the frequency distribution of $C^{ab}(\sigma, \sigma')$, averaged over all possible pairs (a, b) . About 65% of rule table positions are strongly correlated ($C^{ab} = 1.0$), the rest shows correlation values between 0.3 and 1.0. Hence, we find that the space of solutions is restricted, nevertheless there is variability in several rule table positions. To summarize this aspect, the pattern formation mechanism studied in this paper shows considerable robustness against rule mutations, however, a “core module” of rules is always fixed. Interestingly, a similar phenomenon is observed in developmental biology: Regulatory “modules” involved in developmental processes often are evolutionarily very conservative, i.e., they are shared by almost all animal phyla (Davidson, 2001), while morphological variety is created by (few) taxon specific genes (Bosch and Khalturin, 2002) and “rewiring” of existing developmental modules.

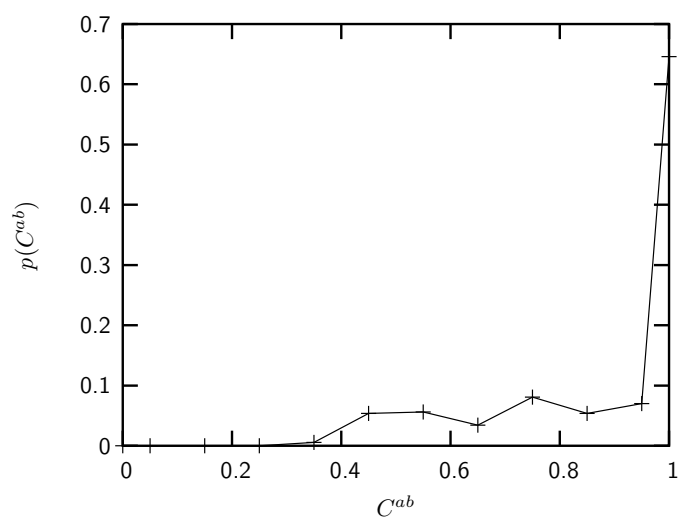


Figure B.2: Frequency distribution $p(C^{ab})$ of two point correlations $C^{ab}(\sigma, \sigma')$ of rule table entries, as defined in Eqn. B.2, averaged over all possible pairs of rule table entries. About 65% of rule table entries have correlation 1.0, for the rest the correlation is lower.

Bibliography

- Abkowitz, J.L., Catlin, S.N. and Gutter, P. (1996). *Evidence that hematopoiesis may be a stochastic process in vivo*. *Nature Medicine* **2**, 190-197
- Akutsu, T., Miyano, S. and Kuhara, S. (1999). *Identification of genetic networks from a small number of gene expression patterns under the boolean network model*. *Pacific Symposium on Biocomputing* **4**, 17-28
- Albert, R. and Barabasi A.-L. (2000). *Dynamics of complex systems: Calculating the period of boolean networks*. *Phys. Rev. Lett.* **84**, 5660 - 5663
- Albert, R. and Barabási, A. L. (2002). *Statistical mechanics of complex networks*. *Rev. Mod. Phys.* **74**, 47
- Albert, R, Jeong, H. and Barabási, A. L. (2000). *Attack and error tolerance in complex networks*. *Nature* **406**, 378-382.
- Alberts, B., Johnson, A., Lewis, J., Raff, M., Roberts, K. and Walter, P. (2002). *Molecular biology of the cell (Fourth edition)*. Garland Science, New York
- Alon, U., Surette, M. G., Barkai, N. and Leibler, S. (1999). *Robustness in bacterial chemotaxis*. *Nature* **397**, 168-171.
- Amaral, L. A. N., Scala, A., Barthélemy, M., and Stanley, H. E. (2000). *Classes of small-world networks*. *Proc. Natl. Acad. Sci. U.S.A.*, **97**(21), 11149–11152.
- Arnone, M.I. and Davidson, E.H. (1997). *The hardwiring of development: organization and function of genomic regulatory systems*. *Development* **124**, 1851-1864
- Artavanis-Tsakonas, S., Rand, M.D. and Lake, R.J. (1999) *Notch Signaling: Cell Fate Control and Signal Interpretation in Development*. *Science* **284**, 770-776
- Arthur, W. (1997). *The Origin of Animal Body Plans - A Study in Evolutionary Developmental Biology*. Cambridge University Press

- Barabási, A. L. and Albert, R. (1999). *Emergence of scaling in random networks*. Science **286**, 509-512
- Barabási, A. L. (2002). *Linked - The New Science of Networks*. Perseus Books.
- Barkai, N. and Leibler, S. (1997). *Robustness in Simple Biochemical Networks*. Nature, **387**, 913-916.
- Barrat, A. and Weigt, M. (2000). On the properties of small-world networks models. *Eur. Phys. J. B*, **13**, 547-560.
- Bastolla, U. and Parisi, G. (1996). *Closing probabilities in the Kauffman model: An annealed computation*. Physica D **98**, 1-25
- Bastolla, U. and Parisi, G. (1998a). *The modular structure of Kauffman networks*. Physica D **115**, 219-233
- Bastolla, U. and Parisi, G. (1998b). *Relevant elements, magnetization and dynamical properties in Kauffman networks: A numerical study*. Physica D **115**, 203-218
- Bhattacharjya, A. and Liang, S. (1996). *Power-Law Distributions in Some Random Boolean Networks*. Phys. Rev. Lett. **77**, 1644-1647
- Bilke, S. and Sjunnesson, F. (2001). *Stability of the Kauffman model*. Phys. Rev. E **65**, 016129
- Becskei, A. and Serrano, L. (2000). *Engineering stability in gene networks by autoregulation*. Nature **405**, 590-593
- Berlekamp, E.R., Conway, J.H. and Guy, R.K. (1982). *What is Life?* Chapter 25 in: *Winning Ways for Your Mathematical Plays, Vol. 2: Games in Particular*. Academic Press.
- Bhalla, U.S. and Iyengar, R. (1999). *Emergent properties of networks of biological signaling pathways*. Science **283**, 381-387
- Bode, P.M. and Bode, H.R. (1984) in *Pattern Formation*, eds. Malacinski, G. and Bryant, S.V. (MacMillan, New York), 213-241
- Bonabeau, E. (1997). *From Classical Models of Morphogenesis to Agent-Based Models of Pattern Formation*. Artificial Life **3**, 191-211, MIT Press 1997
- Bornholdt, S. and Sneppen, K. (1998). *Neutral Mutations and Punctuated Equilibrium in Evolving Genetic Networks*. Phys. Rev. Lett. **81**, 236-239

- Bornholdt, S. and Rohlf, T. (2000). *Topological Evolution of Dynamical Networks: Global Criticality from Local Dynamics*. Phys. Rev. Lett., 84, 6114 - 6117
- Bornholdt, S. and Sneppen, K. (2000). *Robustness as an evolutionary principle*. Proc. R. Soc. Lond. B **267**, 2281-2286
- Bornholdt, S. (2001). *Modeling Genetic Networks and Their Evolution: A Complex Dynamical Systems Perspective*. Biol. Chem. **382**, 1289-1299
- Bosch, T.C.G. (1998). *Hydra*. In: Cellular and Molecular Basis of Regeneration: From Invertebrates to Humans, P. Ferretti and J. Geraudie (Eds.), 111-134. Wiley, Sussex.
- Bosch, T.C.G and Khalturin, K. (2002). *Patterning and cell differentiation in Hydra: novel genes and the limits to conservation*. Canadian Journal of Zoology **80**, 1670-1677.
- Bosch, T.C.G. (2003). *Ancient signals: peptides and the interpretation of positional information in ancestral metazoans*. Comp. Biochem. and Physiol. B, **136**, 185.
- Broun, M., Sokol, S. and Bode, H.R. (1999). *Cngsc, a homologue of goosecoid, participates in the patterning of the head, and is expressed in the organizer region of Hydra*. Development **126**, 5245-5254
- Carey, M., Lin, Y. Green, M.R. and Ptashne, M. (1990). *A mechanism for synergistic activation of a mammalian gene by GAL4 derivatives*. Nature **345**, 361 - 364
- Chen, C.S., Mrksich, M., Huang, S., Whitesides, G.M. and Ingber, D.E. (1997). *Geometric Control of Cell Life and Death*. Science **276**, 1425-1428
- Chapman (2002). *Life Universal Computer*. <http://www.igblan.com/ca/>.
- Chung, F., Lu, L. Dewey, T.G. and Galas, D.J. (2003). *Duplication Models for Biological Networks*. cond-mat/0209008
- Claverie, J.M. (2001). *Gene Number. What If There Are Only 30,000 Human Genes?* Science **291**, 1255
- Collier, J.R., Monk, N.A., Maini, P.K. and Lewis, J.H. (1996). *Pattern formation by lateral inhibition with feedback: mathematical model of delta-notch intercellular signalling*. J. Theor. Biol. **183**, 429-446
- Crutchfield, J.P. and Mitchell, M. (1995). *The Evolution of Emergent Computation*. Proc. Natl. Acad. Sci. (USA), **92**, 10742.

- Das, R., Mitchell, M. and Crutchfield, J.P. (1994). *A Genetic Algorithm Discovers Particle Computation in Cellular Automata*. In: *Parallel Problem Solving from Nature-III*, eds. Davidore, Y. et al., 344, Springer, Berlin.
- Darwin, C. (1859). *On the Origin of Species by means of Natural Selection*. John Murray, London.
- von Dassow, G., Meir, Eli, Muro, E.M. and Odell, G.M. (2000). *The segment polarity network is a robust developmental module*. *Nature* **406**, 188-192
- David, C.N. and Campbell, R.D. (1972). *Cell cycle kinetics and development of Hydra-Attenuata. 1. Epithelial cells*. *J. Cell. Sci.* **11**, 557.
- Davidson, J., Ebel, H., and Bornholdt, S. (2002). Emergence of a small world from local interactions: Modeling acquaintance networks. *Phys. Rev. Lett.*, **88**, 128701.
- Davidson, E.H. (2001). *Genomic Regulatory Systems. Development and Evolution*. Academic Press.
- Dawkins, R. (1976). *The selfish gene*. Oxford University Press
- Dearden, P. and Akam, M. (2000). *Segmentation in silico*. *Nature* **406**, 131-132
- Derrida, B. and Pomeau, Y. (1986). *Random networks of automata: a simple annealed approximation*. *Europhys. Lett.*, **1**, 45-49
- Derrida, B. and Stauffer, D. (1986). *Phase transitions in two-dimensional Kauffman cellular automata*. *Europhys. Lett.*, **2**, 739ff
- Derrida, B. und Flyvbjerg H. (1987). *Distribution of local magnetisations in random networks of automata*. *J. Phys. A*, **20**, L1107-L1112
- Derrida, B., Gardner, E. and Zippelius, A. (1987). *Europhys. Lett.* **4**, 167.
- Derry, L.A., Kaufman, A. and Jacobsen, S.B. (1992). *Sedimentary cycling and environmental change in the late proterozoic - evidence from stable and radiogenic isotopes*. *Geochim. Et Cosmochim. Acta*, **56**, 1317-1329
- Dewey, T.G. and Galas, D. (2001). *Dynamic Models of Gene Expression and Classification*. *Func. Integr. Genomics* **1**, 269-278
- Dorogovtsev, S. N. and Mendes, J. F. F. (2001). *Phys. Rev. E* **63**, 056125, 1-18

- Dorogovtsev, S. N. and Mendes, J. F. F. (2002). *Evolution of networks*. Adv. Phys. **51**, 1079-1187
- Drasdo, D. and Forgacs, G. (2000). *Modeling the Interplay of Generic and Genetic Mechanisms in Cleavage, Blastulation and Gastrulation*. Dev. Dyn. **219**, 182-191
- Ebel, H., Mielsch, L.-I., and Bornholdt, S. (2002). Scale-free topology of e-mail networks. *Phys. Rev. E*, **66**, 035103(R).
- Ebel, H. (2002). *Statistical Physics of Complex Networks and Coevolutionary Games*. PhD Thesis, Kiel University.
- Edelman, G. (1988). *Topobiology. An Introduction to Molecular Embryology*. Basic Books, Inc., New York
- Eggenberger, P. (1997). *Evolving Morphologies of Simulated 3d Organisms Based on Differential Gene Expression*. Proceedings of the Third Conference on Artificial Evolution (EA'97). Springer, Berlin, 251-262
- Endl, I., Lohmann, J.U. and Bosch, T.C.G. (1999). *Head-specific gene expression in Hydra: Complexity of DNA-protein interactions at the promotor of ks1 is inversely correlated to the head activation potential*. Proc. Natl. Acad. Sci. USA **96**, 1445-1450
- Entchev, E.V., Schwabedissen, A. and Gonzales-Gaitan, M. (2000). *Gradient formation of the TGF-beta homolog Dpp*. Cell **103**, 981-991
- Erdős, P. and Rényi, A. (1960). On the evolution of random graphs. *Publ. Math. Inst. Hung. Acad. Sci.*, **5**, 17-61.
- Faloutsos, M., Faloutsos, P., and Faloutsos, C. (1999). On power-law relationships of the internet topology. In *Proceedings of the ACM SIGCOMM '99 Conference of Applications, Technologies, Architectures, and Protocols for Computer Communication*, pages 251-262. ACM Press, New York.
- Flyvbjerg, H. and Kjaer, N.J. (1987). *Exact solution of Kauffman's model with connectivity one*. J. Phys. A, **21**, 1695-1718
- Flyvbjerg, H. (1988). *An order parameter for networks of automata*. J. Phys. A, **21**, L955-L960
- Forman, B.J. and Javois, L. C. (1999). *Interactions between the foot and the head patterning systems in Hydra vulgaris*. Dev. Biol. **15**, 351-366

- Ftterer, C., Colombo, C., Jlicher, F. and Ott, A. (2003). *Morphogenetic oscillations during symmetry breaking of regenerating Hydra vulgaris cells*. *Europhys. Lett.* **7** **64** (1), 137-143
- Furusawa, C. und Kaneko, K. (2000). *Origin of Complexity in Multicellular Organisms*. *Phys. Rev. Lett.* **84**, 6130-6133
- Furusawa, C. und Kaneko, K. (2001). *Theory of Robustness of Irreversible Differentiation in a Stem Cell System: Chaos hypothesis*. *J. Theor. Biol.* **209**, 395-416
- Furusawa, C. und Kaneko, K. (2003). *Robust development as a consequence of generated positional information*. *J. Theor. Biol.* **224**, 413-435
- Gardiner, C.W. (1983). *Handbook of Stochastic Methods*. Springer-Verlag
- Gardner, M. (1983). *The Game of Life, Parts I-III*. In: *Wheels, Life and other Mathematical amusements*, W.H. Freeman and Company.
- Gardner, S. and Collins, J. (2000) *Neutralizing noise in gene networks*. *Nature*, 405, 520-521
- Gerhard, M. and Schuster, H. (1995). *Das digitale Universum - Zelluläre Automaten als Modelle der Natur*. Vieweg Verlag.
- Gierer, A., Berking, S., David, C.N., Flick, K., Hansman, G., Schaller, C.H. and Trenkner, E. (1972). *Regeneration of Hydra from reaggregated cells*. *Nature* **239**, 98-101
- Gierer, A. and Meinhardt, H. (1972). *A Theory of Biological Pattern Formation*. *Kybernetik* **12**, 30-39
- Gilbert, S.F. (1997). *Developmental Biology*. Sinauer Associates, Massachusetts
- Goodwin, B.C., Kauffman, S. and Murray, J.D. (1993). *Is Morphogenesis an Intrinsically Robust Process?* *J. theor. Biol.* **163**, 135-144
- Graner, F. and Glazier, J.A. (1992). *Simulation of Biological Cell Sorting Using a Two-Dimensional Extended Potts Model*. *Phys. Rev. Lett.* **69**, 2013-2016
- Guelzim, N., Bottani, S., Bourguin, P. and Kepes, F. (2002). *Topological and causal structure of the yeast transcriptional regulatory network*. *Nature Genetics* **31**, 60-63
- Gumbiner, B.M. (1996). *Cell Adhesion: the molecular basis of tissue architecture and morphogenesis*. *Cell* **84**, 345-357

- Gurdon, J.B. and Bourillot, P.Y. (2001). *Morphogen gradient interpretation*. Nature **413**, 797-803
- Hanson, J.E. (1993). *Computational Mechanics of Cellular Automata*. PhD thesis, University of California (USA), Berkeley.
- Hanson, J.E. and Crutchfield, J.P. (1997). *Computational Mechanics of cellular automata: an example*. Physica D, **103**, 169-189
- Hasty, J., Isaacs, F., Dolnik, M., McMillen, D. and Collins, J.J. (2001). *Designer gene networks: Towards fundamental cellular control*. Chaos **11**, 207-220
- Hasty, H., Dolnik, M., Rottschäfer, V. and Collins, J.J. (2002) *A synthetic gene network for entraining and amplifying cellular oscillations*. Phys. Rev. Lett., in press
- Hörstadius, S. (1935). Pubbl. Stn. Zool. Napoli **14**, 1
- Hörstadius, S. (1973). *Experimental Embryology of Echinoderms*. Clarendon, Oxford
- Hogeweg, P. (2000). *Evolving Mechanisms of Morphogenesis: on the Interplay between Differential Adhesion and Cell Differentiation*. J. Theor. Biol. **203**, 317-333
- Holland, J.H. (1975). *Adaptation in natural and artificial systems*. University of Michigan Press, Ann Arbor.
- Hopfield, J.J. (1984). *Neurons with graded response have collective computational properties like those of two-state neurons*. Proc. Natl. Acad. Sci. USA **81**, 3088-3092
- Hordijk, W. (1998), Crutchfield, J.P. and Mitchell, M. (1998). *Mechanisms of emergent computation in cellular automata*. In: Parallel problem solving from nature, Eiben, A.E., Bäck, T., Schoenauer, M. and Schwefel, H.-P., eds., 613-622, Springer.
- Hordijk, W. (1999). *Dynamics, Emergent Computation, and Evolution in Cellular Automata*. PhD thesis, University of New Mexico (USA).
- Huberman, B. A. and Adamic, L. A. (1999). Growth dynamics of the world-wide web. *Nature*, **401**, 131.
- The International Human Genome Sequencing Consortium (2001). *Initial sequencing and analysis of the human genome*. Nature **409**, 860 - 921
- Jackson, E.R., Johnson, D. and Nash, W.G. (1986). *Gene Networks in Development*. J. Theor. Biol. **119**, 379-396

- Jeong, H., Tombor, B., Albert, R., Oltvai, Z. N. and Barabási, A. L. (2000). *The large-scale organization of metabolic networks*. Nature, **407**, 651-654
- Jeong, H., Mason, S. P., Barabási, A. L. and Oltvai, Z. N. (2001). *Lethality and centrality in protein networks*. Nature, **411**, 41
- Kamp, C. and Bornholdt, S. (2002). *Co-evolution of quasispecies: B-cell mutation rates maximize viral error catastrophes*. Phys. Rev. Lett. **88**, 068104.
- Kaneko, K. (1994). *Relevance of dynamic clustering to biological networks*. Physica D. **75**, 55-73
- Kaneko, K. und Yomo, T. (1994). *Cell division, differentiation and dynamic clustering*. Physica D **75**, 89-102
- Kaneko, K. und Furusawa, C. (1999). *Robust and Irreversible Development in Cell Society as a General Consequence of Intra-Inter Dynamics*. E-print chao-dyn/9912005
- Kauffman, S.A. (1969). *Metabolic stability and epigenesis in randomly connected nets*. J. Theoret. Biol., **22**, 437-469
- Kauffman, S. (1995). *At Home in the Universe – The Search for the Laws of Self-Organization and Complexity*. Oxford University Press, New York.
- Kauffman, S.A. (1993). *The Origins of Order: Self-Organization and Selection in Evolution*. Oxford University Press
- Kenyon, C. (1995). *A perfect vulva every time: Gradients and signalling cascades in C. Elegans*. Cell, **82**, 171-174
- Kepler, T.B. and Elston, T.C. (2001). *Stochasticity in Transcriptional Regulation: Origins, Consequences and Mathematical Representations*. Santa Fe working paper, May 2001
- Kerszberg, M. and Wolpert, L. (1998). *Mechanisms for positional signalling by morphogen transport: a theoretical study*. J. Theor. Biol., **191**, 103-114
- Kleinberg, J., Kumar, S.R., Raghavan, P., Rajagopalan, S. and Tompkins, A. (1999). *The Web as a graph: Measurements, models and methods*. Proc. Intl. Conf. on Combinatorics and Computing, 1-18
- Klemm, K. and Bornholdt, S. (2003). *Robust gene regulation: Deterministic dynamics from asynchronous networks with delay*. Preprint q-bio/0309013

- Koch, A.J. and Meinhardt, H. (1994). *Biological pattern formation: from basic mechanisms to complex structures*. Rev. Mod. Phys. **66**, 1481-1510
- Kree, R. and Zippelius, A. (1987). Phys. Rev. A, **36**, 4421.
- Kürten, K.E., (1988a). *Critical phenomena in model neural networks*. Physics Letters A, 129(3), 157-160
- Kürten, K.E. (1988b). *Correspondence between neural threshold networks and Kauffman Boolean cellular automata*. J. Phys. A, 21, L615-L619
- Kumar, N.M and Gilula, N. (1996). *What is a gap junction?* Cell **84**, 381-388
- Kyoda, K. and Kitano, H. (1999). *Simulation of genetic interaction for Drosophila leg formation*. Pacific Symposium on Biocomputing **4**, 53-64
- Lander, A.D., Nie, Q., and Wan, F. (2002). *Do Morphogen Gradients arise by Diffusion?* Developmental Cell **2**, 785.
- Langton, C.G.(1986). *Studying Artificial Life with Cellular Automata*. Physica D, **22**, 120-149
- Langton, C.G.(1990). *Computation at the edge of chaos: Phase transitions and emergent computation*. Physica D, **42**, 12-37
- Langton, C.G.(1991). *Life at the edge of Chaos*. In: Langton, C.G., Taylor, C., Farmer, J.D. and Rasmussen, S. (eds.), Artificial Life II, 255-276, Addison-Wesley
- Lawrence, P.A. (1992). *The making of a Fly: The Genetics of Animal Design*. Blackwell, Scientific, Oxford
- Lewin, B. (1997). *Genes VI*. Oxford University Press.
- Lindgren, K. and Nordahl, M.G. (1990). *Universal Computation in Simple One-Dimensional Cellular Automata*. Complex Systems **4**, 299-318
- Luque, B. and Solé, R. (1996). *Phase transitions in random networks: simple analytic determination of critical points*. Phys. Rev. E **55** , 257-260
- Marciniak-Czochra, A. (2003). *Receptor-based models with diffusion-driven instability for pattern formation in Hydra*. J. Biol. Syst. **11**, 293.
- Marnellos, G. and Mjolsness, E. (1998). *A gene network approach to modeling early neurogenesis in Drosophila*. Pacific Symposium on Biocomputing **3**, 18-29

- Marnellos, G., Deblandre, G.A., Mjolsness, E. and Kintner, C. (2000) *Delta-Notch inhibitory patterning in the emergence of ciliated cells in Xenopus: experimental observations and a gene network model*. Pacific Symposium on Biocomputing **5**, 329-340
- Marrs, J.A. and Nelson, W.J. (1996). *Cadherin Cell Adhesion Molecules in Differentiation and Embryogenesis*. Int. Rev. Cyt. **165**, 159-205
- Martinez, D.E., Dirksen, M.L., Bode, P.M., Jamrich, M., Steele, R.E. and Bode, H.R. (1997). *Budhead, a fork head I HNF-3 homologue, is expressed during axis formation and head specification in Hydra*. Dev. Biol. **192**, 523-536
- Maynard Smith, J. and Szathmáry, E. (1995). *The major transitions in evolution*. Oxford University Press, reprinted 1999.
- McCulloch, W.S. and Pitts, W. (1943). *A Logical Calculus of Ideas Immanent in Nervous Activity*. Bull. Math. Biophys. **5**, 115.
- Meinhardt, H. (1978). *Space-dependent cell determination under the control of a morphogen gradient*. J. Theor. Biol. **74**, 307-321
- Meinhardt, H. (1982). *Models of Biological Pattern Formation*. Academic, New York, 1982
- Meinhardt, H. (1983). *Cell determination boundaries as organizing regions for secondary embryonic fields*. Dev. Biol. **96**, 375-385
- Meinhardt, H. (1988). *Models for maternally supplied positional information and the activation of segmentation genes in Drosophila embryogenesis*. Development **104** (Suppl.), 95-110
- Meinhardt, H. (1989). *Models for positional signalling with application to the dorsoventral patterning of insects and segregation into different cell types*. Development **109** (Suppl.), 169-180
- Meinhardt, H. (2000). *Organizer and axes formation as a self-organizing process*. Int. J. Dev. Biol. **45**, 177-188
- Meinhardt, H. and Gierer, A. (2000). *Pattern formation by local self-activation and lateral inhibition*. BioEssays **22**(8), 753-760

- Meinhardt, H. (1993). *A Model for Pattern Formation of Hypostome, Tentacles, and Foot in Hydra: How to Form Structures Close to Each Other, How to Form Them at a Distance*. *Dev. Biol.* **157**, 321-333
- Mendoza, L., Thieffry, D. and Alvarez-Buylla, E.R. (1999). *Genetic control of flower morphogenesis in Arabidopsis Thaliana: a logical analysis*. *Bioinformatics* **15**, 593-606
- Merks, R. (1997). *Evolving "Metazoan" Development*. Masters thesis, Utrecht University, 1997
- McCluskey, E.J. (1956). *Bell Syst. Techn. J.*, **35**, 1417-1444.
- Milgram, S. (1967). The small world problem. *Psych. Today*, **2**, 60-67.
- M. Mitchell, P. Hraber and J.P. Crutchfield (1994). *Evolving Cellular Automata to Perform Computations: Mechanisms and Impediments*. *Physica D*, 361-391.
- Mjolsness, E., Sharp, D.H. and Reintz, J. (1991). *A Connectionist Model of Development*. *J. Theor. Biol.* **152**, 429-456
- Mombach, J.C.M. (1999). *Simulation of embryonic cell self-organization: A study of aggregates with different concentrations of cell types*. *Phys. Rev. E* **59**, R3827-R3830
- Monk, N. A. M., Sherratt, J. A. and Owen, M. R. (2000). *Spatiotemporal patterning in models of juxtacrine intercellular signalling with feedback*. In: *Mathematical Models for Biological Pattern Formation*. (Ed. P. K. Maini and H. G. Othmer), 165-192, Springer.
- Monk, N. A. M. (2000). *Elegant hypothesis and inelegant fact in developmental biology*. *Endeavour* **24**, 170-173.
- Müller, W.A. (1993). *Pattern control in Hydra: basic experiments and concepts*. In: *Experimental and Theoretical Advances in Biological Pattern Formation*, eds. H.G. Othmer, P.K. Maini, and J.D. Murray, Plenum Press, New York.
- Nakamura, I. (2003). *Dynamics of Scale Free Random Threshold Networks*. cond-mat/0302399
- von Neumann, J. (1966). *Theory of Self-Reproducing Automata*. University of Illinois Press (Completed and edited by A.W. Burks)

- Neumann, C. and Cohen, S. (1997). *Morphogens and pattern formation*. *Bioessays* **19**, 721-729
- Newman, M. E. J., Strogatz, S. H., and Watts, D. J. (2001). Random graphs with arbitrary degree distributions and their applications. *Phys. Rev. E*, **64**, 026118.
- Odell, G.M., Oster, G., Alberch, P. and Burnside, B. (1981). *The Mechanical Basis of Morphogenesis. I. Epithelial Folding and Invagination*. *Dev. Biol.* **85**, 446-462
- Oliviero, S. and Struhl, K. (1991). *Synergistic Transcriptional Enhancement does not Depend on the Number of Acidic Activation Domains Bound to the Promoter*. *Proc. Natl. Acad. Sci.* **88**, 224-228.
- Ott, A., Soriano, J. and Colombo, C. (2004). *Symmetriebrechung in einem multizellularen System am Beispiel der Hydra*. *Verhandl. DPG VI 39*, 3/2004
- Owen, M.R., Sherratt, J.A. and Wearing, H.J. (2000). *Lateral induction by juxtacrine signaling is a new mechanism for pattern formation*. *Dev. Biol.* **217**, 54-61
- Pearson, H. (2001). *The regeneration gap*. *Nature*, **414** 388.
- Ransick, A. and Davidson, E.H. (1993). *A Complete Second Gut Induced by Transplanted Micromeres in the Sea Urchin Embryo*. *Science*, **259**, 1134.
- Reboul, J., Vaglio, P., Tzellas, N., Thierry-Mieg, N., Moore, T., Jackson, C., Shin-i, T., Kohara, Y., Thierry-Mieg, D., Thierry-Mieg, J., et al. (2001). *Open-reading-frame sequence tags (OSTs) support the existence of at least 17,300 genes in C. elegans*. *Nature Genetics* **27**, 332-336
- Redner, S. (1998). How popular is your paper? An empirical study of the citation distribution. *Eur. Phys. J. B*, **4**, 131-134.
- Reil, T. (1999). *Dynamics of Gene Expression in an Artificial Genome - Implications for Biological and Artificial Ontogeny*. *Proceedings of the 5th European Conference on Artificial Life*, 457-466, Springer, 1999
- Reil, T. (2000). *Models of Gene Regulation - A Review*. in: Maley, C.C. and Boudreau, E. (eds.), *Artificial Life 7 Workshop Proceedings*, 107-113, MIT Press, 2000
- Reinitz, J., Mjolsness, E. and Sharp, D.H. (1995). *Model for cooperative control of positional information in Drosophila by Bicoid and maternal Hunchback*. *J. Exp. Zool.* **271**, 47-56

- Rohlf, T. (2000). *Networks and Self-organized Criticality*. Masters thesis, Christian-Albrechts-Universität Kiel
- Rohlf, T. and Bornholdt, S. (2002). *Criticality in Random Threshold Networks: Annealed Approximation and Beyond*. *Physica A* **310**, 245-259
- Rohlf, T. and Bornholdt, S. (2003). *Self-organized pattern formation and noise-induced control from particle computation*. Preprint: cond-mat/0312366, submitted
- Rohlf, T. and Bornholdt, S. (2004). *Gene Regulatory Networks: A Discrete Model of Dynamics and Topological Evolution*. In: *Function and regulation of cellular systems - Experiments and Models*, Eds. Deutsch, A., Howard, H., Falcke, M. and Zimmermann, W., Birkhäuser Basel
- Rohlf, T. and Bornholdt, S. (2004). *Morphogenesis by coupled regulatory networks*. Preprint: q-bio.MN/0401024, submitted
- Ruoslahti, E. (1997). *Stretching is Good for a Cell*. *Science* **276**, 1345-1346
- Salazar-Ciudad, I., Garcia-Fernandez, J. and Solé, R.V. (2000). *Gene Networks Capable of Pattern Formation: From Induction to Reaction-diffusion*. *J. Theor. Biol.* **205**
- Salazar-Ciudad, I., Newman, S.A. and Solé, R.V. (2001). *Phenotypic and dynamical transitions in model genetic networks I. Emergence of patterns and genotype-phenotype relationships*. *Evolution & Development* **3**, 84-94
- Salazar-Ciudad, I., Solé, R.V. and Newman, S.A. (2001). *Phenotypic and dynamical transitions in model genetic networks II. Application to the evolution of segmentation mechanisms*. *Evolution & Development* **3**, 95-103
- Samuelson, B. and Troein, C. (2003). *Superpolynomial Growth in the number of Attractors in Kauffman Networks*. *Phys. Rev. Lett.* **90**, 098701
- Sasai, M. and Wolynes, G. (2003). *Stochastic gene expression as a many-body problem*. *Proc. Natl. Acad. Sci. USA*, **100**, 2374-2379
- Schiliro, Forman, B.J. and Javois, L.C. (1999). *Interactions between the foot and bud patterning systems in Hydra vulgaris*. *Dev. Biol.* **209**, 399-408
- Schreckenberg, M., Schadschneider, A., Nagel, K. and Ito, N. (1995). *Discrete stochastic models for traffic flow*. *Physical Review E*, **51**, 2939-2949
- Schuster, H. G. (2001). *Complex Adaptive Systems*. Scator, Saarbrücken.

- Sinha, S., Joshi, N.V., Rao, J.S. and Mookerjee, S. (1984). *A four-variable model for pattern-forming mechanism in Hydra*. *Biosystems* **17**, 15-22
- Simon, P.M and Nagel, K. (1998). *Simplified cellular automaton model for city traffic*. *Physical Review E* **58**, 1286-1295
- Slack, J. (1993). *Embryonic induction*. *Mech. Devel.* **41** , 257, 9100-9107.
- Snel, B., Bork, P. and Huynen, M.A. (2002). *The identification of functional modules from the genomic association of genes*. *Proc. Natl. Acad. Sci. USA*, **99**, 5890-5895
- Smith, A.R. (1971) *Simple Computation-Universal Cellular Spaces*. *J. Assoc. Comput. Mach.* **18**, 339-353.
- Socolar, J.E.S. and Kuuffman, S.A. (2003). *Scaling in Ordered and Critical Random Boolean Networks*. *Phys. Rev. Lett.* **90**, 068702
- Solé, R. and Luque, B. (1995). *Phase transitions and antichaos in generalized Kauffman networks*. *Phys. Lett. A* **196** , 331-334
- Sole, R.V., Salazar-Ciudad, I. and Garcia-Fernandez, J. (1999). *Phase Transitions in a Gene Network Model of Morphogenesis*. Santa Fe working paper, 1999
- Solé, R.V., Salazar-Ciudad, I. und Garcia-Fernandez, J. (2000). *Landscapes, Gene Networks and Pattern Formation: on the Cambrian Explosion*. *Adv. Complex Systems* **1**, 2-28
- Solé, R.V., Salazar-Ciudad, I. und Garcia-Fernandez, J. (2002). *Common pattern formation, modularity and phase transitions in a gene network model of morphogenesis..* *Physica A*, **305**, 640-654
- Steinberg, M.S. and Takeichi, M. (1994). *Experimental specification of cell sorting, tissue spreading, and specific spatial patterning by quantitative differences in cadherin expression*. *Proc. Natl. Acad. Sci. USA* **91**, 206-209
- Storch, V. and Welsch, U. (1996). *Kükenthals Leitfaden für das Zoologische Praktikum, 22. Auflage*. G. Fischer, Stuttgart.
- Strigini, M. and Cohen, S.M (2000). *Wingless gradient formation in the Drosophila wing*. *Curr. Biol.*, **10**, 293-300
- Strogatz, S. H. (2001). *Exploring complex networks*. *Nature*, **410**, 268-276

- Szathmary, E. (2001). *Developmental circuits rewired*. Nature **411**, 143-145
- Takagi, H., Kaneko, K. and Yomo, T. (2000). *Evolution of genetic code through isologous diversification of cellular states*. Artificial Life **6**, 283-305
- Technau, U., Cramer von Laue, C., Rentzsch, F., Luft, S., Hobmayer, B., Bode, H. R., and Holstein, T. W. (2000). Parameters of self-organization in hydra aggregates. *Proc. Natl. Acad. Sci. U.S.A.*, **97**(22), 12127–12131.
- Thieffry, D., Huerta, A., Perez-Rueda, E. und Collado-Vides J. (1998). *From specific gene regulation to genomic networks: a global analysis of transcriptional regulation in Escherichia Coli*. BioEssays, 20, 433-440
- Thieffry, D. and Thomas, R. (1998). *Qualitative Analysis of Gene Networks*. Pacific Symposium on Biocomputing 3, 77-88
- Thieffry, D. and Romero, D. (1999). *The Modularity of Biological Regulatory Networks*. BioSystems **50**, 49-59
- Thomas, R. (1978). *Logical analysis of systems comprising feedback loops*. J. Theor. Biol. **73**, 631-656
- Thomas, R. (1998). *Laws for the dynamics of regulatory networks*. Int. J. Dev. Biol. **42**, 479-485
- Thomsen, S. (2001). *Molekulare Analyse der positionsabhängigen Genaktivierung in Hydra*. Masters thesis, Kiel University, 2001
- Thomsen, S., Till, A., Beetz, C., Wittlieb, J., Khalturin, K. and Bosch, T.C.G. (2003). *Control of foot differentiation in Hydra: in vitro evidence that the NK-2 homeobox factor CnNK-2 autoregulates its own expression and uses pedibin as target gene*. *Mech. Devel.*, in press.
- Toffoli, T. (1984). *Cellular Automata as an alternative (rather than an approximation of) differential equations in modeling physics*. Physica D, **10**, 117-127
- Tracqui, P. (1995). *From passive diffusion to active cellular migration in mathematical models of tumor invasion*. Acta Biotheoretica **43**, 443-464
- Turing, A. (1952). *The chemical basis of morphogenesis*. Phil. Trans. Roy. Soc. Lond. B **237**, 37-72

- Umeda, T. and Inouye, K. (1999). *Theoretical model for morphogenesis and cell sorting in Dictyostelium discoideum*. Physica D **126**, 189-200
- Varea, C., Aragon, J.L. and Bario, R.A. (1997). *Confined Turing patterns in growing systems*. Phys. Rev. E **56**, 1250-1253
- Wagner, A. (1994). *Evolution of gene networks by gene duplications: A mathematical model; and its implications on genome organization*. Proc. Natl. Acad. Sci. USA, **91**, 4387-4391
- Watson, J. and Crick, F. (1953). *A structure for Deoxyribose Nucleic Acid*. Nature **171**, 737
- Watts, D. J. and Strogatz, S. H. (1998). Collective dynamics of “small-world” networks. *Nature*, **393**, 440-442.
- Wessels, L.F.A., van Someren, E.P. and Reinders, M.J.T. (2001). *A Comparison of Genetic Network Models*. Pacific Symposium on Biocomputing **6**, 508-519
- Weaver, D.C., Workman, C.T. and Stormo, G.D. (1999). *Modeling Regulatory Networks with Weight Matrices*. Pacific Symposium on Biocomputing **4**, 112-123
- West, D. B. (2001). *Introduction to Graph Theory*. Prentice Hall, London.
- Wolfram, S. (1983). *Statistical Mechanics of Cellular Automata*. Rev. Mod. Phys., **55**, 601.
- Wolfram, S. (1984a). *Universality and Complexity in Cellular Automata*. Physica D, **10**, 1 (1984);
- Wolfram, S. (1984b). *Cellular automata as models of complexity*. Nature (London), **311**, 419 (1984).
- Wolfram, S. (2002). *A new Kind of Science*. Wolfram Media.
- Wolpert, L. (1969). *Positional information and the spatial pattern of cellular differentiation*. J. Theor. Biol. **25**, 1-47
- Wright, A.P.H. and Gustafsson, J.A. (1991). *Mechanism of Synergistic Transcriptional Transactivation by the Human Glucocorticoid Receptor*. Proc. Natl. Acade. Sci. **88**, 8283-8287.
- Wuensche, A. (1998). *Genomic regulation modeled as a network with basins of attraction*. Pacific Symposium on Biocomputing **3**, 89-102

- Yuh, Chiou-Hwa, Bolouri, H. and Davidson, E.H. (1998). *Genomic Cis-Regulatory Logic: Experimental and Computational Analysis of a Sea Urchin Gene*. *Science* **279**, 1896-1902

Curriculum Vitae

Name: Thimo Rohlf
Geburtstag: 26. Oktober 1973
Geburtsort: Flensburg
Wohnort: Leipzig
Staatsangehörigkeit: deutsch

1980 – 1984 Grundschole Lütjenburg
1984 – 1993 Hoffmann-von-Fallersleben-Gymnasium Lütjenburg
06/1993 Abitur (Note: 1,0)

07/1993 – 09/1994 Zivildienst (Mobiler Sozialer Hilfsdienst in Ascheberg, Holstein)

ab 10/1994 Physik-Studium, Christian-Albrechts-Universität Kiel
10/1996 Vordiplom in Physik (Note: „sehr gut“)
ab 10/1996 Biologie-Studium, Christian-Albrechts-Universität Kiel
10/1998 Vordiplom in Biologie (Note: „sehr gut“)
05/1999 – 05/2000 Diplomarbeit bei Prof. Dr. Stefan Bornholdt am Institut für Theoretische Physik und Astrophysik der Christian-Albrechts-Universität zu Kiel über „Netzwerke und selbstorganisierte Kritizität“
07/2000 Diplom in Physik (Note: „mit Auszeichnung“)

08/2000 – 10/2000 Praktikum bei Siemens Medical Technology, Erlangen
11/2000 – 07/2001 Software-Entwickler bei PPI Financial Systems, Kiel

ab 04/2001 Doktorand bei Prof. Dr. Stefan Bornholdt am Institut für Theoretische Physik und Astrophysik der Christian-Albrechts-Universität zu Kiel über „Statistical Physics of Dynamical Networks and Morphogenesis“ (Doktoranden-Stipendium der Studienstiftung des deutschen Volkes)

06/2002 – 07/2002 Teilnahme an der Complex Systems Summer School des Santa Fe Institute, Santa Fe, New Mexico (USA)

ab 08/2002 Doktorand bei Prof. Dr. Stefan Bornholdt am Interdisziplinären Zentrum für Bioinformatik (IZBI) der Universität Leipzig

Erklärung

Ich erkläre, dass ich die vorgelegte Arbeit – abgesehen von der Beratung durch meinen Betreuer Prof. Dr. Stefan Bornholdt – selbstständig und nur unter Benutzung der angegebenen Literatur angefertigt habe.

Teile der Arbeit wurden in – im Literaturverzeichnis entsprechend aufgeführten – Fachartikeln veröffentlicht bzw. zur Veröffentlichung eingereicht. Die den Veröffentlichungen zugrunde liegende Forschungsarbeit wurde – abgesehen von der Beratung durch meinen wissenschaftlichen Lehrer – ausschließlich von mir selbst durchgeführt.

Außerdem versichere ich, dass die vorliegende Dissertation weder ganz noch zum Teil bei einer anderen Stelle im Rahmen eines Prüfungsverfahrens vorgelegt worden ist.

Frühere Promotionsversuche wurden von mir nicht unternommen.

Leipzig, den 16. März 2004