

Beiträge des Instituts für Umweltsystemforschung
der Universität Osnabrück

Herausgeber: Prof. Dr. Michael Matthies

Beitrag Nr. 22

Modellbildung und -anwendung
in den Wissenschaften IV

Horst Malchow (Hrsg.)

April 2001

**Beiträge des Instituts für Umweltsystemforschung
der Universität Osnabrück**

ISSN 1433-3805

Herausgeber

Prof. Dr. Michael Matthies
Universität Osnabrück
Institut für Umweltsystemforschung
Artilleriestr. 34

D-49069 Osnabrück

Tel. 0541/969-2575

Fax. 0541/969-2599

E-Mail: matthies@uos.de

<http://www.usf.uni-osnabrueck.de>

© USF – Institut für Umweltsystemforschung, Universität Osnabrück

**Dieser Beitrag enthält die schriftlichen Ausarbeitungen der Studierenden zum
Proseminar Angewandte Systemwissenschaft im Sommersemester 2000.**

Inhaltsverzeichnis

| | | |
|----------|--|-----------|
| 1 | Elmar Seestädt, Ole Bahlmann: Chaostheoretische Ansätze bei der Modellierung in den Wirtschaftswissenschaften | 7 |
| 1.1 | Einleitung | 7 |
| 1.1.1 | Kurzzusammenfassung | 7 |
| 1.1.2 | Abstract | 7 |
| 1.1.3 | Motivation | 7 |
| 1.2 | Klassische Modellbildung in den Wirtschaftswissenschaften | 8 |
| 1.2.1 | Charakterisierungskriterien für Modelle | 8 |
| 1.2.2 | Erläuterung der Modelltypen und Anwendungsbeispiele | 8 |
| 1.3 | Einführung in die Chaostheorie | 11 |
| 1.3.1 | Historischer Überblick | 11 |
| 1.3.2 | Begriffe | 12 |
| 1.3.3 | Ein erstes Modell | 20 |
| 1.4 | Einfluss der Chaostheorie auf die Wirtschaftswissenschaften | 21 |
| 1.4.1 | Klassische Theorie für Kapitalmärkte | 21 |
| 1.4.2 | Chaostheoretische Ansätze | 21 |
| 1.4.3 | Befunde für Kapitalmärkte | 23 |
| 1.4.4 | Ein nichtlineares Modell für den Optionsmarkt | 23 |
| 1.4.5 | Grenzen der Vorhersagbarkeit | 24 |
| 1.5 | Schlusswort | 26 |
| 2 | Jan Baudisch: Introduction to Evolutionary Algorithms | 29 |
| 2.1 | Introduction | 29 |
| 2.2 | Biological Background | 29 |
| 2.3 | Why Evolutionary Algorithms? | 30 |
| 2.4 | Genotype / Phenotype | 30 |
| 2.5 | The Genetic Operators | 30 |
| 2.5.1 | Selection | 30 |
| 2.5.2 | Recombination / Crossover | 31 |
| 2.5.3 | Mutation | 32 |
| 2.6 | The Algorithm | 32 |
| 2.7 | History | 33 |
| 2.8 | Types of Evolutionary Algorithms | 33 |
| 2.9 | Areas of Application | 33 |
| 2.10 | Example of a Genetic Algorithm | 34 |
| 2.11 | Variation of Evolutionary Algorithms | 34 |
| 2.12 | Valuation | 35 |

| | | |
|----------|---|-----------|
| 3 | Andreas Focks: Künstliches Leben im Computer | 37 |
| 3.1 | Einleitung | 37 |
| 3.1.1 | Leben | 37 |
| 3.1.2 | Künstliches Leben | 38 |
| 3.1.3 | Grundsätzliches zu Artificial Life | 38 |
| 3.2 | Ergebnisse aus der Erforschung von künstlichem Leben | 39 |
| 3.2.1 | Zellularautomaten | 40 |
| 3.2.2 | Genetische Algorithmen | 44 |
| 3.2.3 | Neuronale Netze | 46 |
| 3.3 | Zusammenfassung | 50 |
| 3.4 | Literatur | 50 |
| | | |
| 4 | Stephan Mankart: Simulation in den Sozialwissenschaften | 52 |
| 4.1 | Überblick | 52 |
| 4.2 | Methodische Einordnung | 52 |
| 4.3 | Ziele sozialwissenschaftlicher Modellierung | 53 |
| 4.3.1 | Qualitatives Verhalten bei willkürlich gewählten Anfangsbedingungen | 53 |
| 4.3.2 | Qualitative Vorhersagen bei realitätsnah gewählten Anfangsbedingungen | 53 |
| 4.3.3 | Quantitative Vorhersagen | 54 |
| 4.3.4 | Modellzweck: Erklären vs. Vorhersagen | 54 |
| 4.4 | Eine Besonderheit sozialwissenschaftlicher Modellierung- Artificial Societies | 55 |
| 4.5 | Probleme | 55 |
| 4.6 | Geschichtlicher Überblick | 55 |
| 4.7 | Simulationsansätze | 56 |
| 4.8 | System dynamics | 56 |
| 4.9 | Mikroanalytische Simulation | 57 |
| 4.10 | Queuing models | 57 |
| 4.11 | Mehrebenenmodelle | 58 |
| 4.12 | Zelluläre Automaten | 59 |
| 4.13 | Multi-Agenten-Modelle | 59 |
| 4.13.1 | Angestrebte Eigenschaften von Agenten | 60 |
| 4.14 | Ausblick | 61 |
| 4.15 | Literatur | 61 |
| 4.15.1 | Verwendete Literatur | 61 |
| 4.15.2 | Angeführte Literatur | 62 |
| | | |
| 5 | Matthias Peußner: Synergetics | 63 |
| 5.1 | Introduction | 63 |
| 5.2 | One Example of Synergetics | 63 |
| 5.3 | Macroscopic structures | 63 |
| 5.4 | Thermodynamic | 67 |
| 5.4.1 | The Second law | 68 |
| 5.4.2 | Contradiction? | 68 |
| 5.5 | Synergetic Systems | 69 |
| 5.5.1 | Definition of (new) words | 69 |
| 5.5.2 | The concept | 70 |

| | | |
|----------|---|------------|
| 5.5.3 | The slaving principle | 71 |
| 5.5.4 | Conclusion | 72 |
| 5.6 | Add-on: Chaotic behavior | 72 |
| 6 | Andreas Manz: A model of the laser | 74 |
| 6.1 | Basic structure of the laser | 74 |
| 6.2 | About light | 75 |
| 6.2.1 | Light in general | 75 |
| 6.2.2 | Light in the resonator | 75 |
| 6.3 | Processes in the laser | 76 |
| 6.3.1 | Loss at the mirror | 76 |
| 6.3.2 | Absorption | 78 |
| 6.3.3 | Spontaneous emission | 78 |
| 6.3.4 | Stimulated emission | 78 |
| 6.4 | How does the amplitude behave? | 79 |
| 6.4.1 | The laser condition | 79 |
| 6.4.2 | What limits the growth of the amplitude? | 79 |
| 6.4.3 | Stability of the amplitude | 80 |
| 7 | Tobias Niemann: Droplet impacts - more than just a few nice pictures | 83 |
| 7.1 | Droplet impacts - what is that? | 83 |
| 7.1.1 | droplet collision with surface | 83 |
| 7.1.2 | droplet collision with droplet | 83 |
| 7.2 | Why research on it? | 84 |
| 7.2.1 | reference to reality | 84 |
| 7.2.2 | theory works in atom and star modeling | 84 |
| 7.3 | Experimental approach | 85 |
| 7.4 | Theoretical approach | 85 |
| 7.4.1 | classical mechanics | 85 |
| 7.4.2 | numerical modeling | 88 |
| 7.5 | reality vs modelling | 91 |
| 7.6 | Conclusions | 93 |
| 8 | Ba Kien Tran: Die Entstehung des Lebens | 95 |
| 8.1 | Was ist Leben? | 95 |
| 8.2 | Hypothesen zur Lebensentstehung | 96 |
| 8.3 | Die „Urerde“ | 97 |
| 8.3.1 | Die Uratmosphäre | 98 |
| 8.3.2 | Die Energiequellen der „Urerde“ | 98 |
| 8.4 | Simulationsversuche zur präbiotischen Stoffsynthese | 98 |
| 8.5 | RNA-Welt | 99 |
| 8.6 | Der Hyperzyklus | 100 |
| 9 | Alexander Meyer: Die Erde lebt - James Lovelocks Gaia-Theorie | 104 |
| 9.1 | Die Gaia - Theorie | 104 |
| 9.1.1 | Was ist Gaia ? | 104 |
| 9.1.2 | Begründer | 105 |
| 9.1.3 | Entstehung | 105 |
| 9.1.4 | Leben | 106 |

| | | |
|-----------|--|------------|
| 9.1.5 | Kritik | 107 |
| 9.2 | Das Daisyworld - Modell | 107 |
| 9.2.1 | Modellannahmen | 107 |
| 9.2.2 | Modell mit zwei Arten | 108 |
| 9.2.3 | Modelle mit mehreren Arten | 109 |
| 9.2.4 | Stabilität des Daisyworlds - Modells | 110 |
| 9.2.5 | Gesamtergebnis | 112 |
| 9.3 | Die Epigenese aus der Sicht der Gaia - Theorie | 113 |
| 9.3.1 | Archaikum | 113 |
| 9.3.2 | Proterozoikum | 116 |
| 9.3.3 | Neuzeit | 117 |
| 9.4 | Physiologie aus Sicht der Gaia - Theorie | 119 |
| 9.4.1 | Schwefel-Kreislauf | 119 |
| 9.4.2 | Kalzium | 119 |
| 9.4.3 | Salz | 120 |
| 9.5 | Der Mensch und Gaia | 121 |
| 9.5.1 | Heutige Probleme | 121 |
| 9.6 | Literaturverzeichnis | 122 |
| 10 | Christian Fiedler: Einführung in die Thermodynamik | 123 |
| 10.1 | Die Klassische Thermodynamik | 123 |
| 10.1.1 | Einleitung | 123 |
| 10.1.2 | Der nullte Hauptsatz | 125 |
| 10.1.3 | Der 1. Hauptsatz | 128 |
| 10.1.4 | Der 2. Hauptsatz | 128 |
| 10.1.5 | Die Gibbsche Fundamentalform | 129 |
| 10.1.6 | Der 3. Hauptsatz | 130 |
| 10.2 | Der Stirlingmotor als Beispiel für eine Wärmekraftmaschine | 130 |
| 10.2.1 | Einleitung | 130 |
| 10.2.2 | Theoretische Grundlagen | 131 |
| 10.3 | Literatur | 133 |
| 11 | Harold Fellermann: Introduction to Self-Organized Criticality | 134 |
| 11.1 | Properties of Complex Systems | 134 |
| 11.2 | The Theory of Self-Organized Criticality | 136 |
| 11.2.1 | catastrophes and $1/f$ -Noise | 136 |
| 11.2.2 | The Critical State | 138 |
| 11.2.3 | Self-Organized Criticality | 139 |
| 11.2.4 | Spatial fractals and SOC | 139 |
| 11.3 | Computer Models of SOC | 140 |
| 11.3.1 | The Canonical Sandpile Model | 140 |
| 11.3.2 | A Model of River Networks | 142 |
| 11.3.3 | Other SOC Computer Models | 144 |
| 11.4 | Further implications of SOC | 145 |
| 11.4.1 | SOC and the Edge of Chaos | 145 |
| 11.4.2 | Possibilities to predict | 145 |

Kapitel 1

Chaostheoretische Ansätze bei der Modellierung in den Wirtschaftswissenschaften

Elmar Seestädt, Ole Bahlmann

12. April 2000

1.1 Einleitung

1.1.1 Kurzzusammenfassung

In diesem Artikel werden wir einen Überblick über die chaostheoretische Modellierung in den Wirtschaftswissenschaften geben. Gefolgt von einer kurzen Einführung in die grundlegenden Begriffe und Methoden der Chaosforschung werden wir zuerst einen Einblick in die klassische Modellierung in den Wirtschaftswissenschaften geben.

Nach den Einführungen in die Grundbegriffe wird ein chaostheoretisch geprägtes Modell für den Optionsmarkt vorgestellt werden. Anhand dieses Modells zeigen wir auch die Unterschiede zu den klassischen Modellen und Hypothesen in den Wirtschaftswissenschaften.

1.1.2 Abstract

In this article we are going to give an overview of the chaos theoretical modelling in the economic sciences. Followed by a short introduction on the basic terms and methods of chaos research we will give an insight into the classic modelling in the economic sciences. After the introduction into the basic terms a chaos theoretic modell of the option market will be introduced. On the basis of this modell we will illustrate the differences to the classic modells and hypothesis in the economic sciences.

1.1.3 Motivation

Wir haben uns im Hinblick auf das Anwendungsfach Betriebswirtschaftslehre für die Modellbildung in den Wirtschaftswissenschaften interessiert. Die Wahl viel auf die chaostheoretischen Ansätze, da dies ein sehr neuer Zweig der Modellierung in den Wirtschaftswissenschaften ist und uns interessiert hat, wie weit dieser Zweig schon gediehen ist. Wir

wollten sehen, was tatsächlich schon möglich ist und wie die Anwendungsreife dieser neuen Ideen für die Modellierung und Analyse von wirtschaftlichen Phänomenen ist.

1.2 Klassische Modellbildung in den Wirtschaftswissenschaften

Die Modellbildung in den Wirtschaftswissenschaften ist aus dem Bedarf entstanden Entscheidungskriterien zu bekommen. Klassisch werden Modelle in den Wirtschaftswissenschaften mittels linearer Zusammenhänge beschrieben. Zu ertragsgesetzlichen Funktionen, die mit Differentialgleichungen beschrieben werden können, sagt Rieper [20, S. 85]: "Formal existieren solche Funktionen Solche Überlegungen bleiben aber formale Übungen ohne praktische Aussagekraft, wenn es keine Anwendungssituation gibt, auf welche die ertragsgesetzlichen Verläufe zutreffen."

Die Aussage von Rieper ist ein klassisches Beispiel dafür, dass auch heute noch lineare Modelle in den Wirtschaftswissenschaften dominieren.

1.2.1 Charakterisierungskriterien für Modelle

Nach Adam [1, S. 81 ff] kann man Modelle in der Betriebswirtschaft nach unterschiedlichen Kriterien einteilen. Die Einteilung ist weder eine Überschneidungsfreie noch Vollständige, sondern stellt die in der Literatur am häufigsten anzutreffenden Einteilungen übersichtlich dar. Adam wählt folgende Einteilung:

- Nach der Art der verarbeiteten Information: qualitative und quantitative Modelle.
- Je nachdem, ob eine Strategie gefunden oder überprüft werden soll: explizite und implizite Modelle
- Nach der Darstellungsform: mathematische, graphische und physikalische Modelle
- Nach dem Einsatzzweck: Entscheidungs- und Erklärungsmodelle
- Nach der Art der Abstraktion:
 - zeitabhängige (dynamische) und zeitunabhängige (statische) Modelle
 - offene und geschlossene
 - und nach der sachlichen bzw. zeitlichen Ausdehnung des Entscheidungsfeldes Partial- und Totalmodelle unterscheiden.

1.2.2 Erläuterung der Modelltypen und Anwendungsbeispiele

Quantitative und qualitative Modelle

Quantitative Modelle beschreiben mit mathematischen Gleichungen oder Ungleichungen strukturelle Zusammenhänge der Zustandsgrößen. Sie haben den Vorteil, dass sie mit geeigneten mathematischen Verfahren bearbeitet und gelöst werden können.

Zur Klasse der quantitativen Modelle gehören Modelle aus der Bestell- oder Absatzpolitik.

Qualitativen Modellen fehlt die Quantifizierbarkeit der kennzeichnenden Grössen. Es wird mit generellen Tendenzen und Richtungen gearbeitet, ohne ein bestimmtes Mass anzugeben.

Solche Modelle spielen eine Rolle in der strategischen Planung, z.B. bei der Positionierung von neuen Produkten am Markt. Ausserdem eignen sie sich zum Ranking zwischen verschiedenen Entscheidungsmöglichkeiten.

Modelle mit implizit und explizit erfassten Strategien

Explizite Modelle ermöglichen es für eine bestimmte Zielsetzung eine optimale Handlungsstrategie zu finden, da sie implizit alle Variablen und Restriktionen für das Modell in den Gleichungen enthalten. Sie haben den Nachteil, dass sie für die Aufstellung der Gleichungen ein hohes Abstraktionsvermögen des Modellierers bedürfen. Dafür gewinnt man aber auch den Vorteil eine vielleicht noch nicht in Betracht gezogene Handlungsstrategie zu erfahren.

Mit einem solchen Modell kann man zum Beispiel ein Bestellsystem modellieren, um herauszufinden, wann Bestellungen in welcher Höhe erforderlich sind.

Implizite Modelle erlauben nur eine Überprüfung von Strategien, mit Hilfe des Modells. Aus ihnen kann man keine neuen Handlungsstrategien ableiten. Sie haben den Vorteil, dass sie relativ leicht zu modellieren sind und Strategien auf Ihre Plausibilität überprüfen können.

Mit solchen Modell kann man zum Beispiel Investitionsstrategien gegeneinander abwägen. Den Zielwert des Modells kann man für unterschiedliche Strategien direkt vergleichen.

In der Praxis empfiehlt es sich beide Methoden miteinander zu kombinieren. D.h. man vergleicht die Ergebnisse der beiden Modelle miteinander, um so Aussagen über die Güte des Modells zu erhalten.

Mathematische, graphische und physikalische Modelle

Mathematische Modelle In der Regel sind die meisten Modelle in den Wirtschaftswissenschaften mathematischer Natur. Sie haben den klaren Vorteil, dass Sie exakt sind und auf wohldefinierten Axiomen basieren. Daher kann man auf sie das gesamte Instrumentarium anwenden, das von der Mathematik zur Verfügung gestellt wird.

Graphische Modelle dienen meist nur der Veranschaulichung von Vorgängen und erleichtern die Einführung von mathematischen Modellen. Sie haben den Nachteil, dass sie nur auf Probleme im zwei- und dreidimensionalen Raum anwendbar sind.

Physikalische Modelle sind äusserst selten in den Wirtschaftswissenschaften anzufinden, da es schwierig ist wirtschaftswissenschaftliche Zusammenhänge in einem physikalischen Modell darzustellen.

Ein Beispiel für ein physikalisches Modell ist das Varignonsche Modell [11, S. 193] zur Wahl eines transportkostenoptimalen Standortes.

Hierzu wird eine Landkarte, die das Zielgebiet markiert auf eine (Tisch-)Platte gelegt und fixiert. Dann bohrt man in die Zielstandorte Löcher und lässt durch jedes der Löcher

einen Faden laufen. Die gleichlangen Fäden werden oberhalb der Platte zusammengeknotet. Nun befestigt man an jedem Faden ein Gewicht, das die Gütermenge, die zu dem entsprechenden Standort transportiert werden soll repräsentiert. Leichtgängigkeit in den Löchern vorausgesetzt bewegt sich der Knoten zu dem Standort mit der geringsten Tonnenkilometerleistung.

Erklärungs- und Entscheidungsmodelle

Erklärungsmodelle versuchen einen Sachverhalt zu Erklären d.h. ihn anschaulich darzustellen. Erst einmal ist natürlich jedes Modell ein Erklärungsmodell.

Entscheidungsmodelle unterscheiden sich von Erklärungsmodellen dadurch, dass aus ihnen Schlussfolgerungen für Entscheidungen gezogen werden sollen. Das Varignonsche Modell ist somit ein Entscheidungsmodell.

Unterschieden werden die beiden Modelltypen nur nach ihrem Modellzweck. Für einen Betrachter kann ein Modell also Erklärungsmodell, für einen anderen Betrachter aber auch zugleich Entscheidungsmodell sein.

Statische und dynamische Modelle

Statische Modelle sind Modelle, in denen der Zeitablauf nicht abgebildet wird. D.h. sie enthalten keine Größen, die sich im Laufe der Zeit ändern können. Da sich reale Systeme aber immer mit der Zeit verändern ist dies ein erheblicher Abstraktionsschritt. Solche Modelle können prinzipiell immer nur einen statischen Gleichgewichtszustand widerspiegeln und sollten nur Anwendung finden, wenn sich das System sehr schnell auf diesen Gleichgewichtszustand einpegelt. Sie sind nicht angebracht, wenn es einer längeren Oszillationsphase bedarf bis sich der Systemzustand einpendelt. Statische Modelle sind daher auch unabhängig von der Dauer der Planungsperiode.

Dynamische Modelle spiegeln im Gegensatz zu den statischen Modellen auch den zeitlichen Verlauf des Systems wider.

Diskrete dynamische Modelle setzen Zeitreihen oder Zeitfunktionen zueinander in Beziehung. Dies unterscheidet sie von den

kontinuierlichen dynamischen Modellen, die einen Zeitverlauf kontinuierlich mit Hilfe von Differentialgleichungen modellieren. Hier ist es mit gegebenen Anfangsbedingungen prinzipiell immer möglich ein Ergebnis zu finden.

Ein wesentlicher Unterscheidungspunkt zu den statischen Modellen ist auch, dass die dynamischen Modelle von der Länge des Betrachtungszeitraums abhängig unterschiedliche Ergebnisse liefern können.

Offene und geschlossene Modelle

Offene Modelle sind Modelle, die von externem, unerklärtem Input abhängig sind. D.h. sie beruhen auf gegebenen Zeitreihen. Ein Modell, das Nachfrageverhalten bei vorgegebenen Preisen untersucht ist ein Offenes, da die Preise von aussen in das System eingespeist werden.

Geschlossene Modelle modellieren den kompletten Systemzusammenhang und verzichten auf externe Daten. Verändert man das oben genannte Beispiel dahingehend, dass man den Preis von der Nachfrage abhängig macht, erhält man ein geschlossenes System.

Partial- und Totalmodelle

Partialmodelle betrachten nur einen Teil der Unternehmung. Das betrachtete Subsystem wird unabhängig von Ergebnissen aus anderen Funktionsbereichen modelliert und untersucht. Historisch hat die Entwicklung der wirtschaftswissenschaftlichen Modellierung ihren Ursprung bei den Partialmodellen. Es wurden Lagerhaltungsmodelle, preispolitische Modelle oder Modelle der klassischen Investitionstheorie betrachtet.

Totalmodelle betrachten die Unternehmung als ganzes. Im extremsten Fall werden hier sämtliche Teilbereiche und Aspekte einer Unternehmung bis zum Ende Ihrer Laufzeit modelliert.

Im Rahmen der grösser gewordenen Leistungsfähigkeit der Rechnersysteme und aufgrund geänderter Bedürfnisse, in Bezug auf die Aussagekraft der Modelle sind Totalmodelle immer mehr in den Vordergrund gerückt.

1.3 Einführung in die Chaostheorie

1.3.1 Historischer Überblick

Noch im 19. Jahrhundert waren Wissenschaftler der Ansicht, dass das Universum und seine Entwicklung vollkommen deterministisch und vorhersagbar sei [15, S. 1]. Geprägt wurde diese Vorstellung von Pierre-Simon de Laplace (1749-1827), der der Meinung war, wenn man sich nur genug Mühe geben würde und genug Geduld aufbrächte die Differentialgleichungen, die das Universum definierten, aufzuschreiben, könnte man die Zukunft mit absoluter Genauigkeit voraussagen.

Der französische Mathematiker Henri Poincaré (1854-1912) wurde durch die Feststellung, dass auch kleine Ursachen grosse Wirkungen haben können zum eigentlichen Gründungs-vater der modernen Chaosforschung. Ausserdem stellte er die Fundamente des deterministischen Weltbildes in Frage, indem er, wie später auch Werner Heisenberg, eine völlig exakte Ermittlung aller Ausgangsvariablen und schlussfolgernd eine Berechnung der Zukunft ausschloss. Noch in den Denkstrukturen seiner Zeit verfangen schaffte Poincaré es nicht seine Überlegungen konsequent bis zum Ende weiterzudenken: "Diese Dinge sind so bizarr, dass ich es nicht aushalte, weiter darüber nachzudenken." [8, S. 22].

1914 zerstörte Albert Einstein mit seiner Relativitätstheorie die klassische physikalische Denkweise und Werner Heisenberg brachte 1927 mit der Heisenbergschen Unschärferelation einen weiteren Beitrag zu neuen Denkstrukturen.

Der Begriff der Chaostheorie selbst wurde 1960 von dem amerikanischen Meteorologen Edward N. Lorenz, der an der Simulation eines globalen Wettermodells arbeitete, geprägt. Dazu, wie Lorenz dem Chaos auf die Spur gekommen ist, gibt es verschiedene Ansichten in der Literatur.

Nach Deser [8, S. 24] hat Lorenz bei der Berechnung seines Wettermodells anstatt von neuem zu rechnen eine Berechnung mit alten, gerundeten Zwischenergebnissen gespeist, um Zeit zu sparen. Bereits nach wenigen Wetterperioden war jedoch jede Ähnlichkeit zu dem vorher berechneten Ergebnis verschwunden, was eine völlige Wertlosigkeit seiner

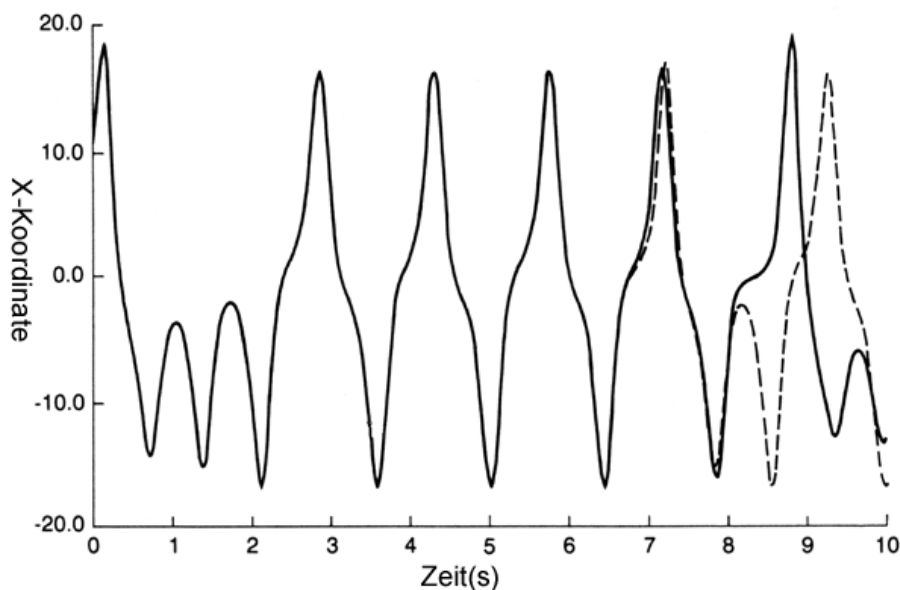


Abbildung 1.1: Diagramm zum Systemverhalten von Lorenz's Wettermodell. Hier ein Plot der X-Koordinate mit den Eingangsvariablen: normale Linie ($X=10$ $Y=10$ $Z=10$), gestrichelte Linie ($X=10,01$ $Y=10$ $Z=10$). Quelle: [[8]].

Berechnungen impliziert hätte. Vergleiche dazu das Diagramm (1.1, S.12). Es ist deutlich zu erkennen, dass das Systemverhalten bei marginaler Änderung der Parameter komplett unterschiedliche Resultate liefert.

Nach Bösen [5] hat Lorenz eine Berechnung einmal mit einer Genauigkeit von sechs und ein zweites mal mit einer von drei Dezimalstellen vollzogen, wobei er zu vollkommen unterschiedlichen Ergebnissen gekommen ist.

Im Kern haben beide Thesen aber die gleiche Aussagekraft. Lorenz hat als erster die Sensibilität von dynamischen Systemen auf die Anfangsbedingungen festgestellt.

1.3.2 Begriffe

Chaos

Der Chaosbegriff ist kein definiertes Gebilde, sondern vor allem eine Umschreibung für ein Phänomen. Mit Chaos ist nicht im klassischen Sinn die Unordnung gemeint, sondern ein komplexes Gefüge von Phänomenen. Das Wörterbuch der deutschen Sprache [26] sagt zu dem Begriff Chaos:

Chaos das; -[lat. chaos < griech cháos = der unendliche leere Raum; die gestaltlose Urmasse (des Weltalls)](bildungsspr.): *Auflösung aller Ordnung, völlige Verwirrung, völliges Durcheinander:*

Knaurs Lexikon von A-Z[16] schreibt:

Chaos s.[gr.] wüstes Durcheinander

Diese Definition, in der Umgangssprache mit dem richtigen Sinn behaftet, spiegeln den Chaosbegriff in seiner wissenschaftlichen Auffassung jedoch nicht, oder nur unzureichend wieder.

Chaotische Eigenschaften können nur nicht-lineare Systeme von Differentialgleichungen bzw. einzelne Differentialgleichungen höherer Ordnung aufweisen. Als nicht-linear werden solche Differentialgleichungen bezeichnet, in denen die Zustandsgrößen in einer Potenz ungleich eins oder ein Produkt aus Zustandsgrößen auftritt [15, S. 3]. Nicht jedes nicht-lineare Modell ist jedoch chaotisch. Wie in 1.3.3 beschrieben kann es sogar vorkommen, dass man nur einen Faktor in seiner Grösse verändert und das Modell keine chaotischen Verhaltensweisen mehr zeigt. Nach Grauwe [15, S. 34] kann Chaos als scheinbar zufälliges und ungleichmässiges Signal, das noch weitere Eigenschaften hat und von einem deterministischen Prozess generiert wird aufgefasst werden. Zu diesen weiteren Eigenschaften zählen:

- Sensitivität gegenüber den Eingangsvariablen
- Das Signal muss einen seltsamen Attraktor haben, d.h. das Phasendiagramm muss eine fraktale Dimension haben.
- ein gleichmässiges Fourier Spektrum
- Mindestens einen positiven Lyapunov Exponenten
- Sie sind aperiodisch, d.h. kein Zustand wird zweimal durchlaufen

Deterministisches Chaos ist ein scheinbar widersprüchlicher Begriff. Seine Bedeutung ist jedoch sehr einleuchtend. Es ist festzustellen, dass sich Systeme nach, oftmals bekannten, deterministischen Regeln verhalten. Diese Systeme sind jedoch äusserst empfindlich, was die Genauigkeit der Eingangsparameter angeht. Eine nur marginale Änderung der Parameter ruft eine faktisch vollkommen unterschiedliche Reaktion hervor. Aufgrund dieser Sensitivität ist es unmöglich über chaotische Systeme beliebig gültige Prognosen abzugeben. Das System verhält sich zwar nach bestimmten Regeln ist jedoch nicht vorhersagbar und wirkt chaotisch. Dieser Tatsache wird mit dem Begriff deterministisches Chaos Rechnung getragen.

Dass kleine Ursachen grosse Wirkungen haben wurde schon vor Poincaré in einem amerikanischen Volkslied besungen[8, S. 28]:

Weil ein Nagel fehlte, ging das Hufeisen verloren, weil ein Hufeisen fehlte, ging das Pferd verloren, weil ein Pferd fehlte, ging der Reiter verloren, weil ein Reiter fehlte, ging die Schlacht verloren, weil die Schlacht verloren war, ging auch das Königreich verloren.

Attraktor

Attraktoren und Attraktionsmengen charakterisieren das Verhalten eines dynamischen Systems im Zeitablauf. Wie Bereits die Bezeichnung impliziert, handelt es sich bei Attraktionsmengen um Bereiche, bei Attraktoren um exakt bestimmbare Punkte, denen sich ein dynamisches System entgegenstrebt [9, S. 42]. Attraktoren werden meist mit einem Diagramm visualisiert, in dem die Zustandsvariablen des Systems gegeneinander im Phasenraum aufgetragen werden. Trotz der schon erwähnten grossen Sensitivität von chaotischen Systemen bleibt das qualitative Erscheinungsbild des Attraktors immer gleich. Der wohl bekannteste aller Attraktoren ist der vielzitierte Lorenzattraktor (1.2, S.14), der bei dem

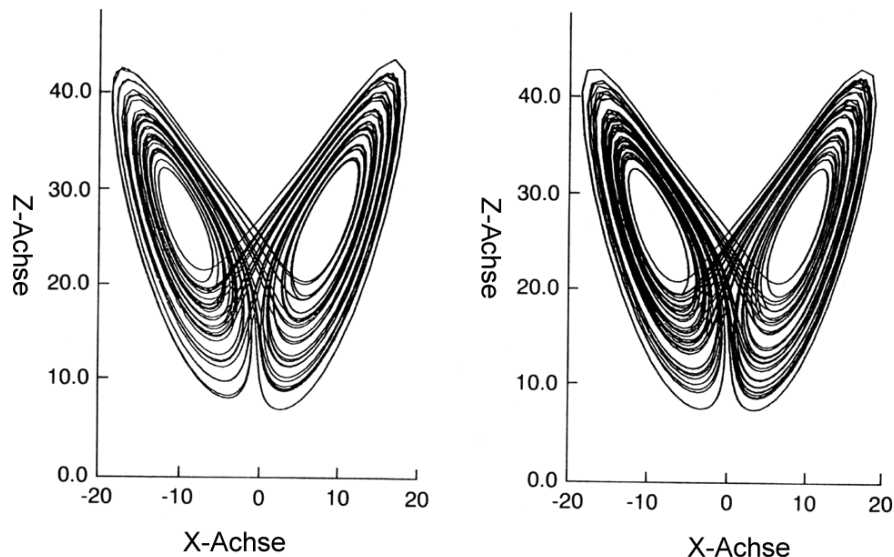


Abbildung 1.2: Lorenzattraktor. Links mit den Eingangsvariablen ($X=10$ $Y=10$ $Z=10$), rechts mit den Variablen ($X=10,01$ $Y=10$ $Z=10$). Quelle: [[8]].

Wettermodell von Lorenz 1.3.3 auftritt. An den beiden Bildern ist auch sehr gut zu erkennen, dass die Qualität gleich bleibt, die einzelnen Variablen in Abb. 1.1 und Abb. 1.11 betrachtet jedoch keine Ähnlichkeit mehr aufweisen.

Trajektorie bezeichnet die Bahn des Systemverlaufs, die um den Attraktor, oder die Attraktionsmenge beschrieben wird.

Bifurkation

Angenommen ein dynamisches System ist von einem Parameter λ abhängig. Bifurkation ist eine qualitative Änderung des Systemverhaltens, dass auftritt, wenn der Parameter λ einen bestimmten kritischen Punkt passiert, den sogenannten Bifurkationspunkt. Der Parameter λ wird Bifurkationsparameter genannt. [23, S. 25]

Bifurkationsdiagramm Mit Hilfe eines Bifurkationsdiagramms wird die qualitative Aufspaltung des Systems dargestellt.

Feigenbaum Konstante 1975 ist der Physiker Mitchell Feigenbaum bei dem Längenvergleich zweier aufeinanderfolgender Äste auf eine konstante Zahl gestossen. Die Länge eine Astes bis zu seiner nächsten Verdoppelung sei mit Δ_n bezeichnet, wobei n die Periodenlänge ist. Die Länge des nächsten Astes ist Δ_{2n} . Feigenbaum stiess bei Annäherung an den kritischen Wert r_{crit} auf die Konstante

$$\begin{aligned} \delta &= \lim_{n \rightarrow \infty} \frac{\Delta_n}{\Delta_{2n}} \\ &= 4,67 \end{aligned}$$

, die nach Feigenbaum benannt wurde. Siehe dazu auch die Grafik 1.4 auf Seite 15. Später stellte Feigenbaum fest, dass δ auch in anderen Systemem mit Periodenverdoppelung

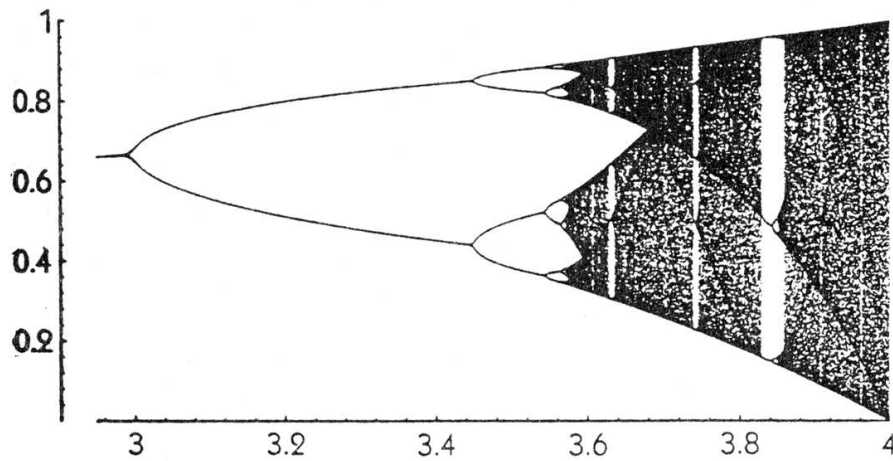


Abbildung 1.3: Bifurkationsdiagramm von $x_{t+1} = r(1 - x_t)x_t$. Quelle: [[17, Fronzek - Von der Ordnung des Chaos, die Feigenbaumsche Universalität, S. 10]].

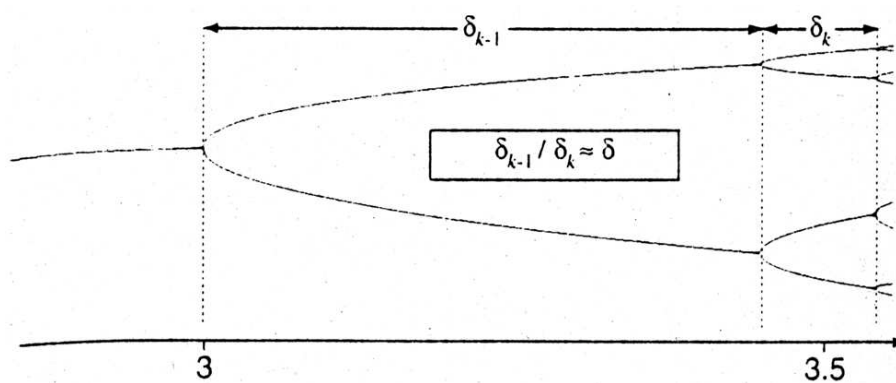


Abbildung 1.4: Das Verhältnis zweier aufeinanderfolgender Äste strebt gegen die Feigenbaumkonstante δ . Quelle: [[17, Fronzek - Von der Ordnung des Chaos, die Feigenbaumsche Universalität, S. 10]].

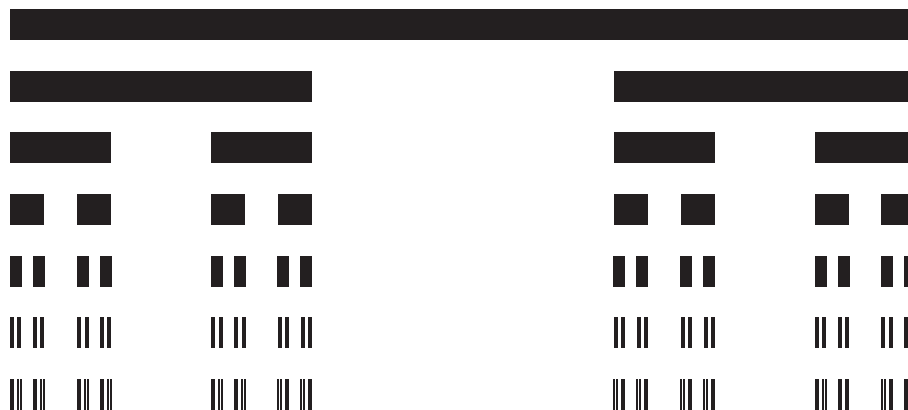


Abbildung 1.5: Die Cantor-Menge erhält man, wenn man ausgehend von einer Linie, Schritt für Schritt jeweils das mittlere Drittel der resultierenden Linien herausschneidet. Quelle: [Selbst erstellte Illustration].

auftritt. man kann behaupten, dass die Feigenbaumkonstante δ eine ähnliche Bedeutung für die Chaostheorie erlangt hat, wie π für die Geometrie. Bei der Untersuchung eines Bifurkationsdiagramms kann man mit Hilfe der Zahl δ feststellen, ab wann chaotisches Verhalten auftreten kann.

Fraktale Dimension

Entstehung Der Begriff der fraktalen Dimension stammt von dem berühmten Mathematiker Benoît Mandelbrot. Mandelbrot entdeckte eine vorher noch nicht wahrgenommene Ordnung in so unterschiedlichen Phänomenen wie der Einkommensverteilung, Preisschwankungen, Übertragungsfehlern in Telefonleitungen bis hin zur Struktur des Weltalls. Mandelbrot stieß auf die fraktale Geometrie, als er 1960 zu einem Vortrag an der Harvard Universität eingeladen wurde. Als er den Vortragssaal betrat stellte er zu seiner Überraschung fest, dass der ihn einladende Professor bereits einen Chart aus seinem Vortrag an der Tafel hatte. Als er jedoch bemerkte, dass der vermeintlich von ihm stammende Chart nicht die Einkommensverteilung sondern die Baumwollpreise widerspiegelte wurde Mandelbrots Neugier geweckt.

Bis dato ging man von einer Unabhängigkeit von kurzfristigen und langfristigen Kurschwankungen aus. Mandelbrot fand jedoch mit Hilfe der Computertechnologie eine symmetrische Ordnung bei der Massstabsänderung.

Später beschäftigte Mandelbrot sich im Rahmen seiner Anstellung bei IBM mit scheinbar zufälligen Störungen bei der telefonischen Übermittlung.

Die bisher mit dem Thema beschäftigten IBM-Ingenieure verfolgten die Problematik nicht weiter, da sie an der Betrachtung scheiterten. Die Fehler traten gehäuft auf, doch je genauer sie betrachtet wurden, desto feiner schien ihre Struktur. Mandelbrot entdeckte jedoch die Struktur, welche nach ihrem Entdecker, Georg Cantor [8, S. 34], Cantor-Menge (1.5, S.16) benannt wird.

Mandelbrot veröffentlichte seine Entdeckungen 1982 zusammenfassend in seinem Werk "The fractal geometry of Nature".

Definition Für die fraktale Dimension gibt es verschiedene Begrifflichkeiten, die auch jeweils anders berechnet werden. Zu den gängigen fraktalen Dimensionen zählen die Haus-

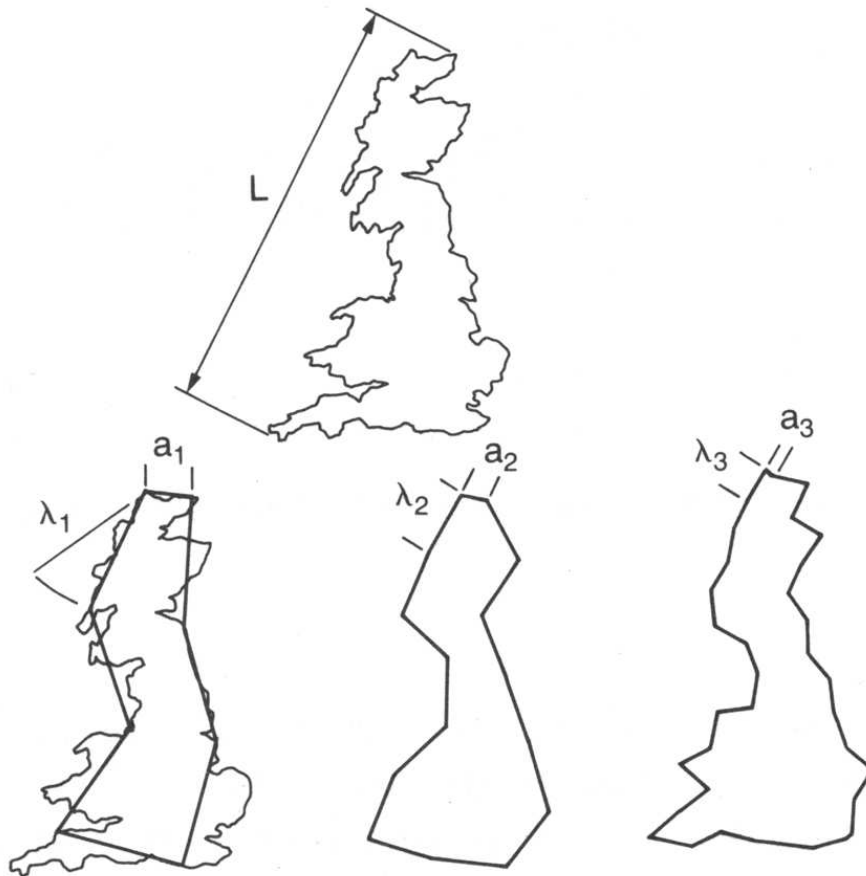


Abbildung 1.6: Finden der Länge der britischen Küstelinie. Es wird ein „Zollstock“ an die Küstelinie angelegt und diese Zahl mit der Länge des „Zollstocks“ multipliziert. Quelle: [15].

dorff Dimension, Boxdimension, Zirkeldimension, intrinsische Dimension und die Kapazitätsdimension. Die Dimensionen werden nicht nur unterschiedlich berechnet, sie liefern auch jeweils etwas andere Ergebnisse, es ist daher von Fall zu Fall abzuwägen, welche Dimension man wählt. Allen gemein ist, dass sie nichts mit der üblichen topologischen Dimension gemein haben, die nur ganze Zahlen annehmen kann. Für die fraktale Dimension gilt, dass sie zwischen der topologischen Dimension der Sache und der nächst niedrigeren topologischen Dimension liegt. D.h. für ein dreidimensionales Gebilde liegt sie zwischen zwei und drei. Für eine Punktmenge, wie der Cantormenge zwischen null und eins, da die topologische Dimension eines Punktes eins ist.

Beispiel Den Fraktalbegriff erläuterte Mandelbrot an der Frage: „Wie lang ist die Küstelinie Grossbritanniens?“. Die Antwort hängt von der Wahl des Betrachtungsmaßstabs ab. Je kleiner man den „Zollstock“ wählt, desto länger wird die Küstelinie (1.6, S. 17). Mandelbrot sagt dazu [8, S. 35]:

“Die Länge einer Küstelinie erweist sich als ein undefinierbarer Begriff, der einem durch die Finger gleitet, wenn man ihn fassen will. Alle Messmethoden führen letztendlich zu dem Schluss, dass die Länge einer typischen Küstelinie sehr gross und so schlecht bestimmt ist, dass sie am besten als unendlich angesehen wird.”

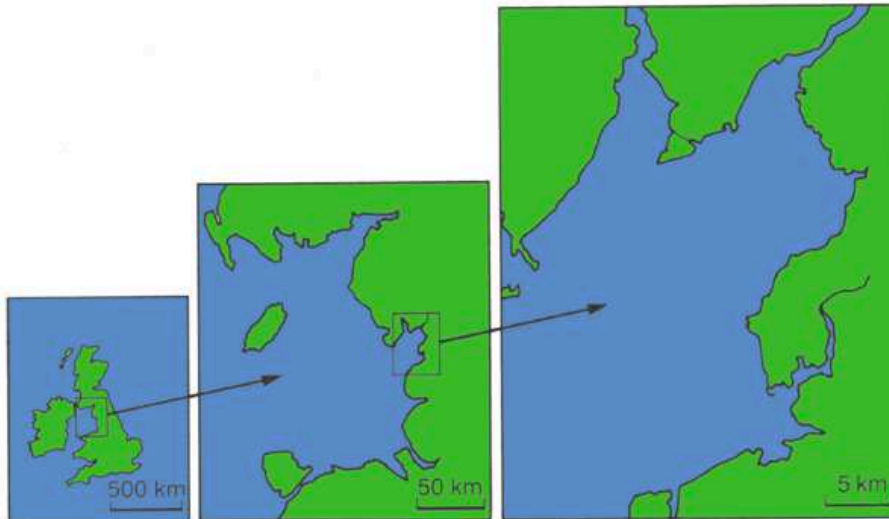


Abbildung 1.7: Selbstähnlichkeit am Beispiel der Küstenlinie Großbritanniens. Quelle: [15].

Selbstähnlichkeit ist ein Begriff, der auch von Mandelbrot geprägt worden ist. Fraktale Strukturen haben die Eigenschaft, dass sie sich quasi selbst reproduzieren. Das bedeutet, vergrößert man den Beobachtungsmaßstab kann man die gleichen Strukturen in sehr ähnlicher, aber nicht in gleicher Weise wiedererkennen. Mandelbrot führt hier als Beispiel wieder die Küstenstruktur Gross Britanniens an (1.7, S.18) Ein weiteres klassisches Beispiel für die Selbstähnlichkeit fraktaler Mengen ist die nach Mandelbrot benannte Mandelbrotmenge. Die Formel zur Berechnung der Mandelbrotmenge ist schon seit langem bekannt hat aber erst durch die graphische Aufarbeitung (1.8, S.19) Berühmtheit erlangt.

Einige fraktale Dimensionen Anhand der Tabelle 1.1 auf Seite 18 lassen sich einige fraktale Dimension von real existierenden Gebilden sowie geometrischen Strukturen und anderen Phänomenen ablesen.

| Phänomen | Abbildung | Quelle | Dimensionsart | fraktale Dimension |
|---------------------|-------------|-------------|---------------|--------------------|
| Cantor Menge | 1.5, S. 16 | [15, S. 28] | Hausdorff | 0,63093 |
| Menschliches Gehirn | keine | [8, S. 37] | keine Angabe | 2,79 |
| Sierpinski Dreieck | 1.10, S. 19 | [15, S. 19] | Hausdorff | 1,58496 |

Tabelle 1.1: Beispiele für fraktale Dimensionen

Lyapunov Exponent

Der Lyapunov Exponent ist nach dem russischen Mathematiker A.M. Lyapunov benannt und ein Maß für das Chaos bzw. die Stabilität in einem dynamischen Prozess. Er soll die Veränderung des Abstands in Abhängigkeit der Iterationen messen.

Es handelt sich um eine reelle Zahl λ , die eine exponentielle Konvergenz ($\lambda < 0$), Neutralität ($\lambda = 0$) oder Divergenz ($\lambda > 0$) zweier eng benachbarter Trajektorien eines

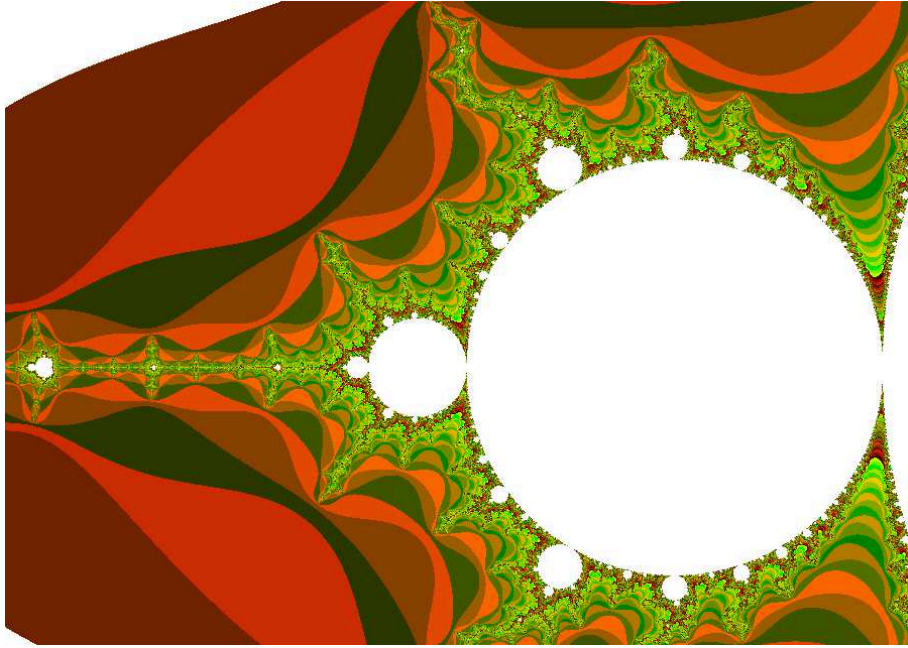


Abbildung 1.8: Mandelbrotmenge. Quelle:[Erstellt mit Fractals Generator 1.3 von Uwe Thiem (uwe@uwix.alt.na)].

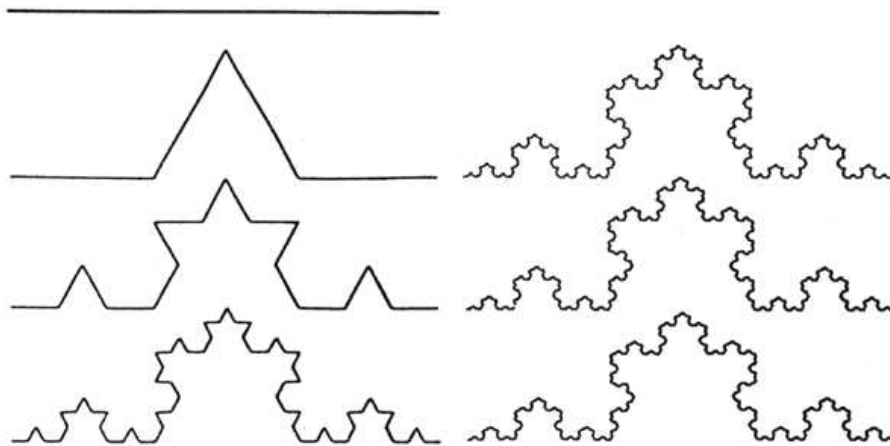


Abbildung 1.9: Koch Kurve. Die Koch Kurve wird konstruiert, indem man ausgehend von einer geraden Linie, das mittlere Drittel “herausschneidet” und durch zwei ebenso lange Stücke ersetzt. In jedem weiteren Schritt wird mit den verbleibenden geraden Teilstücken ebenso verfahren. Quelle:[Selbst erstellte Illustration].

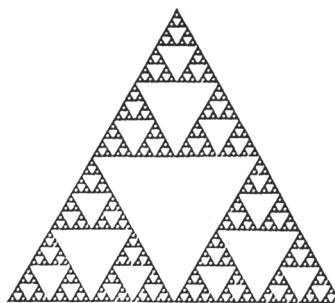


Abbildung 1.10: Sierpinski Dreieck. Quelle:[8].

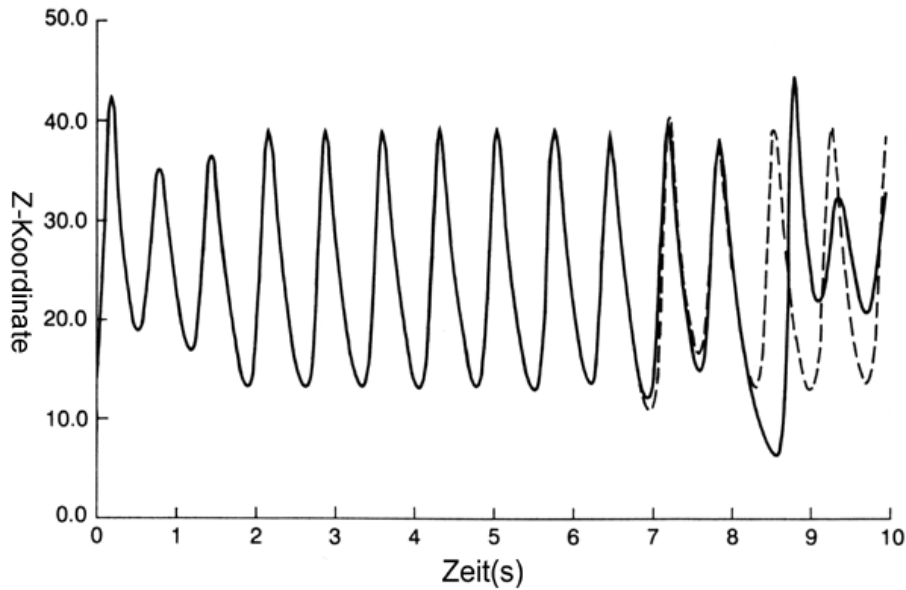


Abbildung 1.11: Diagramm zum Systemverhalten von Lorenz's Wettermodell. Hier ein Plot der Z-Koordinate mit den Eingangsvariablen: normale Linie ($X=10$ $Y=10$ $Z=10$), gestrichelte Linie ($X=10,01$ $Y=10$ $Z=10$). Quelle: [[8]].

dynamischen Systems im zeitlichen Mittel beschreibt.

$$\lambda_{x_0} = \lim_{n \rightarrow \infty} \frac{1}{n} \left(\ln \frac{d f^n(x_0)}{d x} \right)$$

1.3.3 Ein erstes Modell

Wettermodell von Lorenz

Das Wettermodell von Lorenz ist, auch wenn es nicht als solches geplant war, als das erste chaostheoretische Modell anzusehen. Lorenz versuchte mit Hilfe von drei Differentialgleichungen Aussagen über das zukünftige Verhalten des Wetters zu machen.

$$\frac{d X}{d t} = 10(Y - X) \tag{1.1}$$

$$\frac{d Y}{d t} = 28X - Y - XZ \tag{1.2}$$

$$\frac{d Z}{d t} = XY - \frac{8}{3}Z \tag{1.3}$$

Die Variablen X, Y, Z haben keine Relation zu Temperatur, Luftdruck oder Feuchtigkeit, es kann aber angenommen werden, dass das so aufgestellte Modell hinreichend realitätsbezogen ist.

Wie in der Graphik 1.11 auf Seite 20 zu sehen ist, reagiert das System sehr stark auf infinitesimale Änderungen der Eingangswerte. Das bedeutet, dass es sich, wie schon gesagt um ein chaotisches System handelt. Ersetzt man den Faktor 28 in der 2. Gleichung durch den Faktor 12 ist das System nicht mehr sensitiv den Eingangswerten gegenüber und wird daher nicht mehr als chaotisches System bezeichnet [15, S. 7]. Bereits nach sieben Sekunden beginnt das Systemverhalten auseinander zu driften und nach acht Sekunden ist keine Gemeinsamkeit in der Prognose mehr zu erkennen.

1.4 Einfluss der Chaostheorie auf die Wirtschaftswissenschaften

1.4.1 Klassische Theorie für Kapitalmärkte

Theorie der effizienten Märkte

Nach der Theorie der effizienten Märkte, auch mit starker Informationseffizienz bezeichnet, spiegeln die jeweils aktuellen Marktpreise immer alle verfügbaren Informationen wider, die diesen Markt betreffen.

Schwach effiziente Märkte zeichnen sich dadurch aus, dass es Insider gibt, die private Informationen nutzen können und somit der Marktpreis nicht die verfügbaren Informationen enthält.

Informationseffizienz im strengen Sinne liegt erst vor, wenn der Marktpreis sowohl alle öffentlichen als auch die privaten Informationen reflektiert.

Random-walk Hypothese

Die Random-walk Hypothese leitet sich aus der Theorie der effizienten Märkte ab. Unter der Annahme, dass der Markt immer alle Informationen birgt, können Preisänderungen nur durch neue Informationen bewirkt werden. Neue Informationen, sind jedoch vorher unbekannt und daher vom Zufall bestimmt. Daher muss auch die Preisentwicklung zufallsbestimmt sein. Die Preisentwicklung folgt demnach einem Zufallspfad (random-walk).

1.4.2 Chaostheoretische Ansätze

Allgemeine Vorgehensweise

Nach heutigen Ansätzen geht man davon aus, dass Märkte nicht zufallsbestimmt sind, sondern sich deterministisch, also chaotische Verhaltensweisen zeigen. Streng genommen kann man jedoch nicht beweisen, dass sich Märkte chaotisch verhalten, man kann nur Indizien finden, bzw. den Zufall ausschliessen.

Dies ist auch die von Uhlig [25, S. 25] gewählte Vorgehensweise. Es gilt die Hypothese:

“ Die Marktdaten lassen sich als Folge eines Zufallsprozesses erklären.”

zu widerlegen.

Dies geschieht, indem realistische Marktdaten mit Zufallsfolgen verglichen werden, die die gleichen statistischen Kennzeichen (Varianz, Mittel, etc.) aufweisen.

Nachweismethoden für deterministisches Verhalten

Im folgenden werden verschiedene Verfahren vorgestellt, mit denen man statistische Daten auf ihre zufallsbestimmtheit untersuchen kann.

Lineare Statistik

Lineare Korrelation Mit Hilfe der Linearen Korrelation kann berechnet werden, ob zwei Messreihen voneinander abhängig sind. Die Abhängigkeit muss natürlich noch auf ihren Kausalität überprueft werden. So sind ZUsammenhänge wie: “Wenn viele Störche fliegen, werden auch viele Kinder geboren.” eventuell statistisch mit der Korrelation zu begründen, ein realer Zusammenhang besteht allerdings nicht.

Fourier Analyse Ein wichtiges Verfahren der linearen Statistik ist die schnelle Fourier Transformation, auch FFT (**F**ast **F**ourier **T**ransformation) genannt.

Bei der Fourier Transformation wird der Umstand genutzt, dass sich Kurven durch Überlagerung von Sinus- und Cosinusschwingungen darstellen lassen. Dies gilt auch für sehr komplizierte Kurven, die keine Regelmässigkeit mehr erkennen lassen. Das Phasenspektrum dieser Analyse lässt sich dann entsprechend interpretieren.

Wavelet Analyse Die Wavelet Analyse ist wie die Fourier Transformation ein lineares Analyseverfahren. Bei der Bildverarbeitung ist sie der Fourier Transformation jedoch überlegen. Sie soll im Bereich hochfrequenter Schwingungen besonders gut Regelmässigkeiten erkennbar machen.

Im Gegensatz zur Fourier Transformation ist die Wavelet Analyse jedoch kein eindeutig bestimmtes Verfahren. Vielmehr muss man eine sogenannte Mutterfunktion wählen.

Nicht-lineare Statistik

Chi-Statistik Die Chie-Statistik erlaubt es beliebige Abhängigkeiten zu messen. Weiterhin ist es möglich mit ihrer Hilfe Signifikante Assoziationen zu bestimmen.

Entropieberechnung Die Entropie ist ein Mass für die allgemeine Unsicherheit eines Systems.

Das bedeutet die Entropie ist umso kleiner, je unregelmässiger die Wahrscheinlichkeiten auf die Zustände verteilt sind. Ebenso wie bei der Chi-Statistik kann man beliebige, also auch nicht-lineare Zusammenhänge herausfinden.

Transinformation oder bedingte Entropie Die Transinformaion gibt Auskunft darüber, ob die Einführung einer zusätzlichen Variable die Unsicherheit für eine Prognose senkt und wenn ja, wie stark diese Auswirkung ist.

Zweidimensionale Kolmogroff Smirnoff Statistik Die (KS)-Statistik erlaubt, im Gegensatz zu der eindimensionalen Variante den Vergleich zweidimensionaler Verteilungen.

Uhlig verwendet sie, um die Wochenänderung eines Marktes bzw. einer Zufallsreihe mit den Änderungen in ein, zwei, drei oder vier Wochen später zu vergleichen. Die späteren Werte fliessen dann jeweils in die zweite Dimension ein.

Test auf kurzfristige Abhängigkeit Diese Methode misst die Wahrscheinlichkeit, mit der aufeinanderfolgende Werte einer Zeitreihe einander ähnlich sind.

Nach der Auffassung von Robert Savit und Matthew Green ist dies der Fall für deterministische Systeme. Bei zufällig bestimmten Systemen ist die Dauer der Abhängigkeit kleiner. Als Massstab für die Änderung wird die Standardabweichung der Zeitreihe genutzt.

1.4.3 Befunde für Kapitalmärkte

Nach Uhlig [25, S. 91] lassen sich folgende Befunde feststellen:

- Die Marktdynamik ist nicht linear ¹
- Die Erträge der Märkte sind nicht normalverteilt ²
- Die Marktdynamik ist nicht zufallsbestimmt ³

Auch andere Autoren sind für verschiedene Märkte zu dem eindeutigen Befund gekommen, dass sie nicht zufallsbestimmt sind.

Ein erstes Fazit

Da die Märkte offensichtlich nicht zufallsbestimmt sind kann die Chaostheorie behilflich sein verlässlichere und damit bessere Prognosen, als sie mit der klassischen Statistik möglich sind, zu ermöglichen.

1.4.4 Ein nichtlineares Modell für den Optionsmarkt

Die Anwendung eines nichtlinearen Modells auf einen Markt beruht immer auf mindestens zwei Annahmen.

Zum einen, dass der modellierte Markt kein stochastisches sondern ein deterministisches, also nichtzufälliges, chaotisches, Verhalten zeigt.

Zum anderen, dass es sich um einen heterogenen Markt handelt. Das heißt, dass der Markt durch verschiedene Eigenschaften der Teilnehmer geprägt ist. Dazu zählen unter anderem eine Ungleichverteilung von Informationen unter den Teilnehmern, resultierend aus unterschiedlich guter Kapitalausstattung. Dies ist auch der Grund für eine unterschiedliche Risikobereitschaft bei den Teilnehmern. Dazu kommt noch die Unterscheidung zwischen langfristig und kurzfristig orientierten Anlegern [25, S. 135].

Modellbeschreibung

Ausgehend von diesen Annahmen haben *Michael de la Maza* und *Deniz Yuret* vom MIT ⁴ ein adaptives Marktmodell vorgestellt, das mit Hilfe von heterogen beschaffenen Agenten und genetischen Algorithmen den amerikanischen Index-Optionenmarkt simuliert.

An ihrem Modell überprüften sie die Einflüsse auf die Gewinnaussichten. Getestet wurden

1. Verrauschte Signale
2. Zwangsliquidationen
3. Unterschiedlicher Informationsstand der Teilnehmer
4. Veröffentlichungen von Erfolgsstrategien

¹Anhand einer Korrelationsrechnung gezeigt

²Anhand der (KS)-Statistik gezeigt

³Getestet mit Chi-Test, Entropieberechnung, bedingter Wahrscheinlichkeit und Widerkehrabbildung

⁴Massachusetts Institute of Technology, Cambridge, U.S.A.

Die verrauschten Signale hatten dabei die geringsten Auswirkungen auf die Teilnehmer. Sie wurden also besonders gut kompensiert. Zwangsliquidationen aufgrund geringer Liquidität und Unterschiede in den Informationsständen führten immer bei den Benachteiligten zu Verlusten und umgekehrt bei den nicht Betroffenen bzw. besser informierten zu deutlichen Gewinnsteigerungen. Wurden die Erfolgsstrategien allen Teilnehmern zugänglich gemacht, so sanken die Gewinne schnell gegen Null [25, S. 134].

Ergebnis dieses Ansatzes

Michael de la Maza und *Deniz Yuret* halten es für möglich auf heterogenen Märkten, wie den, den sie in ihrem Modell simuliert haben, über lange Zeit kontinuierlich Gewinne zu erwirtschaften.

Zumindest haben Sie mit den Erfahrungen aus ihren Marktmodellen 1993 an einer Investmentmeisterschaft für Optionen in den U.S.A. teilgenommen und mit einem Gewinn von 44% den fünften Platz belegt. Ein beachtlicher Erfolg, wenn man bedenkt, dass am Optionsmarkt in der Regel 90% der Teilnehmer Geld verlieren.

1.4.5 Grenzen der Vorhersagbarkeit

Kommen wir noch mal zurück zur Datenanalyse und dem Versuch mit Mitteln aus der Chaosforschung Voraussagen über das Verhalten deterministischer Systeme wie dem Aktienmarkt zu machen.

Dazu muss zuerst das komplexe System durch ein Modell mit möglichst wenig Dimensionen abgebildet werden. Dann muss mit dem Modell die Dynamik des Systems rekonstruiert und ein Attraktor gefunden werden. Dies ist ein kompliziertes Unterfangen, das sehr stark abhängig von den vorliegenden Daten ist. *Uhlir* führt dies in seinem Buch [25] an einem Beispiel vor und erläutert die Probleme. Das Vorhandensein eines Attraktors und damit die Möglichkeit zur Berechnung von lokalen Lyapunov-Exponenten ist Voraussetzung für die Vorhersagbarkeit eines Systems.

Befindet sich das System in der Nähe eines Attraktors und ist der lokale Lyapunov-Exponent an diesem Punkt klein, so ist auch die Empfindlichkeit von den Anfangsbedingungen gering. Mit steigendem Lyapunov-Exponenten steigt auch die Empfindlichkeit von den Anfangsbedingungen und damit wird der Zeitraum, in dem eine verlässliche Prognose möglich ist, kürzer.

Unterschiede in den lokalen Lyapunov-Exponenten eines Systems kennzeichnen sogenannte Vorhersagbarkeitsfenster. Dabei handelt es sich um Zustandsräume in denen genauere Prognosen möglich sind.

Um also seine Erfolgsaussichten in einem Markt zu verbessern ist es notwendig die Vorhersagbarkeitsfenster aufzuspüren. Das ist aber aus zwei Gründen durchaus nicht trivial. Zum einen können die Fenster durchaus kurz sein und zudem selten auftreten. Zum Anderen ist die Berechnung der lokalen Lyapunov-Exponenten im Vergleich zum globalen Lyapunov-Exponenten, der ja einen Durchschnittswert über einen vergleichsweise langen Zeitraum darstellt, deutlich schwieriger. *Uhlir* hat versucht die Lage der Vorhersagbarkeitsfenster zu bestimmen, indem er die Lage des Systems auf dem Attraktor (siehe Abbildung 1.12) mit den Prognosefehlern an den jeweiligen Positionen (siehe Abbildung 1.13) in einer Grafik aufgetragen hat.

Es fällt auf, dass die beiden Grafiken annähernd komplementär sind. Anhand dieses Feh-

Quartalsänderungen am deutschen Aktienmarkt

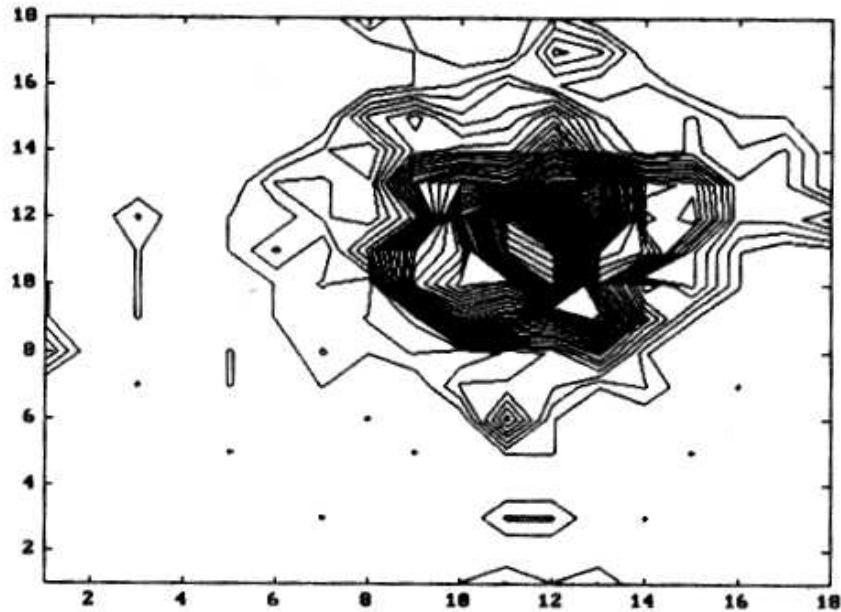


Abbildung 1.12: Das Bild zeigt die Häufigkeitsverteilung des Attraktormodells. Die relative Häufigkeit des Aufenthalts ist als Konturgrafik aufgezeichnet. Zunehmende Dichte der Linien entspricht zunehmender Häufigkeit des Aufenthalts. Quelle: [25, S. 181]

Prognosefehler für Quartalsänderungen am deutschen Aktienmarkt

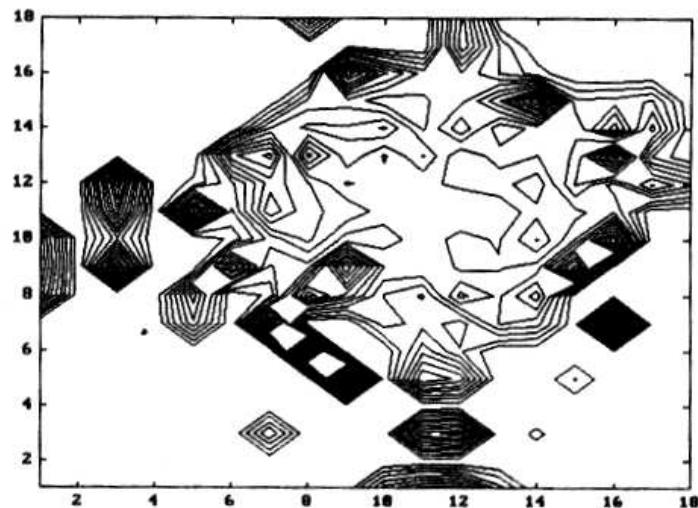


Abbildung 1.13: Hier ist die Größe des Prognosefehlers für den gleichen Bereich wie eben aufgezeichnet. Zunehmende Dichte der Linien entspricht zunehmendem durchschnittlichem Fehler. Quelle: [25, S. 181]

lerprofils für den deutschen Aktienmarkt, in dem die Kursänderungen im Quartalsabstand dargestellt sind, kann man erkennen, dass die Fehler dort häufiger sind, wo sich das System seltener aufhält.

Anzumerken ist noch, dass die Form des Fehlerprofils natürlich auch mit Resultat des verwendeten Optimierungsverfahrens ist. *Uhlig* benutzt hier ein neuronales Netz. Das hat den Vorteil dass häufig vorkommende Zustände besonders gut prognostiziert werden, während bei seltenen Zuständen größere Fehler in Kauf genommen werden.

Weitere Einschränkungen für die Vorhersagbarkeit

Es gibt noch weitere Gründe, die eine Vorhersage bei Systemen im allgemeinen und bei Märkten im speziellen, einschränken.

Zum einen darf man nicht vergessen, dass die Teilnehmer an den Märkten Menschen sind, die unter Umständen, bewusst oder unbewusst, nichtrational handeln.

Zum anderen sind Märkte auch stark von politischen Ereignissen und Umweltkatastrophen beeinflusst. Solche Geschehnisse sind zweifelsohne im Rahmen eines Marktmodells nicht vorhersagbar und schränken deshalb die Möglichkeit zur zuverlässigen Voraussage eines Marktverhaltens stark ein. Man sollte hier aber noch *Bossel* [6] beachten, der in einem solchen Fall der Ansicht ist, dass Systemverhalten sei zwar nicht genau vorhersagbar, bewege sich aber in jedem Fall in angebbaren Grenzen.

Ob diese Grenzen dann aber so gesteckt sind, dass trotzdem ein Vorteil aus der Voraussage zu ziehen ist, ist aber stark anzuzweifeln.

1.5 Schlusswort

Die Literaturliste ist nicht nur als Quellenangabe für die entsprechenden Textverweise zu verstehen, sondern auch als Liste mit weiterführender Literatur für den interessierten Leser. Alle Bücher aus der Literaturliste haben wir im im Laufe der Vorbereitungen bearbeitet. Viele Ideen und Gedanken entstammen aus ihnen, auch wenn sie nicht direkt als Verweis im Text auftauchen. Für uns waren die Bücher [25, 8, 5, 15] besonders interessant.

Literaturverzeichnis

- [1] Adam Dietrich, (1996) 4. Auflage. *Planung und Entscheidung. Modelle - Ziele - Methoden. Mit Fallstudien und Lösungen*, Gabler, Wiesbaden.
- [2] Ahmad Syed, (1991). *Capital in economic theory: neo-classical, Cambridge and chaos*, Edward Elgar Publishing, Hants.
- [3] Beckenbach Frank, Diefenbacher Hans, (1994). *Zwischen Entropie und Selbstorganisation*, Metropolis-Verlag, Marburg.
- [4] Benhabib Jess, (1991). *Cycles and chaos in economic equilibrium*, Princeton University Press, Princeton.
- [5] Bösen Martin, (1995). *Investitions-Controlling unter Berücksichtigung von Selbstorganisation und Chaos in der Unternehmung und Umwelt. Entwicklung eines prozessorientierten Controlling-Modells*, Phillips-Universität zu Marburg, Marburg.
- [6] Bossel Hartmut, (1994) 2. Auflage. *Modellbildung und Simulation. Konzepte, Verfahren und Modelle zum Verhalten dynamischer Systeme*, Vieweg, Braunschweig.
- [7] Brock W.A., Malliaris A.G., (1989). *Differential Equations Stability and Chaos in Dynamic Economics*, North-Holland, Amsterdam.
- [8] Deser Frank, (1996). *Chaos und Ordnung im Unternehmen. Chaosforschung als ein Beitrag zum besseren Verständnis von Unternehmen als Nichtlinearen Dynamischen Systemen*, Physica Verlag, Heidelberg.
- [9] Elsner Jürgen, (1996). *Chaos und Zufall am deutschen Aktienmarkt: eine Studie über nicht lineare Dynamiken, Volatilitäten und Effizienz*, Physica Verlag, Heidelberg.
- [10] Finkenstädt Bärbel, (1995). *Nonlinear dynamics in economics: a theoretical and statistical approach to agricultural markets*, Springer, Berlin.
- [11] Francis R.L., White J.A., (1974). *Facility layout and location. An analytical approach.*, Engelwood Cliffs, N.J.
- [12] Freiling C., (1989). *Lexikon der Rechnungslegung und Abschlussprüfung*, W. Lück, Marburg.
- [13] Gabisch Günter, Lorenz Hans-Walter, (1989) 2nd Edition. *Business Cycle Theory. A Survey of Methods and Concepts*, Springer Verlag, Berlin.
- [14] Gandolfo Giancarlo, (1996) 3rd Edition. *Economic Dynamics*, Springer Verlag, Berlin.

- [15] Grauwe Paul de, Dewachter Hans, Embrechts Mark, (1993). *Exchange rate theory: chaotic models of foreign exchange markets*, Blackwell Publishers, Oxford
- [16] Zöckler Paul Dr., (1951). *Knaurs Lexikon von A-Z*, Droemersch Verlag, München.
- [17] Malchow Horst, (1998). *Modellbildung und Anwendung in den Wissenschaften*, Institut für Umweltsystemforschung Universität Osnabrück, Osnabrück.
- [18] Perrings Charles, (1990). *Economy and Environment. A Theoretical Essay on the Interdependence of Economic and Environmental Systems*, Cambridge University Press, Cambridge.
- [19] Hashem Pesaran M., Potter Simon M. (1993) *Nonlinear dynamics chaos and econometrics*, John Wiley & Sons Ltd., Chichester.
- [20] Rieper Bernd, Witte Thomas, (1995) 3. Auflage. *Grundwissen Produktion*, Peter Lang, Frankfurt am Main.
- [21] Ruth Matthias, Hannon Bruce, (1997). *Modeling Dynamic Systems*, Springer, New York.
- [22] Smith Peter, (1998). *Explaining chaos*, Cambridge University Press, Cambridge.
- [23] Steeb Willi-Hans, (1991), *A handbook of terms used in chaos and quantum chaos*, BI-Wissenschafts-Verlag, Zürich.
- [24] Turnheim Georg, (1993) 2. Auflage. *Chaos und Management*, Gabler, Wiesbaden.
- [25] Uhlig Hans, (1999). *Finanzmarktanalyse: neue Ansätze aus der Chaosforschung*, mehrere Verlage
- [26] Drosdowski Günther, Köster Rudolf, Müller Wolfgang, (1977). *Das grosse Wörterbuch der deutschen Sprache*, Dudenredaktion, Mannheim, Wien Zürich.
- [27] Wöhe Günter, (1993) 18. Auflage. *Einführung in die allgemeine Betriebswirtschaftslehre*, Vahlen, München.

Kapitel 2

Introduction to Evolutionary Algorithms

Jan Baudisch

26. April 2000

2.1 Introduction

Evolutionary Algorithms, the generic term for modeling natural evolution by computers, gained meaning in recent years. This survey about Evolutionary Algorithms tries to explain the differences of the different types of Evolutionary Algorithms, and will present their history. Among Genetic Algorithms, which are by far the most commonly used, other concepts will be presented and compared.

2.2 Biological Background

In order to simulate evolution with computers, it is necessary to understand the evolution of nature. Therefore, the basics will be repeated briefly.

According to Darwin, evolution bases on the following principles:

- evolution is not directed or intended, but a permanent process of adaption to the changing environment.
- every individual has a different genotype and therefore differs from the others.
- every generation come much more creatures into existance than were necessary to replace the dead ones. Under well environmental conditions, the population grows exponentially. The growth is limited by the at most linary growing resources. Therefore, the individuals are under pressure for selection.
- Natural selection favors individuals which are best adapted to the environmental conditions (“survival of the fittest“).

2.3 Why Evolutionary Algorithms?

Evolutionary Algorithms are most commonly used for optimizing some parameters. There are different approaches in doing this:

- One possibility could be to try out all combinations of parameters. However this is often too costly, especially if you compute with real numbers, or if you have a lot of variables.
- Another method is to perform random search: Many times, the parameters are chosen randomly and the best solutions are picked out. In a large search-space, random search is also pretty unefficient.
- A better procedure is to use informations about the multi-dimensional function you want to optimize, like the gradient, which is the direction of the greatest increase. Every step you move a small distance in the direction of the gradient. After a while, you are situated in a maximum of the function. This procedure is called “hillclimbing“.
It fails, if the function has many local, suboptimal maximums, because then the process stays in that suboptimums. Another problem is that the function has to be differentiable.
- A solution for this problem can be Evolutionary Algorithms. They imitate natural evolution for finding a satisfying solution. Evolutionary Algorithms maintain a population of individuals, which are potential solutions for the problem which is to be solved. In order to get fit individuals, you use the genetic operators selection, mutation and crossover.

2.4 Genotype / Phenotype

As in nature, you can distinguish between a genotype and a phenotype. In biological creatures, the genotype is the genome of the individual, that means the complete genetic representation of his properties in the DNA. In Evolutionary Algorithms, the genotype is a kind of datastructure as a bitstring or a tree. The realisation of the genotype is called the phenotype. Often it is the value of the string, which has to be specified by a evaluation function.

2.5 The Genetic Operators

2.5.1 Selection

Roulette Wheel selection The parents which are destined for reproduction are selected with a probability which is proportional to the fitness of the individual. This way of selection is called “Roulette Wheel Selection“. It is possible, that the best individual does not reproduce itself, and that even poor individuals are selected, even if the probability might be low.

Elite Selection Evolution Strategies use a deterministic method which is called “Elite Selection“. It selects only the best individuals. Here it is guaranteed, that the genetic information of the best individual is being maintained.

The fitness can be determined by a simple function, or you let the individuals compete with each other, and select the ones which hold their ground best, which corresponds more to the natural example.

2.5.2 Recombination / Crossover

In order to exchange genetic information between individuals, the recombination - operator is being applied. It is regarded as the most important one at Genetic Algorithms.

Single Point Crossover The most simple mode is Single Point Crossover. Single Point Crossover takes two parents each, chooses a random position in the string, divides the two strings at this position, and combines the init of the first string with the tail of the second to a new string. Just the same you proceed with the beginning of the second and the end of the first. Thus, the offspring inherits properties of both parents.

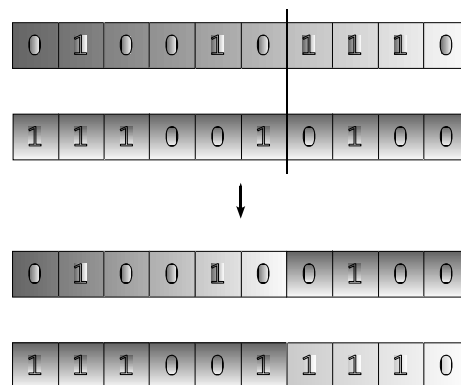


Abbildung 2.1: Single point crossover

Multi Point Crossover Single point crossover has some disadvantages. For example, if one string has quite good properties at its beginning and its end, and another string in the center, it is not possible to obtain an individual which inherits the best properties of the two parents.

To avoid this problem, you can cut the individuals at several positions, which is called multi point crossover. Yet, the principle of this problem stays, if you choose a number of cutting points which is smaller than the number of bits. The best results are being obtained, if you exchange each bit with a certain probability.

2.5.3 Mutation

The mutation-operator is responsible for innovation and variability. If just the recombination-operator is being used, it can happen, that after a while all individuals have the same bit at a certain position. This is called “Convergence“. So, the search is only performed in a subspace or the real searchspace.

To prevent this the heritable information is modified at each position with a low probability, for example by inverting a bit.

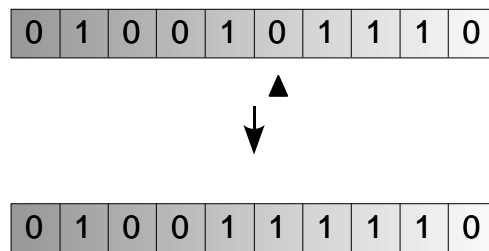


Abbildung 2.2: The mutation operator.

Evolution Strategies use just mutation and selection, and not recombination.

2.6 The Algorithm

The pseudocode of a Genetic Algorithm looks like that, but it is basically the same in all Evolutionary Algorithms.

```
create a random initial population
evaluate individuals
while (solution is not satisfying) do
    calculate fitness for each individual
    compute the probability for selection
    select parents of new offspring
    cross individuals to new offspring
    mutate individuals
end
```

If you have done everything well, the parents are replaced by better and better children. The evaluation function plays an important role, namely the same of the environment in organic evolution. It is the connection between the algorithm and the problem to be solved. The evaluation function gets a string as argument and returns a value, which is a measure for the performance of the chromosome.

2.7 History

First attempts of simulating evolution with computers were made at the end of the fifties by Richard Friedberg. His approaches failed among other things because of a lack of powerful computers in that time. At the beginning of the sixties, Evolutionary Algorithms were developed, which were successfully utilized in practical applications. John Holland invented at the University of Michigan the Genetic Algorithms, which are nowadays the most often used and well-known Evolutionary Algorithms. Genetic Algorithms use Bit-strings as a representation for the solution. In a Theory of GA's, Holland also tried to explain why GA's do successfully work. However, this theory is pretty controversial.

In the same time, two students at the Technical University of Berlin, Ingo Rechenberg and Hans-Paul Schwefel, searched for the optimal shape of bodies in a flow. He used a mutation operator for it and founded in doing so the Evolution Strategies. Now, Rechenberg is professor for Bionics at TU Berlin, and still teaches Evolution Strategies.

The youngest branch of EA's is genetic Programming. John Koza uses the programming language Lisp to represent the genetic information in trees of functions. In so doing, it is possible to produce dynamical source code, and model very complex behaviors.

2.8 Types of Evolutionary Algorithms

- Genetic Algorithms (GA)
- Evolutionary Programming (EP)
- Evolution Strategies (ES)
- Genetic Programming (GP)

Somewhat confusing might be the different types of Evolutionary Programming. The methods emphasize the genetic operators in a different way. In Genetic Algorithms, recombination is regarded as the most important operator, while in Evolutionary Programming this operator does not even exist.

Different is also the representation of the solution. Genetic Algorithms use primary Bit-strings, Evolutionary Programming and Evolution Strategies operate on vectors of real numbers, and Genetic Programming tries to optimize Computer Programs.

Furthermore, selection is realized in a different manner. In Evolution Programming, the selection is deterministic, because it always chooses the best individuals for selection. Genetic Algorithms have a random selection, so that even poor individuals have a chance to reproduce themselves.

Fourthly, the population size is constant at GA's and variable at EP and ES.

Evolutionary Programming are the least used and investigated Evolutionary Algorithms.

2.9 Areas of Application

- simulation of artificial life
- controlling of robots
- production of neuronal networks (selection of architecture)

- scheduling and transportation problems
- strategies for games, e.g. prisoner's dilemma
- travelling salesman problem

the problem is how to link a number of cities to a closed route, so that the covered distance is as short as possible. this is not a real application, yet in theoretical informatics it is regarded al a very difficult problem and a benchmark for techniques in computer science. Evolutionary Algorithms solve that problem quite well.

2.10 Example of a Genetic Algorithm

| maximisation of the function $f(x) = 2x$, $0 \leq x \leq 63$, x whole - numbered | | | | | |
|--|-------------------|---------------------------|------------------------------------|------------------------|----------------|
| initial population | decoded value | $f(x)$ | relative probability for selection | cumulative probability | random numbers |
| 0 1 0 1 0 0 | 20 | 40 | 0.16 | 0.16 | 0.834 |
| 0 0 1 0 1 1 | 11 | 22 | 0.09 | 0.25 | 0.505 |
| 1 1 0 1 0 0 | 52 | 104 | 0.41 | 0.66 | 0.017 |
| 1 0 1 0 1 1 | 43 | 86 | 0.34 | 1.00 | 0.283 |
| | | \sum 252 | | | |
| offspring | copies of parents | offspring after crossover | offspring after mutation | decoded value | $f(x)$ |
| 1 | 0 1 0 1 0 0 | 0 1 0 1 0 0 | 0 1 0 1 1 0 | 22 | 44 |
| 0 | 1 1 0 1 0 0 | 1 1 0 1 0 0 | 1 1 0 1 0 0 | 52 | 104 |
| 2 | 1 1 0 1 0 0 | 1 1 1 0 1 1 | 1 1 1 0 1 1 | 59 | 118 |
| 1 | 1 0 1 0 1 1 | 1 0 0 1 0 0 | 1 0 0 1 0 0 | 36 | 72 |
| | | | | | \sum 338 |

Abbildung 2.3: simple example of one loop of a genetic algorithm, taken from [3]; notice that the total fitness of the population raised from 252 to 338

2.11 Variation of Evolutionary Algorithms

Recently, the concept of *coevolution* gained meaning. That method maintains two or more populations. The fitness of one specie depends on the fitness of the other as, for example in predator - prey - systems.

Often used are *Hybrid Algorithms*. They you combine Evolutionary Algorithms with other methods of optimation, like fuzzy systems or neuronal networks. For example, Genetic Algorithms are used to select a good architecture for neuronal networks or to train neuronal networks.

You can also achieve some improvement by introducing a spacial dimension. You maintain a lot of "local distributed" partial populations, which exchange individuals once in a while. This way, the population converges more slowly to a suboptimum.

2.12 Valuation

Even if Evolutionary Algorithms have shown some promising results, and are gaining more and more meaning, there is a lack of a comprehensive theory, which explains, *why* Evolutionary Algorithms work and in which cases. Sometimes, Evolutionary Algorithms totally fail, because the most important thing is to find a clever representation of the problem, and sometimes this is very difficult. Therefore, you still depend on experiences and intuition. John Holland has begun to develop a theory of Genetic Algorithms, but he did not bring it to a favorable issue, and is heavily criticized.

On the other hand, Evolutionary Algorithms are robust against phenomena as uncertainty, which is a common property of soft computing methods like fuzzy systems and neuronal networks. If a component of the system as a chromosome in the Evolutionary Algorithms or a neuron in a Neuronal Network fails, the whole system will work anyway.

In any case, Evolutionary Algorithms will be more and more important in future.

Literaturverzeichnis

- [1] Back, T. (1996). *Evolutionary Algorithms in theory and practice*. Oxford Univ. Press, New York.
- [2] Michalewicz, Z. (1996). *genetic algorithms + data structures = evolution programs* Springer, Berlin.
- [3] Nissen, V. (1997). *Einführung in evolutionäre Algorithmen: Optimierung nach dem Vorbild der Evolution*. Vieweg, Braunschweig.

Kapitel 3

Künstliches Leben im Computer

Andreas Focks

3. Mai 2000

3.1 Einleitung

Dem Wesen des Lebens auf die Spur zu kommen, versuchen Wissenschaftler seit einer Reihe von Jahren. Die Faszination liegt vor allem darin, das der Mensch ständig sehen muss, wie hilflos er letztlich Leben und Tod gegenübersteht. Ewig zu leben ist immer noch ein Menschheitstraum, weil auch heute die Entstehung des Lebens auf der Erde weder geklärt ist noch nachvollzogen werden kann. Wissenschaftler in allen Disziplinen versuchen deshalb, den Begriff des Lebens in eine für sie fassbare Struktur zu bringen, mit der sie forschen, arbeiten und sich mit anderen austauschen können.

Das menschliche Leben, der menschliche Körper wurde zunächst von Medizinern untersucht, Biologen und Chemiker befassen sich mit der Erfassung und Charakterisierung von lebenden Organismen und deren strukturellen Voraussetzungen. Philosophen und Soziologen interessieren sich für Struktur und Auswirkung des Zusammenlebens auf der Erde. In diesem Zusammenhang existiert seit wenigen Jahren ein neuer Forschungszweig, der sich mit der Entstehung und Struktur lebender Systeme beschäftigt: Artificial Life.

3.1.1 Leben

Fragt man sich, was man selber eigentlich unter Leben versteht, fällt es einem nicht schwer, Beispiele für Leben zu finden. Jeder Mensch lebt selbstverständlich, auch Pflanzen und Tiere. Abgrenzungen zu finden fällt dagegen schon schwer und führt teilweise zu Auseinandersetzungen in Forschungskreisen. So haben ein Mikrobiologe und ein Ökologe bestimmt unterschiedliche Ansichten darüber, ob ein Virus lebt oder nicht. Die entscheidende Frage in diesem Zusammenhang ist, ob ein Virus als lebendig bezeichnet werden kann, obwohl es sich anderer Organismen bedienen muss, um sich fortzupflanzen. Noch schwieriger wird die Abgrenzung, wenn man sich an der Klassifikation von Einheiten versucht, die nicht im klassischen Sinne als Lebewesen gelten, aber durchaus Eigenschaften lebender Systeme aufweisen. Kann man zum Beispiel, wie in der Gaia-Theorie beschrieben, die Erde als Lebewesen betrachten? Lebt das Internet, insbesondere auch Figuren aus dem Cyberspace? Um eine Grundlage zur Abgrenzung von lebenden zu nicht lebenden Systemen zu ermöglichen, gibt es verschiedene Forderungen, die eine Einheit erfüllen sollte, um als lebendig

gelten zu können ([2], S.33).

Ein lebender Organismus soll

- sowohl im Raum als auch in der Zeit existieren
- sich in einem eigenen oder in einem verbundenen Organismus reproduzieren
- Informationen über sich selbst speichern
- einen Stoffwechsel haben, innerhalb seiner Umgebung aktiv werden
- aus voneinander abhängigen Teilen bestehen
- bei Veränderung der Umweltbedingungen stabil bleiben
- sich entwickeln, wachsen oder sich ausdehnen

Ob ein System dabei allen oder nur einigen dieser Anforderungen genügt, soll keinen Einfluss darauf haben, ob man dieses System nun als lebendig bezeichnet oder nicht. Eine qualitative Aussage über die jeweilige Lebendigkeit ist anhand dieser Liste ohnehin nicht möglich. Die Forderungen sollen nur die bisherigen Definitionen von Leben, die z.B. in der Biologie bisher verwendet wurden, erweitern, so dass man versuchen kann, auch künstliches Leben anhand dieser Forderungen zu überprüfen.

3.1.2 Künstliches Leben

Ob die Entstehung des Lebens auf der Erde nicht einem einfachen Zufall entspringt, kann niemand wissen. Es drängt sich einem jedoch mit zunehmenden wissenschaftlichen Erkenntnissen in allen Bereichen der Forschung der Eindruck auf, dass Leben als ein logisches Produkt der komplexen Wechselwirkungen zwischen den verschiedenen abiotischen Faktoren der Umwelt in der Frühzeit unseres Planeten betrachtet werden kann. Stellt man sich nun einmal vor, die Evolution, wie sie mit der Bildung der ersten genetischen Information begann, ließe sich noch einmal abspielen, lässt man die Äonen vor seinem geistigen Auge im Zeitraffer ablaufen, kann man sich vorstellen, dass das Ergebnis dieses Gedankenexperimentes identisch mit unserer heutigen Welt sein sollte? Das Leben, wie wir es heute auf der Erde vorfinden, ist manifest, ein Ausdruck einer langen Entwicklung und wir kennen nur dieses Leben. Deshalb wird es auch als *life-as-we-know-it* bezeichnet. Im Unterschied dazu kann man künstliches Leben bezeichnen als *life-as-it-could-be*. Die Verwendung des Begriffes *Leben* für die Ergebnisse der bisherigen AL-Forscher muss deshalb immer in diesem Zusammenhang verstanden werden als Ergebnis einer künstlichen Evolution, die sich versucht, am einzigartigen natürlichen Vorbild zu orientieren. Dasselbe gilt für die Verwendung von Begriffen aus verschiedenen Bereichen der Biologie.

3.1.3 Grundsätzliches zu Artificial Life

Artificial Life (AL) ist immer die Arbeit mit Experimenten. So wie ein Experiment im Chemielabor aus einem Versuchsaufbau, aus verschiedenen chemischen Verbindungen und der Zufuhr von Energie besteht, so besteht ein Experiment im Computer aus einer definierten Umgebung, strukturierter Information und Zeit - vor allem Rechenzeit als Energieform. Die Schöpfung künstlichen Lebens erfolgt nach verschiedenen Ansätzen, die jedoch alle

eines gemeinsam haben: Kein AL-Forscher wird versuchen, einen digitalen Frankenstein hervorzubringen, indem er sich eine Kreatur vorstellt, die er dann Stück für Stück zusammensetzt, bis sein Bild Wirklichkeit geworden ist. Die Stärke von AL liegt im Gegenteil, und damit auch im Widerspruch zum Ansatz der Forschungen zur künstlichen Intelligenz (KI) im Bottom-up-Prinzip. Durch lokale Interaktionen zwischen einfachen Strukturen entstehen globale Interaktionen, entsteht Komplexität in einer Weise, die man schlecht vorhersagen kann. Ein typischer Versuchsaufbau eines AL-Experimentes könnte zum Beispiel darin bestehen, einen wenige Zeilen umfassenden Algorithmus als Programm in einen Bereich des Arbeitsspeichers eines Computers zu setzen. Dabei hat der Algorithmus zum Ziel, sich selbst möglichst oft zu kopieren. Da dieser Algorithmus dabei mit anderen Algorithmen konkurrieren muss, welche das gleiche Ziel haben, nämlich ebenfalls Rechenzeit und Speicherplatz des Computers für sich in Anspruch nehmen zu wollen, und weil man es den Organismen gestattet, sich gemäss einiger Gesetzmässigkeiten der Evolution zu verändern, entwickeln sich im Speicher dieses Computers immer perfektere kleine Organismen.

Thomas Ray hat zu Beginn seines Tierra-Projektes eben ein solches Experiment gestartet und nach einiger Zeit festgestellt, das einige der Algorithmen kürzer waren, als sie eigentlich sein mussten, um sich selbst kopieren zu können. Als er diese Algorithmen näher betrachtete, stellte er fest, das sie sich der Kopierfunktion anderer Bewohner des Arbeitsspeichers bedienten, um sich selbst zu kopieren. Es hatte sich also eine Art Parasitismus entwickelt, ohne das Thomas Ray dies zu Beginn seines Experimentes hätte voraussehen können.

Die Verhaltensweisen einer digitalen Kreatur sind einfach, da sie algorithmisch erfasst und abgearbeitet werden können. Durch die Reproduktion dieser Kreaturen und durch die Notwendigkeit, einer wie auch immer gearteten Selektion zu entgehen und den dadurch entstehenden Druck, sich selbst in Hinblick auf diese Selektion zu optimieren, entstehen in einem evolvierenden Prozess Strukturen, die man sonst nur in lebenden Systemen beobachten kann. Somit wird der Computer als universelle Rechenmaschine, als ein Werkzeug zur Abarbeitung von algorithmischen Anweisungen, zu einem Labor, in dem sich eine hohe Anzahl von Organismen in einer definierten Umgebung zeitlich nahezu unbegrenzt entwickeln kann. Dabei bedeutet die Tatsache, dass heute immer mehr Rechenleistung zur Verfügung steht, nicht, dass damit selbstverständlich künstliche Lebewesen mit einer höheren Komplexität entstehen. Im Gegenteil zeigen die Experimente, bei denen wirklich interessante Phänomene auftreten, diese oft schon nach kurzer Zeit. Es kommt also auf die richtigen Anfangsbedingungen an. Welche Bedingungen gelten müssen, um in einem System Komplexität und Leben entstehen zu lassen wird heute im Bereich der Komplexitätsforschung untersucht. Grundsätzlich hat heute jeder, der einen Schreibtisch-PC und genug Interesse besitzt, einen Experimentierkasten zur Erforschung von Komplexität und künstlichem Leben zur Verfügung, der den Pionieren der AL-Forschung nicht zur Verfügung stand.

3.2 Ergebnisse aus der Erforschung von künstlichem Leben

Lebendige Organismen im Sinne des Artificial Life haben gemeinsam, dass sie in Strukturen auftreten, in denen sich Informationen innerhalb einer bestimmten Umgebung mit Hilfe von verschiedenen Methoden verbreiten. Die folgenden Abschnitte sollen einen Über-

blick über die wichtigsten Ergebnisse im Bereich des künstlichen Lebens geben und dabei beschreiben, welche Methoden und Strukturen bisher verwendet wurden.

3.2.1 Zellularautomaten

Der selbstreproduzierende Automat

Die Grundlagen zur Erforschung von künstlichem Leben legte der ungarische, in die USA emigrierte Wissenschaftler John von Neumann. Er fasste das Leben als rekonstruierbare Kette von Ereignissen und Wechswirkungen auf und vertrat die Ansicht, Leben beruhe auf logischen Regeln. Er war fasziniert von der Möglichkeit, eine Theorie aufzustellen, die die ausserordentliche Komplexität von Lebewesen und deren bislang ungeklärte Ursache erklären könnte. Um diese Theorie zu erarbeiten versuchte sich von Neumann zunächst mit der gedanklichen Konstruktion eines selbstreproduzierenden Automaten, da er die Selbstreproduktion als die komplizierteste biologische Fähigkeit einschätzte. Auf Vorschlag des ungarischen Wissenschaftlers Stanislaw Ulam konstruierte er die zunächst als kinematisches Modell auf einen imaginären See postierte Kreatur als zelluläres Modell eines selbstreproduzierenden Automaten auf einem horizontlosen Schachbrett. Die Felder oder Zellen dieses Schachbrettes konnten verschiedene Zustände einnehmen, dadurch wurde der Automat definiert. War er einmal auf dem Schachbrett verankert, unterlag er bestimmten Regeln, genauer gesagt begann jede einzelne Zelle der Kreatur wie eine Finite-state-machine den ihr zugeordneten Regeln zu folgen.

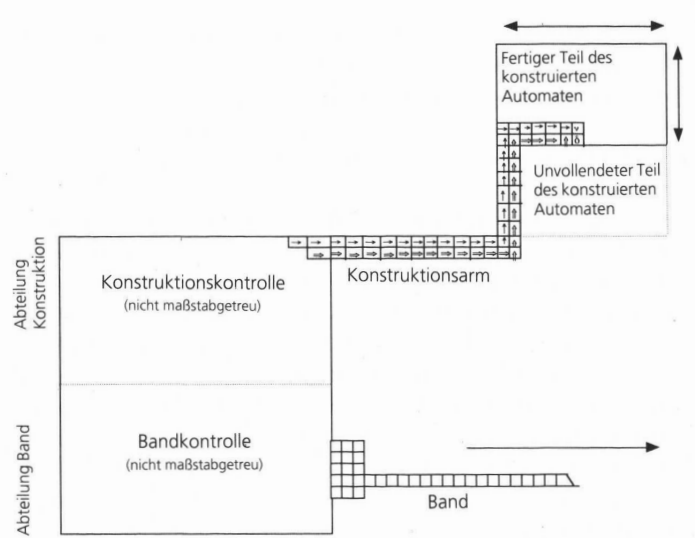


Abbildung 3.1: ([1],S.59) Von Neumanns selbstreproduzierender Automat.

Durch das Zusammenwirken der lokalen Verhaltensweisen sollten dann globale Verhaltensmuster entstehen können. Der amerikanische Wissenschaftler John G. Kemny, Entwickler der Programmiersprache BASIC, kam ebenso wie von Neumann selbst zu dem Schluss, dass eine solche Maschine einen evolutionären Prozess durchlaufen könnte.([1], S.61) Von Neumann starb jedoch 1957, bevor er seine Arbeiten an diesem Modell eines selbstreproduzierenden Automaten beenden konnte.

Trotzdem hatte von Neumann mit dieser Arbeit, die er neben seiner eigentlichen Aufgabe als Wissenschaftler der US-Luftwaffe ausführte, den Grundstein für die Erforschung künstlichen Lebens geschaffen: Der zelluläre Automat war geboren.

Life als universelle Rechenmaschine

Der nächste bedeutsamen Schritt erfolgte Ende der sechziger Jahre auf Initiative des britischen Mathematikers John Conway. Dieser Wissenschaftler hatte sich in den Kopf gesetzt, von Neumanns Automat, der als universelle Rechenmaschine fungieren konnte, also in der Lage war, jedes beschreibbare Problem anderer Maschinen durch die Benutzung einer Anzahl logischer Regeln abzuarbeiten, zu vereinfachen. Conway glaubte, dass ein Zellularautomat mit universeller Rechenleistung einfacher darzustellen sei als der von Neumanns, welcher aus 200 000 Zellen mit jeweils neunundzwanzig verschiedenen Zuständen bestand. Sein Ansatz beinhaltete, dass eine Zelle nur zwei verschiedene Zustände, die Werte 1 und 0 annehmen, lebendig oder tot sein kann. Schlüssel zum Erfolg jedoch waren, nachdem Conway Anzahl der Zustände reduziert hatte, die Regeln, nach denen sich im zellulären Automaten entschied, ob eine Zelle im nächsten Zeitschritt lebt oder nicht. Es waren zwei Jahre Arbeit notwendig, um diese Regeln zu finden, die schliesslich bekannt werden sollten als *Conways Game of Life*. Diese Regeln beinhalteten den Keim zu einer Entwicklung die, natürlich abhängig von der jeweiligen Ausgangskonfiguration des zellulären Automaten, weder ein chaotisches Verhalten zeigte noch zu einem sofortigen Aussterben oder einer sofortigen vollständigen Bevölkerung des zellulären Automaten führte. Conway und seine Kollegen von der mathematischen Fakultät der Universität Cambridge testeten diese Regeln, indem sie im Gemeinschaftsraum auf kariertem Papier per Hand quadratische Figuren und Muschelschalen als Symbole für die beiden möglichen Zustände der Zellen des Automaten plazierten und dann Zelle für Zelle entsprechend der Regeln in ihren nächsten Zustand überführten, indem sie entweder die Zelle unverändert liessen oder eine quadratische Figur oder eine Muschelschale auf ihr plazierten. Sie versuchten anschliessend, Conways zellulären Automaten darauf zu testen, ob er tatsächlich eine universelle Rechenmaschine darstellt. Nachdem Mitarbeiter der Massachusetts Institute for Technologie (MIT) eine auf dem Game of Life basierende Rechenmaschine entworfen hatten ([1], S,75), war diese Bedingung erfüllt. Conway hatte es also wirklich geschafft, den Automaten von Neumanns zu vereinfachen und trotzdem die Fähigkeiten des selbstreproduzierenden Automaten beizubehalten. Die Bits in der Rechenmaschine bestanden aus sogenannten Gleitern. Gleiter sind Muster in Conways zellulärem Automaten, die sich im Raum fortbewegen können, die also sozusagen über das Spielfeld gleiten.

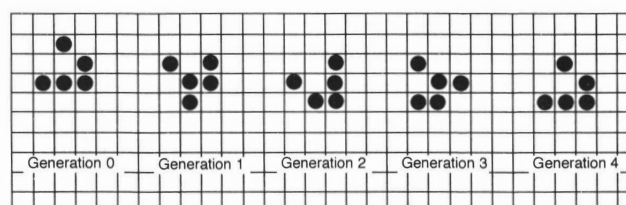


Abbildung 3.2: ([1], S.73) Gleiter als sich fortbewegendes Muster aus Life.

Klassifizierung der Komplexität

Die Arbeiten von Neumanns und Conways fanden erst 1982 eine Fortsetzung, als Stephen Wolfram anfang, über die Verwendbarkeit von zellulären Automaten als Grundlage für Experimente im naturwissenschaftlichen Bereich nachzudenken. Er nahm an, dass er zelluläre Automaten dazu benutzen könnte, Erkenntnisse über komplexe Systeme im

Allgemeinen zu erhalten. Die Vorstellungen, die Wolfram selbst von der Entstehung von Komplexität aus sehr einfachen Ursprüngen hatte, wollte er anhand von zellulären Automaten untersuchen.

Er verwendete dazu einen selbstentworfenen eindimensionalen zellulären Automaten. Der Vorteil für Wolfram bestand vor allem darin, dass sein Automat mit 256 möglichen Ausgangskonfigurationen einfacher zu überschauen war als auch Conways vereinfachter Automat. Wolfram wertete Probelaufe des Automaten mit allen möglichen Ausgangskonfigurationen aus, indem er die entstehenden Muster in vier Klassen einteilte:

- homogene, unveränderliche Muster
- sich periodisch ändernde Muster
- chaotische Muster
- komplizierte, räumlich voneinander getrennte Muster, die sich zeitlich unbegrenzt weiterentwickeln, ohne in eine der drei anderen Klassen von Mustern zu münden

Die vierte dieser Klassen, welche auch auf mehrdimensionale zelluläre Automaten übertragbar sind, beinhaltet dabei alle Zellularautomaten, die als universeller Computer angesehen werden können, zutreffend also auch auf Conways Game of Life. Interessant ist aber auch die dritte von Wolframs Klassen, deren Muster ein chaotisches Verhalten zeigen. Betrachtet man ein Muster aus dieser Klasse, würde man annehmen, dass es durch Zufall entstanden sein muss, da keine Regel, nach der das Muster aufgebaut ist ersichtlich ist. Es gibt jedoch diese Regeln im zellulären Automaten, und sie sind einfach - genau so einfach wie die Regeln, unter denen sich die Muster der Klasse vier und letztlich auch biologische Lebensformen aufbauen. Diese hoffte Wolfram zu finden, bevor er mit der Entwicklung der Computeralgebra-Software Mathematica begann und sich aus der Erforschung künstlichen Lebens zurückzog. Dabei finden sich von Wolframs zellulärem Automaten erzeugte Muster fast identisch in der Natur wieder, so zum Beispiel auf der Schale von bestimmten Meeresschnecken.

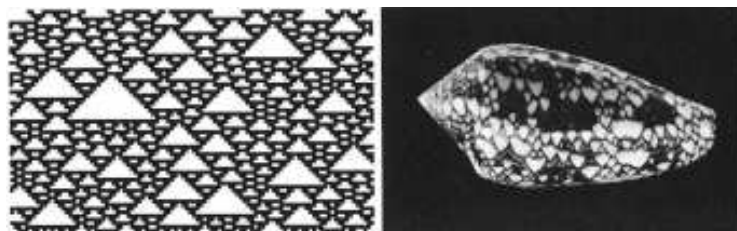


Abbildung 3.3: ([1], S.95) Schale einer Meeresschnecke und eindimensionaler Automat der Klasse 3

Bottom-up und der λ -Parameter

Der nächste Wissenschaftler, der zelluläre Automaten für seine Forschungen verwendete, war der Amerikaner Chris Langton. Er verwendete einen Automaten mit 94 Zellen und acht möglichen Zellzuständen, um etwas zu erschaffen, was er als selbstreproduzierende Schleife bezeichnete. Aufbauend auf einer bestimmten Ausgangskonfiguration und basierend auf einem Regelwerk, an dem Langton lange gearbeitet hatte, war dieser Automat in der Lage, eine genaue Kopie von sich selbst zu produzieren und die Kopie von

sich abzutrennen, so dass in Folge zwei Schleifen begannen, Kopien von sich herzustellen. Diese Schleife erfüllte damit alle Kriterien, die man an ein Lebewesen stellen kann, geht man davon aus, das ein Lebewesen keinen organischen Stoffwechsel betreiben muss, um als lebendig gelten zu dürfen, sondern das die Umwandlung von Information in Rechenzeit, also Energie, als synthetisches Pendant des bei organischen Lebewesen vorhandenen Stoffwechsels gelten darf.

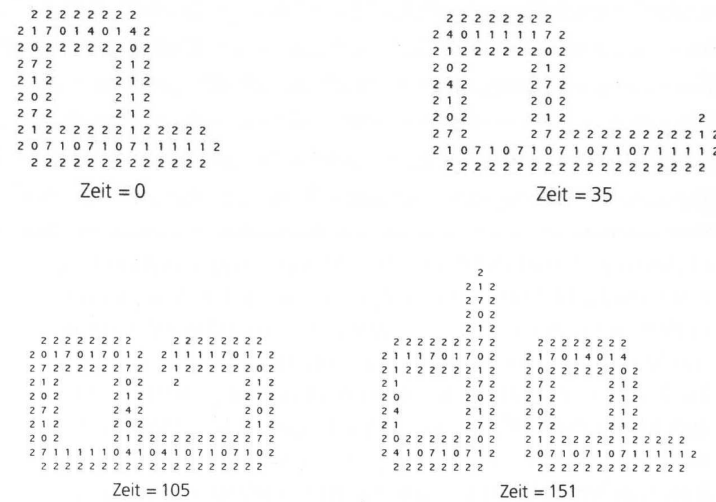


Abbildung 3.4: ([1], S.124/125) Langtons selbstreproduzierende Schleife.

Als Langton dann feststellte, dass sich die Schleifen, während sie sich vervielfältigten, eine Ordnung einhielten, die der von Korallen stark ähnlich ist, war er überzeugt, einen wichtigen Schritt in der Frage nach dem Ursprung der Komplexität und der Herkunft des Lebens getan zu haben. Während nachfolgenden Arbeiten mit durch Zustände in benachbarten Zellen simulierten Ameisen in einem zellulären Automaten, in der die Individuen verblüffend genau die unter anderem durch Pheromone erfolgende Kommunikation unter realen Ameisen wiedergaben, beschäftigte er sich längere Zeit mit entomologischen Publikationen und fasste seine bisherigen Erfahrungen schliesslich unter dem Begriff *Bottom-up* zusammen. Bottom-up steht für die sowohl in der Natur wie auch in Langtons Zellulardautomatensimulation vorkommende fehlende Steuerung von komplexen Verhaltensweisen. Schlüssel für ihn war in dieser Beziehung der Satz „Daher muss keine der Termiten mit einem Plan in der Hand als Aufseher fungieren“, den er in einer Publikation las und der ihn nicht mehr loslies, weil er ihn an das Verhalten seiner virtuellen Ameisen erinnerte ([1], S.132). Das Konzept Bottom-up beschreibt die Kollektivkraft kleiner Aktionen, die sich nach oben ausbreiten, mit anderen Aktionen verbinden und häufig rekursiv verlaufen, so dass als Ergebnis globale Verhaltensweisen oder auch emergente Eigenschaften entstehen, die man nur aus Kenntnis der lokalen Aktionen nicht hätte voraussagen können.

Im folgenden machte Langton sich an die Arbeit, herauszufinden, welche Voraussetzungen für ein Universum, also auch für einen zellulären Automaten gelten müssen, damit es Leben unterstützt. Dabei mass er dem Informationsfluss in einem System eine entscheidende Bedeutung zu, so entscheidend, dass er in seinen Experimenten einen eigenen Parameter für die Dynamik des Informationsflusses verwendete, den λ -Parameter. Dieser λ -Parameter bildet mathematisch die Dynamik des Austausches von Information, die durch ein bestimmtes Regelwerk gegeben wird, auf einen Zahlenwert zwischen Null und

Eins ab. Langton stellte schnell fest, dass der Bereich zwischen einem zu hohen Informationsfluss, welcher den zugehörigen zellulären Automaten in einen chaotischen Zustand stürzte, und einem zu geringen Informationsfluss, der sich in langweiligen Anordnungen des zugehörigen zellulären Automaten äusserte, der interessante Bereich war. Diese Qualität des Informationsflusses, deren zugehörige Automaten zur Klasse vier der von Wolfram eingeführten Klassifizierung gehören, schien also die Entstehung von Leben zu begünstigen. Langton fand eine Art Phasenübergang von einem Zustand in einen anderen vor, den schon von Neumann für komplexe Systeme vermutet hatte:

„Bei komplexen Systemen gibt es eine kritische Grösse, bei deren Unterschreitung sich der Prozess der Synthese in Abbau umkehrt und bei deren Überschreitung das Phänomen der Synthese explosiv werden kann.“ ([1], S.137)

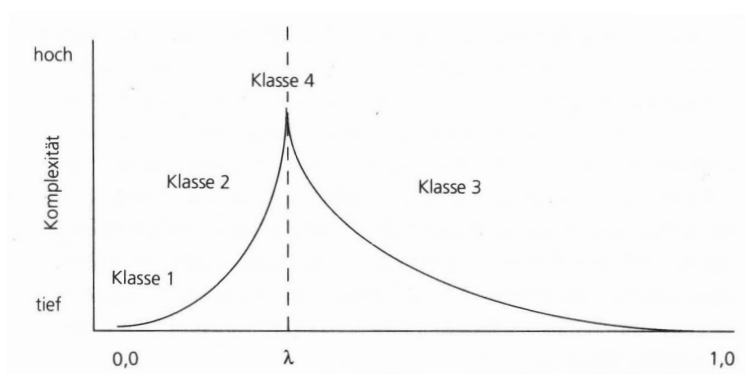


Abbildung 3.5: ([1], S.139) Zusammenhang zwischen Komplexität des zellulären Automaten und der Intensität des Informationsflusses.

Beeindruckt von diesen Ergebnissen berief Langton 1987 eine Konferenz in Los Alamos ins Leben, die später als erster Workshop über künstliches Leben bekannt wurde.

3.2.2 Genetische Algorithmen

Aus nichts etwas machen

Das Verfahren des genetischen Algorithmus geht zurück auf den amerikanischen Wissenschaftler John Holland. Es stützt sich darauf, dass Information in der natürlichen Evolution auf zweierlei Weise betrachtet wird: Als genetische Information, die verdoppelt wird und als genetische Instruktion, die ausgeführt werden muss. John Holland entwickelte nun zwischen 1962 und 1965 zusammen mit Doktoranden an der Universität von Michigan einen Algorithmus, der Computerprogramme in einer Art und Weise optimiert, wie sie in der biologischen Vererbung von Genen verwendet wird ([1], S.199). Die Fitness eines Computerprogramms wird jedoch, anders als in der natürlichen Vererbung, in der die Tauglichkeit durch das Überleben in der natürlichen Umwelt bestimmt wird, durch einen auf die Lösung eines bestimmten Problems bezogenen Fitnesswert bestimmt. Die Programme, welche den höchsten Fitnesswert bezogen auf das zu lösende Problem besitzen, haben in Hollands genetischem Algorithmus die höchste Wahrscheinlichkeit, ihr genetisches Material, nämlich den Programmcode, eine Generation weiterzugeben. Beim Schritt in die nächste Generation, oder auch der Vererbung, werden nach natürlichem Vorbild auch Mutation und Crossing-Over als Werkzeuge eingesetzt.

Holland vertraute bei der Entwicklung dieses Werkzeuges darauf, dass die Kräfte, welche die natürliche Evolution antreiben auch in der künstlichen Umgebung eines Computers

wirken müssen, wenn man es geschickt anstellt. Er hoffte zudem darauf mit der Simulation eines biologischen Systems auch bisher nicht ermöglichte Einblicke in den dynamischen Prozess während der Evolution dieses Systems zu bekommen.

Wie wichtig sind Mutationen?

Ein Beispiel dafür, wie der genetische Algorithmus arbeitet, lieferte David Jefferson von der University of Californias (UCLA). Er benutzte in einem Versuch den genetischen Algorithmus dazu, aus Information bestehenden Ameisen das bei ihren natürlichen Artgenossen zu beobachtende Duftspurverhalten beizubringen. Zu diesem Zweck konstruierte der Wissenschaftler ein torusförmiges Netz von 32×32 Quadraten, auf denen eine Folge von 89 markierten Quadraten den Weg markiert, dem die Ameisen folgen sollen. Diese Ameisen bestehen eigentlich aus Ketten mit jeweils 450 Bits, die vom Computer als Finite state machines angesehen werden. Er liess nun eine Population von 65536 dieser Ameisen, die aus zufällig aneinander gereihten Nullen und Einsen bestanden, auf dem Netz über 200 Zeitschritte laufen. Gemäss den Gesetzen der Selektion wurden die Ameisen ermittelt, welche möglichst viele Teile des Pfades erfolgreich absolviert hatten. Die besten zehn Prozent der Urpopulation wurde so oft kopiert, bis wieder 65536 Individuen vorhanden waren. Danach wurden die Bitketten, aus denen jedes Individuum bestand, mittels der Vefahrens des Crossing-over und der Mutation neu gemischt. Die Mutationsrate hatte bei einem ersten Lauf einen Wert von einem Prozent, bei einem nächsten Lauf einen Wert von einem Promill. Die Tatsache, das sich die Ergebnisse trotzdem nicht voneinander unterschieden, hatte John Holland schon vermutet([1], S.209). Er behauptete, dass den Mutationen im Bereich der Biologie sehr viel mehr Bedeutung beigemessen wird, als ihnen eigentlich in Bezug auf die Optimierung von genetischer Information eigentlich zusteht.

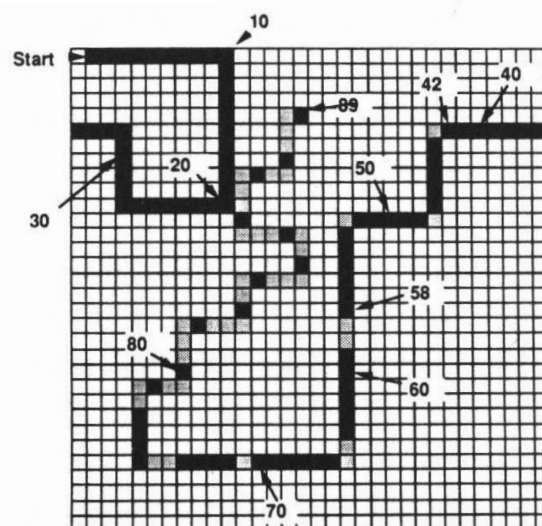


Abbildung 3.6: ([1], S.206) Der Pfad, dem die Ameisen in Jeffersons zellulärem Automaten folgen sollten.

Eine der erfolgreichen Ameisen in Jeffersons Netz war *Champ 100*. Dieser Organismus konnte nicht nur den John-Muir-Pfad genannten, markierten Quadraten folgen, sondern auch Hindernisse wie unmarkierte Felder hinter sich lassen. Diese Fähigkeit hatte Champ 100 nach hundert Zeitschritten gelernt.

Erkundung des Genetic Space

Der Evolutionsbiologe Richard Dawkins war einer der wenigen Biologen, welcher von den Möglichkeiten der genetischen Programmierung Gebrauch machte. Er entwickelte in der Vorbereitungszeit seines Buches „*Der blinde Uhrmacher*“ ein Programm, welches dem Benutzer gestattet, aus einfachen botanischen Strichzeichnungen entstandene pflanzen- und insektenähnliche Formen, die er Biomorphe nannte, zu erzeugen. Dabei spricht Dawkins bei diesem Prozess nicht von einem Schöpfungsakt, sondern er vertritt die Auffassung, man entdecke diese Biomorphe nur, da sie als Permutationen eines vorhandenen Satzes von Genen bereits existieren ([1], S.216). In der Biologie spricht man von diesen Permutationen als *genetic space*. Ein Biomorph bestand aus neun Genen, die mit jeweils neunzehn möglichen Allelen besetzt waren, die für Merkmale wie Verzweigung, Gliederung und Symmetrie verantwortlich waren. Dawkins stellte die Biomorphe her, indem er aus einer in einem Schritt erzeugten Anzahl verschiedener Biomorphe zwei auswählte, die er rein äusserlich besonders hübsch oder interessant fand, und verwendete deren genetisches Material im nächsten Schritt wieder, um neue Varianten der Ursprungs-Biomorphe zu erzeugen. Dies geschah unter Verwendung von Werkzeugen der genetischen Programmierung. Dawkins hielt sich dabei nicht möglichst nahe an die Evolution, sondern versuchte eher, optisch ansprechende Formen zu erzeugen. So entdeckte er in einer Schlüsselsituation ein Biomorph, dessen Bild er als *Zarathustra-Insekt* bekannt machte. Diese Kreatur hatte sich während der Bearbeitung des ursprünglichen Gensatzes durch Reproduktion und Mutation 29 Schritte von seinem Urahn entfernt. Für jeden einzelnen Schritt gab es etwa eine halbe Billion Permutationen. Man kann sich also kaum vorstellen, wie gross die Anzahl der möglichen Biomorphe ist, die das Zarathustra-Insekt während seines Entdeckungsprozesses passiert hat.

Aus diesem Hintergrund heraus folgerte Dawkins, dass Evolution ein einflussreicher Mechanismus ist, um den *genetic space* zu durchsuchen. Wissenschaftler wie Holland oder Jefferson hatten vorher mit ihren genetischen Algorithmen etwas ähnliches geschafft. In ihrem Fall spannte die jeweilige Problemstellung den sogenannten *problem space* auf, den die Wesen in ihren kleinen Universen auf der Suche nach der besten Lösung durchsuchten. Sie waren ähnlich erfolgreich, wie Dawkins mit Hilfe seiner genetischen Programmierung auf der Suche nach naturähnlichen Formen.

3.2.3 Neuronale Netze

Epistatische Gene und Coevolution

Der amerikanische Wissenschaftler Daniel Hillis legte mit seiner Dissertation den Grundstein für eine Reihe von weiteren Experimenten im Bereich des AL. Er entwarf einen Parallelrechner, der mit 64K, also 65536 Prozessoren arbeiten sollte. Gegen den ungläubigen Protest der Informatikwelt schaffte er es auch, dieses Gerät von seiner eigens gegründeten Firma bauen zu lassen. Insbesondere für militärische Zwecke interessant, wurden die Rechner für einen Stückpreis von bis zu 50 000 000 Dollar verkauft. Hillis aber hatte den Rechner eigentlich nur zu einem Zweck konstruiert: Er wollte einen leistungsfähigen Rechner zu Verfügung haben, leistungsfähig genug, um dem Geheimnis des Lebens auf die Spur zu kommen. Hillis interessierte sich schon früh für bestimmte Zusammenhänge in der Biologie und hatte seine eigenen Thesen. Einmal im Besitz der Verknüpfungsmaschine, wie er den Parallelrechner genannt hatte, begann er, seine Thesen anhand von Experimenten zu überprüfen.

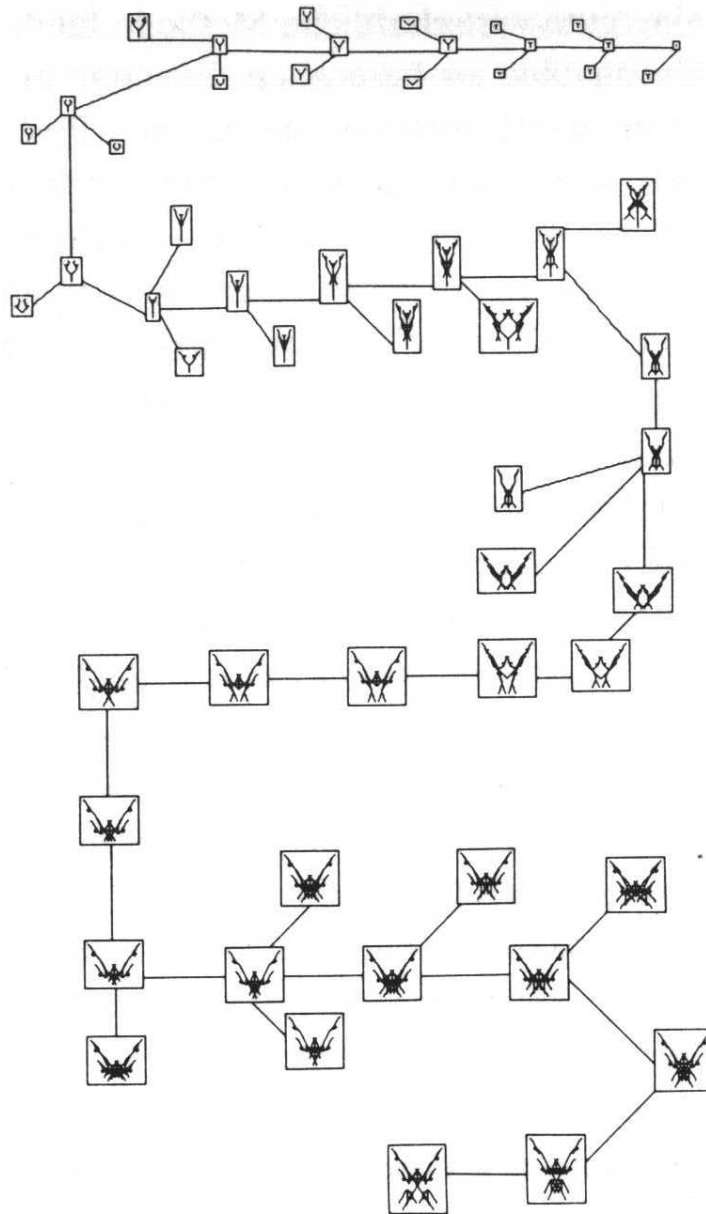


Abbildung 3.7: ([1], S.217) Entwicklungsbaum des Zarathustra-Insekts

Der Vorteil, den Hillis dabei gegenüber den Biologen hatte, war immens. Er konnte nicht nur die Phänotypen seiner künstlichen Populationen untersuchen, sondern er konnte sich jederzeit den gesamten Genpool einer Population anschauen, das heisst, er musste ihn sich nicht einmal anschauen, sondern konnte sich vom Rechner z.B. Daten über die Häufigkeit eines Gens in der ganzen Population gezielt ausgeben lassen. Dank der immensen Geschwindigkeit, welche die Verknüpfungsmaschine bei der Simulation von Populationen einer Grösse von 64 K erreichte, konnte Hillis Populationen hunderttausende Generationen in einer Zeit laufen lassen, die in keinem Verhältnis zu der Zeit, die Biologen für Versuche z.B. mit *Drosophila* benötigten, stand.

Zunächst fand er heraus, dass es für die Fitness einer bestimmten Art vorteilhaft ist, wenn die Art sich gegenüber Parasiten oder Räubern behaupten muss. Diese unter Biologen schon unter dem Namen *Rote-Königin-Effekt* bekannte Vermutung belegte er anhand von sogenannten Rampen, kleine Programmen, deren Fitness dadurch bestimmt war, mit wie wenig Schritten sie in der Lage waren, zwölf Zahlen zu sortieren. Ohne Konkurrenten schafften die besten seiner Rampen es, in 65 Schritten die Zahlen zu sortieren. Der beste von Menschen erdachte Algorithmus brauchte zu der Zeit 62 Schritte. Als Hillis die Rampen mit einer Art Parasit, den Antirampen sich gemeinsam entwickeln liess, kam er in einem kürzeren Evolutionszeitraum auf eine Art, welche das Sortierproblem in 61 Schritten bewältigten. Hillis Triumph, als er glaubte, seine Organismen hätten den besten menschlichen Programmierer geschlagen, war beendet, als er von einem Programmierer erfuhr, dessen Algorithmus die Sortierung in nur noch 60 Schritten schaffte ([1], S.250). Trotz der Enttäuschung darüber blieb Hillis die Gewissheit, dass er einen unter Evolutionbiologen lange vermuteten Effekt bestätigen konnte. Er verhalf weiterhin den Biologen, das Konzept der *Hill-climbing-techniques*, welches die Fitness von Populationen als eine Art Hügellandschaft betrachtet, in der Individuen mit unterschiedlichen Höhen unterschiedliche Fitnesswerte erreichen, zu erweitern. Die Existenz von coevolvierenden Individuen hatte Hillis Untersuchungen nach zur Folge, dass an lokalen Fitnessmaxima sitzende Genkombinationen von ihren Gipfeln vertrieben werden und so erst die Möglichkeit erhalten, noch höhere Fitnesswerte zu erreichen.

Durch die Möglichkeit, die Genotypen ebenso einfach analysieren zu können wie die Phänotypen kam Hillis im weiteren Verlauf seiner Forschungen einem anderen Phänomen auf die Spur, welches Biologen seit längerem beschäftigte. Unter dem Begriff *unterbrochenes Gleichgewicht* fassten Wissenschaftler die Tatsache zusammen, dass Arten erst nach einer Phase der genetischen Stabilität wieder ein qualitativer Sprung in Bezug auf ihre Tauglichkeit gelingt. Biologen machten dafür eine erfolgte Anpassung an die Umwelt oder Mutationen verantwortlich, welche die Population mit einer höheren Fitness ausstattet. Hillis konnte diesen Erklärungsversuchen nicht so recht Glauben schenken und beschäftigte sich mit der Frage, wie die Evolution so etwas Komplexes wie ein Auge hervorbringen kann, wenn dafür eine ganze Anzahl von Genen verantwortlich sein muss, die einzeln keinen Fitnessvorsprung bedeuten. Die Biologen sprechen bei diesem Sachverhalt von *epistatischen Genen*, die in einer ausgewogenen Kombination vorliegen müssen, um ein komplexes Organ hervorbringen zu können. Mit Hilfe seiner Verknüpfungsmaschine untersuchte er eine Population von Rampen, denen er es gestattete, sich frei zu entwickeln. Anhand von Testaufgaben konnte er die Fitness der Rampen überprüfen. Wie sich zeigte, entwickelten sich die Rampen in Hillis Experiment sprungweise. Als Hillis jedoch die Genotypen der einzelnen Individuen untersuchte, stellte er fest, dass diese sich auch in den stabilen Phasen, in denen sich ihre Fitnesswerte nicht änderten, ständig veränderten. Hillis fand mit Hilfe des Biologen Eric Lander heraus, dass Gene oder Genkombinationen,

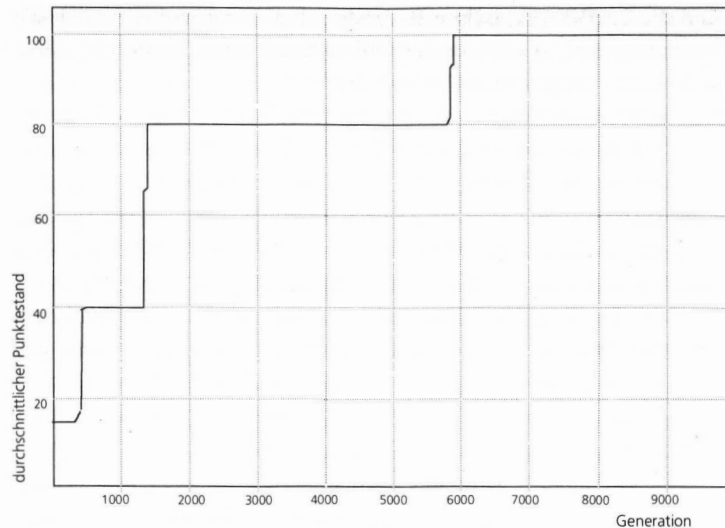


Abbildung 3.8: ([1], S.254) Entwicklung der Fitnesswerte der Rampen.

die Vorteile für eine gesamte Population mit sich bringen, sich erst langsam bis zu einem Prozentsatz in der Population verbreiten müssen, der $\frac{1}{e^2}$ entspricht ([1], S.255), bis sich die vorteilhafte Genkombination sprunghaft in der gesamten Population ausbreitet. Damit hatte Hillis in seiner Simulation eine andere Erklärung gefunden als die, welche Biologen bislang für das unterbrochene Gleichgewicht einer Population verantwortlich machten.

Erlerntes und robustes Verhalten

Ein Beispiel für ein echtes künstliches Lebewesen hat der amerikanische Wissenschaftler Randall D. Beer erschaffen, indem er ein der Amerikanischen Grossschabe *Periplaneta americana* nachempfundenenes künstliches Insekt namens *Periplaneta computatrix* konstruierte. Die Besonderheit dieses Lebewesens lag darin, dass sie über ein künstliches Nervensystem in Form eines neuronalen Netzes verfügte. Dieses Nervensystem war in der Lage, Eindrücke aus der Umwelt so zu verarbeiten, dass es das Verhalten der Kreatur steuern konnte. Dabei verwertete die Schabe zum Beispiel für ihre motorischen Fähigkeiten Rückmeldungen ihrer Beinen, um diese steuern zu lernen. Beer installierte eine Grundgangart, gab dem neuronalen Netz jedoch die Möglichkeit, dazuzulernen. Interessanterweise ergab sich bei einem seiner Versuche, dass die Schabe das Durchtrennen einiger Nerven, symbolisiert durch Trennen von Verbindungen im neuronalen Netz, überwand, indem es die zunächst beeinträchtigte Gangart schnell wieder stabilisierte ([1], S.362). Weiterhin konnte Beer durch das Erhöhen von Aktionspotentialen in bestimmten Knoten des Netzwerkes die Schabe in verschiedene Gangarten versetzen, die Beer ihr nicht beigebracht hatte, die aber alle von dem Entomologen E.O. Wilson als charakteristisch für natürliche Insekten bestimmt wurden. In einem nächsten Versuch setzte Beer die Schabe in einem Käfig aus, der eine Futterquelle und eine Barriere enthielt. Die Schabe orientierte sich in dieser für sie unbekanntem Situation, indem sie zunächst Kurs auf die Futterquelle nahm, dann an der Barriere entlanglief um zu deren Ende zu gelangen. Dort angekommen, drehte sie sich, um den Kontakt zur verlorenen Futterquelle wiederzusuchen und gelangte schliesslich an eine Aussenwand, der sie so lange folgte, bis die Futterquelle wieder in ihrem Witterungsbereich lag. Nach einem kurzen Moment des Zögerns folgte das künstliche Insekt seiner Witterung und nahm Kurs auf die Futterquelle.

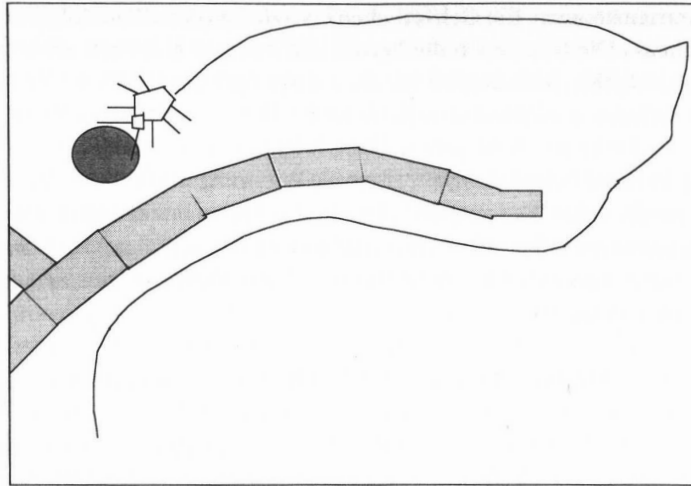


Abbildung 3.9: ([1], S.364) Schematische Darstellung des Weges, den *P.computatrix* im Verlauf von Beers Experiment zurücklegte.

Entscheidend an diesem Versuch ist nicht, dass die Schabe die Futterquelle wirklich erreichte, entscheidend war für Beer vielmehr, dass sich die Beobachter des Experimentes während dessen Verlauf mehrmals fragten: Was wird das Insekt wohl als nächstes tun?

3.3 Zusammenfassung

Weitere Beispiele aus dem Bereich des Artificial Life zu nennen, fiel nicht schwer. Inzwischen gibt es eine Menge von Wissenschaftlern, Studenten und Interessierten, die sich mit den Fragestellungen des AL beschäftigen. Fest steht, dass der Forschungszweig AL zwar immer noch unter Vorurteilen zu leiden hat, jedoch immer mehr Anerkennung genießt. Dies resultiert nicht zuletzt aus der Faszination, die für viele aus der Möglichkeit resultiert, die Mechanismen der Evolution näher untersuchen zu können. Auch heute kann man sicher nicht Ergebnisse von Experimenten im Computer mit natürlicher Evolution gleichsetzen, jedoch hat sich an vielen Beispielen gezeigt, dass AL brauchbare Ergebnisse liefert, die Vermutungen von Biologen erhärten oder auch widersprechen. AL ist in den Worten von Stuart Kauffman, einem weiteren Pionier im Bereich des AL : „...*der Weg herauszufinden, wie komplexe Systeme so etwas wie Selbstorganisation, Anpassung, Evolution, Koevolution, Stoffwechsel und all diese Dinge hervorbringen können. Es ist eine Nachahmung der Biologie, wenngleich Biologen das bisher noch nicht wissen.*“ ([1], S.155) Darüberhinaus gibt es AL-Projekte, die für sich in Anspruch nehmen, wirkliches künstliches Leben hervorzubringen. Ob dem nun so ist oder nicht, darüber mögen sich Philosophen streiten. Wer für sich eine Entscheidung in dieser Frage treffen will, verwendet am Besten eines der zahlreich kursierenden Software-Programme wie z.B. Tierra oder Avida und experimentiert selbst.

3.4 Literatur

[1] Levy, Steven (1993) *KL- Künstliches Leben aus dem Computer* Droemer Knaur, München.

- [2] Thro, Ellen (1994) *Künstliches Leben : eine spielerische Entdeckungsreise*
- [3] Boden, Margaret A. (1996) *The philosophy of artificial life*

Kapitel 4

Simulation in den Sozialwissenschaften

Stephan Mankart

17. Mai 2000

4.1 Überblick

In dieser Arbeit werden nach einer methodischen Einordnung von Simulation, der Vorstellung von Zielen, Besonderheiten und spezifischen Problemen sozialwissenschaftlicher Modellierung in den Kapiteln 1-6, in den Kapiteln 7-14, verschiedene Simulationsansätze vorgestellt. Die historische Entstehung der Ansätze ist in Abb. 1 auf Seite 5 festgehalten. Ziel der Arbeit ist es einen Überblick über wesentliche Aspekte sozialwissenschaftlicher Modellierung und gängige Simulationsansätze zu geben. Die Größe des Themengebiets führt dazu, dass im vorgegebenen Rahmen viele Aspekte und Details unerwähnt bleiben.

4.2 Methodische Einordnung

In einer Kritik an sozialwissenschaftlicher Simulation bezeichnet Ostrom Simulation als das 'Third Symbol System' (Ostrom 1988: 381). Diese Einordnung ist auch bei Troitzsch zu finden und räumt der Simulation eine eigenständige Position, neben dem 'First Symbol System' verbale Argumentation und dem 'Second Symbol System' Mathematik ein (Troitzsch 1997a: 46). Ostrom grenzt somit Simulation methodisch von der Mathematik ab. Dort werde Simulation zwar angewendet, etwa beim Auffinden von Lösungen für nichtlineare Gleichungssysteme, es hänge ihr aber der Ruf an, nur 'schlechte' Mathematik zu sein. Diese Abgrenzung erscheint leicht verständlich, in der Vergangenheit hat dieser Punkt innerhalb der Sozialwissenschaften aber öfters für Verwirrung gesorgt (vgl. Alkner 1974: 139). Nachvollziehen lässt sich diese Verwirrung, wenn man in Betracht zieht, dass ältere Simulationstechniken tatsächlich wenig anderes als die Weiterführung von Mathematik mit erweiterten (Rechner-) Kapazitäten darstellen. Im Hinblick auf neuere Ansätze trifft dies allerdings nicht zu (vgl. Kapitel 12 und 13). Stellt man die drei Symbol Systems einander gegenüber und untersucht ihre jeweiligen Vorzüge, bzw. Schwächen, wird ersichtlich, dass Simulation im Vergleich zu verbaler Argumentation aufgrund eines höheren Formalisierungsgrades den Vorteil besitzt, dass Annahmen und Theorien leichter auf Konsistenz zu überprüfen sind. Verglichen mit Mathematik bietet Simulation eine Erweiterung der Lösungsmöglichkeiten. In Fällen, in denen die analytischen Methoden der

Mathematik versagen, besteht mit Hilfe von Simulation die Möglichkeit weitere Lösungen zu finden. Nach Ostrom können die beiden ersten Symbol Systems in der Regel in das Dritte, also die Sprache der Simulation überführt werden. Wobei jedoch ausdrücklich nicht gilt, dass jede verbale Argumentation auch mathematisch abgebildet werden kann. Hinsichtlich des üblichen Methodenspektrums der Sozialwissenschaften bietet Simulation die Möglichkeit zwischen den beiden gängigen Ansätzen Induktion und Deduktion zu wechseln. Dies erleichtert die Überprüfung von Theorien und stellt somit eine fruchtbare Ergänzung dar.

4.3 Ziele sozialwissenschaftlicher Modellierung

Troitzsch formuliert als Ziel von Simulation die Vorhersage möglichen Systemverhaltens (Troitzsch 1997a: 47). Dieses Ziel differenziert er nochmals und kommt unter der Voraussetzung, dass das betrachtete Problem angemessen in ein Modell umgesetzt werden kann, schließlich zu folgenden drei Fragestellungen:

- Welche Arten von Verhalten können bei willkürlich gewählten Parametern und gegebenen Anfangsbedingung auftreten? (qualitatives Verhalten bei willkürlich gewählten Anfangsbedingungen)
- Welche Art von Verhalten wird das System in naher Zukunft aufweisen? (qualitatives Verhalten bei realitätsnah gewählten Anfangsbedingungen)
- Welchen Zustand wird das System in naher Zukunft annehmen? (quantitatives Verhalten)

4.3.1 Qualitatives Verhalten bei willkürlich gewählten Anfangsbedingungen

Bei dieser Betrachtung des qualitativen Verhaltens geht es darum herauszufinden welches Systemverhalten prinzipiell, bei willkürlich gewählten Parametern und Werten von Zustandsgrößen, auftreten kann. Dabei erweist sich Simulation als geeignetes Mittel, um einerseits das Übersehen von nicht offensichtlichen Zusammenhängen, wie es bei der Betrachtung von rein verbal repräsentierten Systemen geschehen kann, zu vermeiden und, um andererseits Lösungen zu errechnen, die auf analytischem Wege bei einer mathematischen Repräsentierung nicht zu finden sind. Beachten muss man hierbei allerdings, dass nicht gewährleistet ist, dass alle Lösungen mittels Simulation gefunden werden. Aus diesem Grund ist es erforderlich, Simulationsläufe mit möglichst allen Anfangswertkombinationen durchzuführen. Zusätzlich dazu sollte schon in diesem Zusammenhang eine Sensitivitätsanalyse des Systems durchgeführt werden, d.h. die Frage geklärt werden, ob Parameter existieren, die besonders großen Einfluss auf das Systemverhalten haben.

4.3.2 Qualitative Vorhersagen bei realitätsnah gewählten Anfangsbedingungen

Es werden realitätsnahe Anfangsbedingungen, von denen man annimmt, dass sie annähernd mit den in der Realität beobachtbaren Größen übereinstimmen gewählt und das

Systemverhalten in zahlreichen Simulationsläufen beobachtet. Untersucht wird insbesondere, ob Fixpunkte auftreten, gegebenenfalls welcher Art sie sind, ob chaotisches Verhalten auftritt und ob emergenter Strukturen wie Schichtbildung, Polarisierung oder Clusterbildung vorkommen.

Schwierigkeiten bereitet die Quantifizierung der Zustandsgrößen und Parametern i. d. R. bei analytisch lösbaren Modellen komplexer sozialer Phänomene, da einzelne Parameter in solchen Fällen zu viele verschiedene, oft schon einzeln schwer messbare Einflüsse wiedergeben müssten. Außerdem kommen im Zielsystem häufig Größen vor, die zu unterschiedlichen Zeiten verschiedene Werte annehmen, aber aus Gründen der Vereinfachung im Modell als konstant angenommen werden. Die an dieser Stelle eigentlich notwendige Erweiterung um eine Zustandsgröße, die den betreffenden Parameter variabel repräsentierte, würde wiederum zum Verlust der analytischen Lösbarkeit führen.

4.3.3 Quantitative Vorhersagen

Quantitative Vorhersagen erfordern ebenfalls die Quantifizierung der wichtigsten Parameterwerte und Zustandsgrößen. Von besonderem Interesse ist dabei die Frage nach der Steuerbarkeit des Systems durch bestimmte Parameter.

4.3.4 Modellzweck: Erklären vs. Vorhersagen

Der Modellzweck spielt eine wesentliche Rolle bei der Modellbildung, von der Entscheidung für eine Zielsetzung hängt vieles ab. Troitzsch setzt sich explizit mit zwei Zwecken auseinander, zum einen das Erklären und zum anderen die Vorhersage von Systemverhalten (Troitzsch 1997:6).

Erklärende Modelle sind i. d. R. einfach gehalten und es wird Wert darauf gelegt, die sozialen Prozesse des Zielsystems, nachzubilden. Durch Weglassen aller für diese Prozesse nicht relevanten Größen und Zusammenhänge werden relativ überschaubare Modelle erstellt, die sich dann zumeist gut untersuchen lassen. Mit Hilfe solcher Modelle werden sog. 'would-be-worlds' erklärt (vgl. Casti 1996).

Besteht der Modellzweck darin quantitative Vorhersagen zu treffen, so wird versucht möglichst alle, das System beeinflussende Größen in das Modell miteinzubeziehen, was zwangsläufig zu komplexeren Modellen führt. Bei diesen Modellen wird von den Prozessen des Zielsystems insofern abgewichen, als einzig der Output des Modells gleich der in der Realität beobachtbaren Werte sein muss und der Weg, auf dem diese Größen innerhalb des Modells erzeugt wurden, nicht zwingend mit den realen Prozessen übereinstimmen muss. Grundsätzlich führt der Entstehungsprozess von Vorhersage-Modellen, über die Erklär-Modelle, da selbstverständlich auch für die Vorhersage Verständnis für die Wirkungszusammenhänge zwischen den Zustandsgrößen des Modells nötig ist. Dieses Verständnis spiegelt sich auch in einem anderen, der Modellbildung innewohnenden Aspekt wider, der von Troitzsch als 'theory building' bezeichnet (Troitzsch 1997: 8). Die Erarbeitung eines einen Wirklichkeitsausschnitt repräsentierenden Modells erfolgt schrittweise, angefangen bei einer recht simplen Näherung bis hin zu einer vielleicht sogar vollständigen Erfassung der Realität. Die theoretischen Kenntnisse über den Realitätsausschnitt verfeinert sich entsprechend schrittweise.

4.4 Eine Besonderheit sozialwissenschaftlicher Modellierung- Artificial Societies

Ein Spezifikum sozialwissenschaftlicher Simulation ist die Beschäftigung mit sog. Artificial Societies. Hierbei wird nicht versucht, möglichst exakt die Realität menschlicher Gesellschaften abzubilden, sondern man löst sich von diesen Gegebenheiten und beschäftigt sich allgemein mit Gesellschaften interagierender Individuen. Ziel hierbei ist die Erforschung allgemeingültiger Regeln, um die Ergebnisse mit Phänomenen in menschlichen Gesellschaften zu vergleichen, und dadurch zu einem tieferen Verständnis der Zusammenhänge innerhalb Gesellschaften zu kommen. Beispiele hierfür sind Studien über Auswirkungen von eingeschränkter Kommunikation auf Märkten (vgl. Alvin und Foley 1992) oder Studien über Gesellschaften in denen alle Individuen perfektes Wissen über die Zukunft haben (vgl. Doran 1998). In letzterer Untersuchung wurde insbesondere versucht eine Antwort auf die Frage zu finden, ob Individuen trotz ihres Wissens noch über Entscheidungsspielraum bei der Wahl ihres Verhalten verfügen.

4.5 Probleme

Grundsätzliche Probleme mit denen sozialwissenschaftliche Simulation zu kämpfen hat, sind die große Komplexität sozialer System, der Mangel an Daten und die Schwierigkeiten bei der Datengewinnung. Letztere ist i.d. R. aufwendig und teuer. Außerdem sind Umfragedaten, häufig sehr schwierig auf ihre Aussagekraft zu überprüfen, was die Arbeit mit ihnen erschwert. In den wenigen Gebieten sozialwissenschaftlicher Forschung, in denen ausreichend Daten zur Verfügung stehen, wie beispielsweise der Demographie, der Wahlforschung und der Makroökonomie, ist die Verwendung von Modellen aber durchaus gebräuchlich.

4.6 Geschichtlicher Überblick

Erste vereinzelte Versuche Simulation die den Sozialwissenschaften einzusetzen gab es schon in den 60iger Jahren. In den 70igern sorgte dann der System dynamics-Ansatz für ein erhöhtes Interesse an dieser Methode, Weltmodelle wie die von Forrester, die im Zusammenhang mit der Club of Rome- Studie 'Die Grenzen des Wachstums' entstanden sind, wurden einer breiten Öffentlichkeit bekannt. Als sich aber bald darauf die Vorhersagen dieser Modelle als unhaltbar erwiesen, führte dies auch zu einem etwas abebbenden Interesse an jenen. Weitere aus dieser Zeit stammende Ansätze, die sich bis heute gehalten haben sind die Queuing Modells (vgl. Kapitel 10) und die Mikroanalytische Simulation(vgl. Kapitel 9). Neue Impulse erhielt die sozialwissenschaftliche Modellierung dann erst wieder Anfang der 90iger, als damit begonnen wurde, Ansätze aus anderen Wissenschaftsbereichen auf sozialwissenschaftliche Fragestellungen zu übertragen. Dies trifft auf das Konzept der Zellulären Automaten und insbesondere auf den aus der K.I.-Forschung stammenden Ansatz der agentenbasierten Modellierung zu.

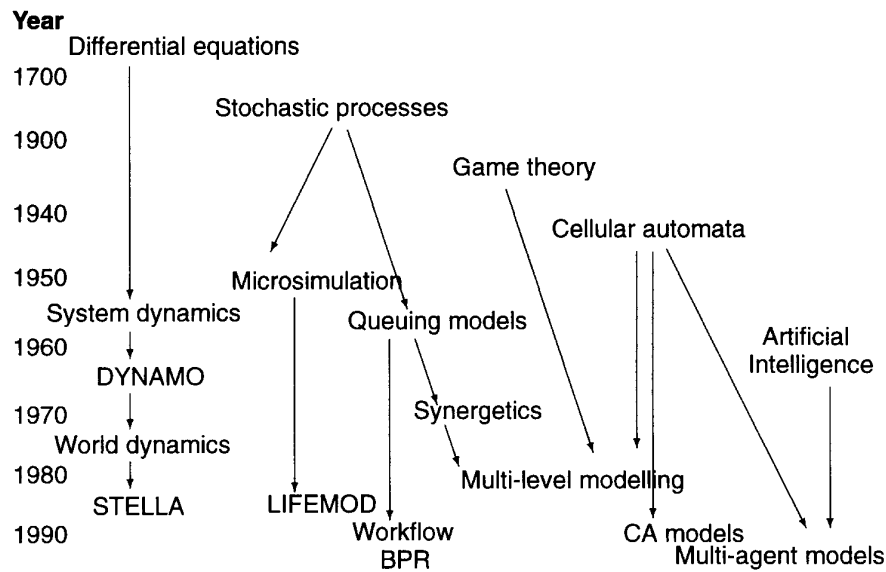


Abbildung 4.1: Geschichtliche Entwicklung der Ansätze (entnommen aus Gilbert/Troitzsch, 1999: 7)

4.7 Simulationsansätze

In den folgenden Kapiteln werden die in den Sozialwissenschaften gängigen Simulationsansätze kurz vorgestellt, wobei auch auf Grenzen und Probleme der Ansätze hingewiesen wird. Abschließend sind nochmals markante Eigenschaften der einzelnen Ansätze übersichtlich in einer Tabelle festgehalten.

4.8 System dynamics

Der System dynamics-Ansatz hat seine Ursprünge in der Behandlung von Differentialgleichungen (DGL). Durch Zustandsvariablen und Raten wird das Zielsystem unteilbar auf Makroebene beschrieben. Dabei werden zur Zustandsberechnung anstelle von kontinuierlichen DGLs zeitdiskrete Differenzgleichungen benutzt. Durch die Wahl von sehr kleinen Zeitschritten kann aber ein 'quasi-kontinuierliches' Modellverhalten erreicht werden. Ein wichtiger Unterschied bei der Zustandsberechnung zwischen Differenzgleichungen und DGLs ist die Zulassung von unstetigen Funktionen bei den Differenzgleichungen. Der System dynamics-Ansatz verfügt über eine umfassende, intuitive Symbol-Sprache, die bei Nutzung einer der zahlreichen Softwareumgebungen zu einer recht übersichtlichen und leicht verständlichen Anwendung führt (Bsp. für Software: Powersim, Dynamo, Stella...). Beispiel für System dynamics- Modelle, die in Osnabrück im Grundstudium behandelt wurden, sind das Weltmodell aus der Einführung in die Systemwissenschaft und das Hawks, doves and law-arbiters-Modell nach Coll, bzw. die erweiterte Version nach Troitzsch, das im Rahmen der Vorlesung Umweltsysteme vorgestellt wurde (Gilbert/Troitzsch 1999: 31).

Als Schwäche des System dynamics- Ansatz muss das Übergehen der Struktur sozialer Systeme, die i. d. R. aus mehreren Ebenen bestehen, angesehen werden, bzw. die Unmöglichkeit der Einbeziehung mehrerer Ebenen. Außerdem sind die erstellten Modelle

im Allgemeinen unteilbar und eine konzeptionelle Veränderung eines Teiles des Modells erzwingt eine Anpassung des gesamten Modells.

4.9 Mikroanalytische Simulation

Dieser Ansatz berücksichtigt die Existenz verschiedener Betrachtungsebenen, Individuen und eine höher aggregierte Ebene können unabhängig von einander betrachtet werden, was sozialen Strukturen und Prozessen näher kommt, als die 'eindimensionale' Sichtweise des System dynamics-Ansatz. Ziel des mikroanalytischen Ansatzes ist insbesondere eine Vorhersage des quantitativen Systemverhaltens.

Es werden zunächst Wirklichkeitsausschnitte, die von Interesse sind, ausgewählt und mit Hilfe von repräsentativen Klassen und Eigenschaften, die für den Modellzweck relevant sind, auf Mikroebene abgebildet. Damit erhält man ein sog. microdata file, welches die Grundlage der Simulation bildet und neben den Klassen und Eigenschaften auch eine Gewichtung dieser nach dem Vorkommen im Zielsystem enthält. Bei der Simulation wird für jede Klasse einzeln der Zustand zum Folgezeitpunkt errechnet. Anschließend werden die Ergebnisse anhand der Gewichtungen auf die Makroebene hochgerechnet. Ein Beispiel aus der Finanzpolitik, die neben der Sozialpolitik das Hauptanwendungsgebiet des mikroanalytischen Ansatzes darstellt, ist ein einfaches Steuermodell, mit dem die Auswirkungen einer gestaffelten Steuererhöhungen auf das Gesamtaufkommen vorhergesagt werden sollen. Die Klassen des Modells bestehen in diesem Fall aus den Einkommensgruppen, zusätzliche, für die Simulation wichtige Eigenschaften sind Informationen wie Familienstand, Anzahl Kinder,... . Für die einzelnen Einkommensgruppen wird nun das Steueraufkommen errechnet und anschließend unter Berücksichtigung der Gewichtung (Vorkommen der Gruppen in der Bevölkerung) das Gesamtsteueraufkommen bestimmt. Eine bereits erwähnte, grundsätzliche Schwierigkeit, die bei der mikroanalytischen Simulation besonders zu Tage tritt, ist die starke Datenabhängigkeit. Ein weiteres Defizit ist, das Fehlen von Wechselwirkungen zwischen den Ebenen, was den Einsatz der Modelle auf eine kurzfristige Analyse beschränkt. So hat beispielsweise die Entwicklung des Gesamtsteueraufkommens bei obig beschriebenem Modell keine Auswirkungen auf die Steuersätze, was natürlich eine sehr unrealitätsferne Annahme darstellt.

4.10 Queuing models

Queuing models werden zur Beschreibung von, in Events gegliederten, Abläufen benutzt. Sie bestehen aus zumindest drei Objekten: Customern, Servern, Queues. Zusätzlich gibt es noch eine sogenannte Agenda auf der die zeitliche Abfolge der Events festgehalten sind. Mit Hilfe von Abbildung 2, die ein einfaches Queuing model zeigt, kann man sich den Aufbau verdeutlichen.

Der generelle Simulationsablauf sieht wie folgt aus: Customers reihen sich vor einem Server in eine Queue ein, die sie nach einer gewissen, ev. zufällig erzeugten Wartezeit verlassen. Danach werden sie in einer gewissen, ev. zufällig erzeugte Zeit von einem Server bedient. Schließlich wurde der Customer abgefertigt und kann sich in eine weiteren Queue einreihen. Das Eintreten von Customern in diesen Prozess kann ebenfalls stochastisch bestimmt sein. Es wird diskrete, eventorientierte Zeit verwendet, pro Zeitschritt werden nur einige Zustandsvariablen des Modells verändert.

Queuing models dienen zur Ablaufbeschreibung und -optimierung in Umgebungen mit mehreren Schaltern, bzw. Servern, wie z.B. Flughäfen, Behörden, usw. . Dazu gehört auch das Auffinden von mögliche Engpässen und Versorgungslücken. Schwächen des Ansatzes liegen im 'starren' Verhalten der Customer, adaptive Anpassungen an die Umgebung sind nämlich nicht vorgesehen. So verteilen sich die Customer beispielsweise nicht selbstständig auf mehrere äquivalente Schalter.

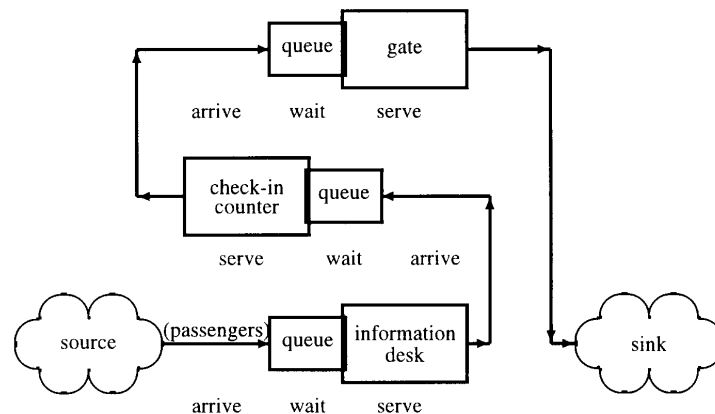


Abbildung 4.2: Einfaches Queuing model (entnommen aus Gilbert/Troitsch, 1999: 78)

4.11 Mehrebenenmodelle

Bei der Mehrebenenmodellierung wird versucht, die Entstehung neuer Strukturen, die auf Makroebene allein nicht erklärt werden kann, durch nichtlineare stochastische Prozessen nachzubilden. Ziel ist somit die Vorhersage des qualitativen Systemverhaltens.

Bei Mehrebenenmodellen werden Wechselwirkungen zwischen Ebenen zugelassen, d.h. aggregierte Ebenen wirken auf darunterliegende Ebenen ein. Zudem ist direkte Interaktion zwischen den Individuen einer Ebene möglich. Um Rückkoppelungsschleifen zu vermeiden werden i.d. R. jeweils abwechselnd Schritte, in denen die aggregierten Ebenen aus den niederen Ebenen berechnet werden, bzw. umgekehrt, durchgeführt.

Ein einfaches Mehrebenenmodell ist beispielsweise das von Troitzsch vorgestellte Drei-Ebenen-Modell zur Überwindung der Geschlechtertrennung in Lehrerkollegien (Troitzsch 1999b: 326). Mit Hilfe dieses Modells kann die Entwicklung des Frauenanteils in den rheinland-pfälzischen Lehrerkollegien in den Jahren 1950- 1990 simuliert werden (vgl. Abb. 3). Auf der obersten Ebene steht dabei das Landesschulsystem, auf der mittleren einzelne Schulen und auf der untersten die einzelnen Lehrerinnen und Lehrer.

Das Zulassen von Rückkoppelungen zwischen den Ebenen und Interaktion zwischen den Individuen ist ein großer Vorteil der Mehrebenenmodellierung gegenüber der mikroanalytischen Simulation. Jedoch sind diese Möglichkeit noch sehr eingeschränkt, es ist nur eine einzige Art der Interaktion vorgesehen. Neuere Ansätze gehen weit darüber hinaus und bieten weitaus vielfältigere Möglichkeiten der Interaktion (vgl. Kapitel 13).

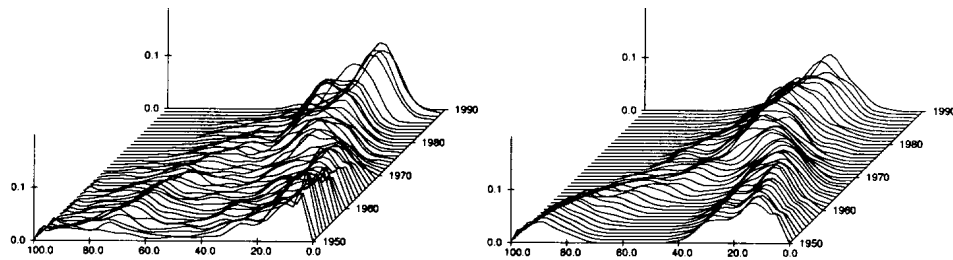


Abbildung 4.3: Verteilung der Frauenanteile in den Lehrerkollegien von 150 rheinland-pfälzischen Gymnasien von 1950-1990. Links: empirische Daten. Rechts: Drei-Ebenen-Simulation (Lehrer - Schulen - Land) (entnommen aus Gilbert/Troitsch, 1999: 7)

4.12 Zelluläre Automaten

Durch die Zellulären Automaten kommt der bisher außer Acht gelassene Aspekt der räumlichen Verteilung ins Spiel. Auf einem ein-, zwei oder dreidimensionalen diskreten Gitter sind Zellen, die diskrete Zustände annehmen, verteilt. Die Zustandsänderungen der Zellen wird mittels vorgegebener, ev. stochastischer Regeln in diskreten Zeitschritten berechnet. In einfachen Zellulären Automaten findet Beeinflussung des Zustandes, bzw. Interaktion nur zwischen benachbarten Zellen des Gitters statt, wobei Nachbarschaft unterschiedlich definiert sein kann.

Mit Hilfe eines sehr einfachen Zellulären Automaten kann beispielsweise die Ausbreitung von Gerüchten simuliert werden. Dabei hängt der Zustand der Zelle, der entweder besagt, dass die Zelle das Gerücht kennt oder, dass die Zelle das Gerücht noch nicht kennt, von den Zellen in der Nachbarschaft ab. Eine einzige informierte Zelle in der Nachbarschaft genügt, um den Zustand der Zentralzelle zu verändern, falls diese das Gerücht noch nicht kennt. Erweiterungen in Form einer Abhängigkeit des Zustandes der Zentralzelle von einer größeren Umgebung oder von allen Zellen des Zellulären Automaten sind möglich. Ein Beispiel hierfür ist ein Immigrationsmodell, in dem die Zustandsberechnung einer Zelle von der Summe der Nachbarzellen abhängt. Zusätzlich sind in diesem Modell auch Bewegungen der Individuen im Modellraum möglich, womit auch der Fall auftreten kann, dass sich mehrere Individuen auf einer Zelle befinden, was durch gesonderte Zustände repräsentiert wird.

Mit Zellulären Automaten lassen sich schnell eindrucksvolle Simulationen erstellen, schwierig wird es allerdings, wenn es an die Auswertung der Ergebnisse geht. Die erzeugten Resultate lassen sich nur schwer analysieren und interpretieren. Es ist ausgesprochen schwierig, die Resultate zu verwerten, um mit ihrer Hilfe konkreten Vorgaben für die simulierten Systeme zu erstellen (vgl. Vortrag von Sabine Dormann vom 5. Juni '00).

4.13 Multi-Agenten-Modelle

Dieser sehr junge Ansatz hat seine Ursprünge in der K.I.-Forschung, insbesondere im Bereich der Verteilten K.I., welche sich beispielweise mit der Fragestellung auseinandersetzt, wie eine Gruppe von Individuen (Agenten), die alle verschiedene Fähigkeiten besitzen, konzipiert werden muss, damit die Gruppe in der Lage ist, gemeinsam an der Lösung eines Problems zu arbeiten, welches für einzelne Individuen jeweils unlösbar ist. Ziel des

M.-A.-Ansatzes ist stark vereinfacht gesprochen die Erschaffung von Programmen, die intelligent mit ihrer Umgebung interagieren.

Eine einheitliche Definition des Agentenbegriffs existiert nicht, eine von Huhns und Singh stammende Annäherung findet man bei Troitzsch. Agenten sind demnach: 'self-contained programs which can control their own actions based on their perceptions of their operating environment' (Huhns and Singh 1998, entnommen aus Troitzsch 1998: 159). Dabei ist das Verhalten der Agenten durch Regeln bestimmt. Beim Einsatz von M.A.-Modellen in den Sozialwissenschaften steht i. d. R. die qualitative Vorhersage von Systemverhalten im Vordergrund.

Agenten können sowohl einzeln als auch in Gruppen auftreten, einfache Beispiele sind Suchagenten, die das world wide web nach bestimmten Informationen durchforsten und ihre Resultate rückmelden, oder kleine Marktmodelle in der Ökonomie, die das Verhalten von Käufern und Verkäufern auf Märkten simulieren (zur Zeit arbeitet Franz Beckenbach an der Universität Kassel an solchen Modelle).

Auf menschliches Handeln übertragen, wird das Agency-Konzept meist dazu genutzt, um dessen zweckgerichtete Natur auszudrücken. Somit steht es in Beziehung zu Konzepten, wie Absicht, freier Wille und Intentionalität. Außerdem ist man bemüht Agenten Eigenschaften wie Überzeugungen, Bedürfnisse, Motive und Emotionen beizubringen. Wobei es selbstverständlich schwierig ist, diese Konzepte und Eigenschaften in Zusammenhang mit Computerprogrammen zu bringen. Man muss sich darüber im klaren sein, dass Agenten nicht über diese Eigenschaften verfügen, sondern man lediglich bemüht ist, diese zu simulieren.

4.13.1 Angestrebte Eigenschaften von Agenten

(vgl.: Gilbert, N /Troitzsch, K. G. 1999a:161)

- Wissen und Überzeugungen
Handlungen der Agenten basieren auf Wissen über ihre Umgebung, dabei können einige Informationen die Agenten besitzen durchaus falsch oder fehlerhaft sein, in solchen Fällen wird von Überzeugungen gesprochen.
- Schlussfolgerungen
Aus einer Menge von Überzeugungen können Agenten Folgerungen ableiten. Bsp.: Agent A ist überzeugt davon, dass Agent B gerade gegessen hat, daraus folgert er, dass es in der Nähe von B 'Nahrung' gibt.
- Soziale Modelle
Agenten können in der Lage sein, etwas über die Interaktionsstruktur ihrer Umgebung zu lernen (z.B. wer mit wem kommuniziert) und sich dadurch ein Bild ihrer Umgebung erschaffen.
- Ziele
Agenten können Ziele und Subziele zu verfolgen. Beispiel: Ziel: Überleben, Subziele: Energie aufnehmen, lebensgefährlichen Bedrohungen auszuweichen.
- Planen
Agenten müssen um Ziele erreichen zu können in der Lage sein, ihr Verhalten über mehrere Schritte hinweg zu planen.

- Sprache
Agenten können mit ihrer Umgebung und untereinander direkt oder indirekt kommunizieren.

Nicht alle Agenten verfügen über die selben Fähigkeiten, nach Troitzsch kann man sie in die drei folgenden Gruppen einteilen, wobei aber allen gemeinsam ist, dass sie autonom agieren und ihr Verhalten nicht direkt gesteuert wird. Dabei gilt, dass bei dieser Einteilung die letztgenannten Gruppen jeweils auch über die Fähigkeiten der vorgenannten verfügen (vgl.: Gilbert, N / Troitzsch, K.G. 1999: 160).

- Reactive agents
Reagieren auf Nachrichten ihrer Umgebung durch Aktualisierung ihres Zustandes und durch Verschicken von Nachrichten an andere Agenten.
- Intentional agents
Wenden 'Metarules' an und sind in der Lage Ziele zu formulieren, beispielsweise abhängig von ihrer Motivation oder ihren Bedürfnissen, Prioritäten zu setzen, wenn sich Ziel nicht miteinander vereinbaren lassen und Pläne aufzustellen um Ziele zu erreichen.
- Social agents
Verfügen über Modelle von anderen Agenten und der Kommunikationsstruktur zwischen Agenten und können somit andere Agenten mit in ihre Planung einbeziehen.

Als Schwäche des Agenten-Ansatzes muss man die bereits bei den Zellulären Automaten erwähnten und auf alle regelbasierten Ansätze zutreffenden Probleme bei der Auswertung der Ergebnisse ansehen. Dennoch ist der Agenten-Ansatz insgesamt sehr vielversprechend und die Grenzen seiner Anwendbarkeit scheinen noch nicht klar abzusehen zu sein.

4.14 Ausblick

Man kann sicherlich behaupten, dass Simulation sich auch in den Sozialwissenschaften etabliert hat, dies gilt insbesondere für die oben schon angesprochenen Bereiche Demographie, Finanz- und Sozialpolitik. Dabei werden zumeist die 'klassischen' Ansätze aus den Kapiteln 8-11 angewendet. Durch die neueren Ansätze, welche die alten in sich aufnehmen, gleichzeitig aber weit über sie hinaus gehen, scheinen zudem neue Anwendungsbereiche offenzustehen. So können beispielsweise mit Hilfe des Agenten-Ansatzes alle Modelle die mit Ansätzen aus den Kapiteln erstellt wurden realisiert werden (vgl. Troitzsch 1997b). Zudem ergeben sich durch die explizite Simulation von kognitiven Prozessen und Kommunikation zwischen den Individuen, die bei diesem Ansatz in Mittelpunkt stehen, vielfältige neue Möglichkeiten. So kann über die Simulation von Strukturen sozialer Ordnung hinausgegangen werden und auch die Frage wie soziales Verhalten entsteht aufgegriffen werden. (Troitzsch 1999b: 337)

4.15 Literatur

4.15.1 Verwendete Literatur

- Gilbert, N und Troitzsch, Klaus G. (1999) Simulation for the Social Scientist. Open University Press.

- Troitzsch, Klaus G. (1990) Modellbildung und Simulation in den Sozialwissenschaften. Westdeutscher Verlag, Opladen.
- Troitzsch, Klaus G. (1999) Dynamische Modelle komplexer sozialer Systeme: Was leisten Computersimulationen? In Klaus Mainzer, Hrsg.: Komplexe Systeme in Natur und Gesellschaft. Komplexitätsforschung in Deutschland auf dem Weg ins nächste Jahrhundert, Springer, Berlin, Seite 321-338.
- Troitzsch, Klaus G. (1997a) Social Science Simulation– Origins, Prospects, Purposes, in: Rosaria Conte, Rainer Hegselmann, Pietro Terna, eds.: Simulating Social Phenomena, Springer, Berlin, pp. 41-54.
- Troitzsch, Klaus G. (1997b) Can Agents Cover All the World?, in: Rosaria Conte, Rainer Hegselmann, Pietro Terna, eds.: Simulating Social Phenomena, Springer, Berlin, pp. 55-72.

4.15.2 Angeführte Literatur

- Alkner, Jr, Hayward R. (1974) Computer simualations: Inelegant mathematics and worse social science? International Journal of Mathematical Education in Science and Technology, 5: 139-155.
- Alvin, P und Foley, D. Decentralized, dispersed exchange without an auctioneer. Journal of Economic Behaviour and Organization, 18: 27-57.
- Casti, John L. (1996) Would-Be Worlds. How Simulation Is Changing the Frontiers of Science. Wiley, New York.
- Doran, J.E. (1997) Foreknowledge in artifical societies. In R. Conte, R. Hegselmann und P. Terna, Simulating Social Phenomena, Springer, Berlin, pp. 457-470.
- Huhns, M. und Singh, M.P. (1998) Reading in Agents. Morgan Kaufmann, San Mateo, CA, 1998.
- Ostrom, Thomas. Computer simulation: The third symbol system. Journal of Experimental Social Pschology, 24: 381-392.

Kapitel 5

Synergetics

Matthias Peußner

24. Mai 2000

5.1 Introduction

Synergetics is the science of cooperation and interaction. The science was introduced by *Hermann Haken*, Institute for theoretical physics at the university of Stuttgart, Germany. Hermann Haken, the father of Synergetics, first mentioned the word during a lecture in 1969. The aim of Synergetics is to find some principles and laws in different branches of science. This helps to explain or understand several phenomena in our environment. The phenomena can be observed in the so called self-organized systems. So Haken developed a new concept and new methods.

5.2 One Example of Synergetics

So first I will show some examples of Synergetic Systems. The first example is fluid dynamics (see fig. 5.1).

In these five pictures you see a flow of fluid round a cylinder. The first one shows the flow of fluid at a very low speed. Here nothing interesting can be found. At a few higher speed, behind the cylinder two circles appear. If we increases the speed once more, we notice a moving pattern. These structures seems to appear suddenly, directed by a magic hand. The next picture shows us fluctuations and chaos. Final, at a very high speed we will only recognize undefined chaos.

5.3 Macroscopic structures

So we can divide macroscopic structures into some categories:

- Man made structuers: Such as an engine of a car, or the structures of new, especially American, cities.
- Structures in systems in thermal equilibrium (this word will be explained later), like a crystal.

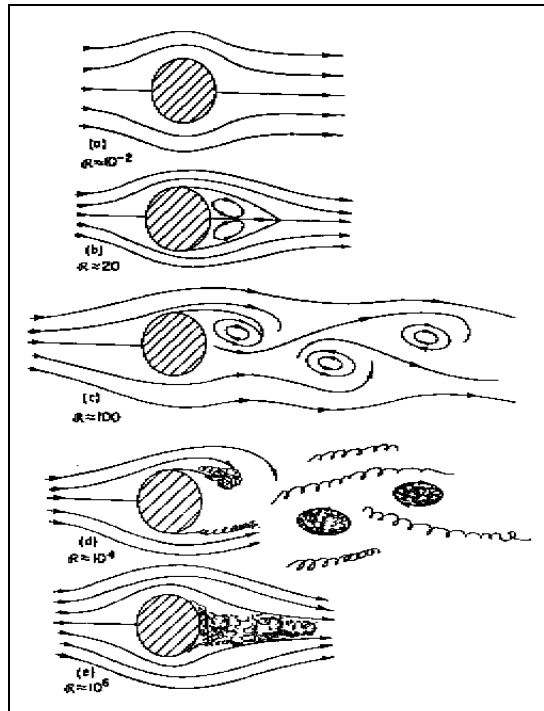


Abbildung 5.1: Fluid dynamics around a cylinder (From [2]).

- Self-organized Systems: (see the following figures from 5.2 to 5.8).

Other examples of self-organized systems

- Belousov-Zhabotinski-reaction
- 22 years period of the sun spots
- the magnetic field of the earth
- decision-making in the society
- our brain (really complex one)
- a forest
- The paradigm for Synergetic Systems is the **laser**. At this system Hermann Haken developed his new concept.

Hermann Haken classifies all systems into three classes:

1. closed systems in thermal equilibrium
2. systems near to thermal equilibrium
3. systems far from thermal equilibrium

In the first class are systems with no input nor output of energy or matter. These systems are treated by the Thermodynamic. The systems in the next class can be treated with methods of the Irreversible Thermodynamic. In the third class there are system I am talking about. Later we will see some fundamental differences between these classes.



Abbildung 5.2: Sand dune (From [1]).

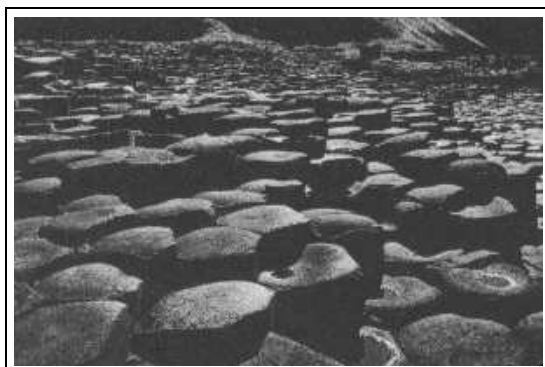


Abbildung 5.3: Basalt pillars (From [1]).

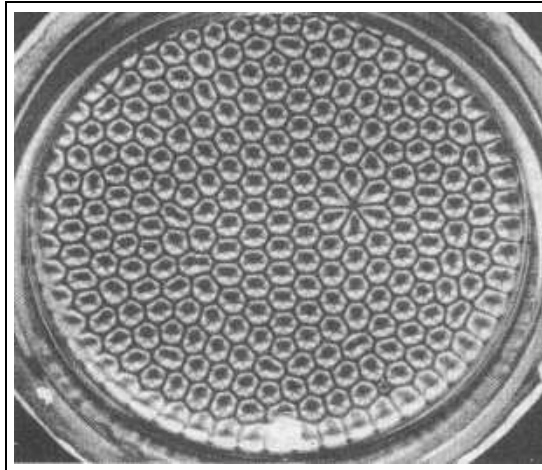


Abbildung 5.4: Bénard instability (From [1]).



Abbildung 5.5: Karman's whirl street (From [1]).

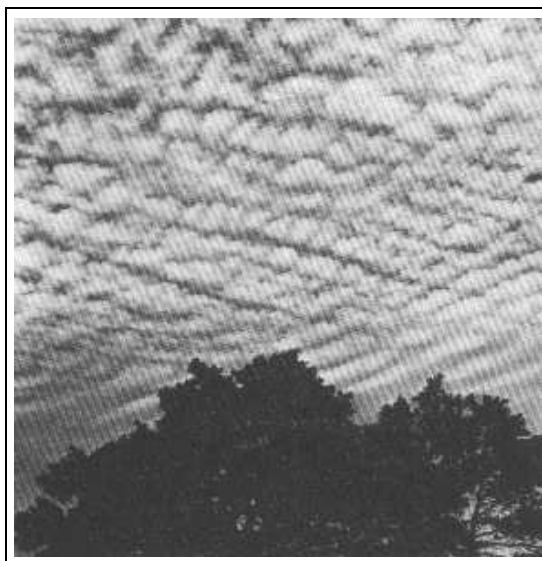


Abbildung 5.6: Cloud streets (From [1]).

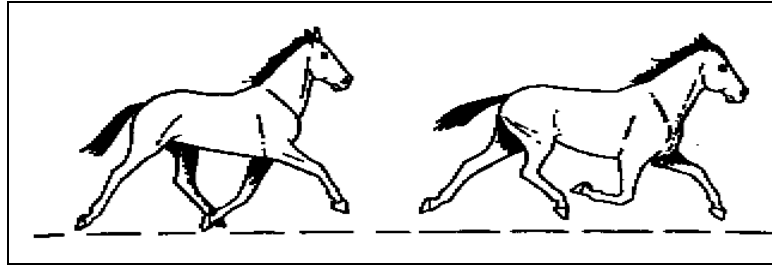


Abbildung 5.7: Walk of a horse. Here you see a moving pattern, which will change at specified speeds. (From [1]).



Abbildung 5.8: Very large structures in the atmosphere of the Jupiter (From [1]).

5.4 Thermodynamic

First we have to know about thermal equilibrium. A system is in thermal equilibrium when we have at every place the same temperature, we have a very macroscopic homogeneous state. For example when two gases are totally mixed up (see fig. 5.9), the system is in thermal equilibrium.

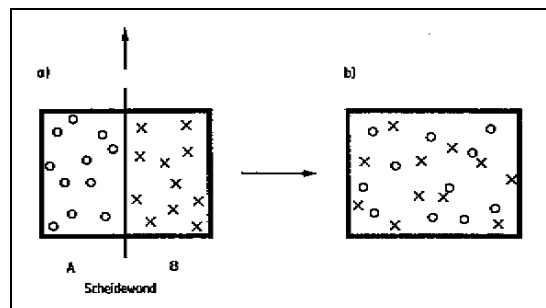


Abbildung 5.9: a) two gases are divided by a wall; b) without a wall the gases are totally mixed up (From [1]).

In Thermodynamic we treat several macroscopic properties, like the temperature, pressure, amount of heat and the entropy. The Thermodynamic is based on two theories or assumptions. These theories cannot be proved, they are only recognized .

The First Law: This law is the thesis of conservation of energy. That means, that the total amount of energy in a closed system, is always the same.

The Second Law says, that in a closed system, the change of entropy is always positive.

The entropy is a measure for the degree of disorder. A system with a higher entropy is more disordered.

5.4.1 The Second law

There are some more descriptions of the Second Law, which can be more easily imagined.

- Heat cannot go by its own from a body of lower temperature to a body of a higher temperature (see also fig. 5.10).
- All natural processes are irreversible, that means, we cannot get exactly a previous state of a system without work.
- The change of entropy in a closed system is always positive.
- In a closed system the entropy always increases to its maximal value, at its maximal value, the system is in thermal equilibrium (see fig. 5.11).

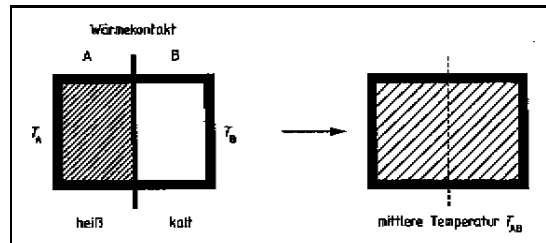


Abbildung 5.10: on the left side we have still two different temperatures; after some time we will get one middle temperature for the hole box (From [1]).

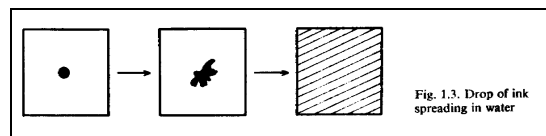


Abbildung 5.11: In a closed system the entropy always increases to its maximal value (From [2]).

5.4.2 Contradiction?

So there seems to be a contradiction between the two laws of Thermodynamic and the Synergetic systems which form stable macroscopic structures and so are far away from thermal equilibrium.

Many scientists thought that there is contradiction, and they tried to find other concepts like the one of negative entropy. But Haken showed that the structure forming systems are not in contradiction to the Second Law. The methods of the Thermodynamic and the Irreversible Thermodynamic are at their limit, and therefore we need a new allegory to examine the systems of the 3. class. So that is why Haken introduced the Synergetics.

5.5 Synergetic Systems

Here are important properties of such systems:

- they are open Systems and have a permanent flux of energy or matter
- structures only appear inside of narrow boundaries for the energy, that means, we do not have to put too few nor too much energy in the system (please compare to the first example)
- they are far from thermal equilibrium
- they consists of many manyparticles (elements/subsystems)
- there is correlation between the elements

5.5.1 Definition of (new) words

Phase transition

A phase transition is e.g. in analogy to systems in thermal equilibrium the changing point of the aggregate states e.g. of water. At a temperature of 100 °C liquid water suddenly changes into gas. Or a ferro magnet loses his property as a magnet at a specified (higher) temperature (see fig. 5.12). All this examples show a dramatic changes of the macroscopic properties. These critical points, the phase transition, are observed in self-organized systems.

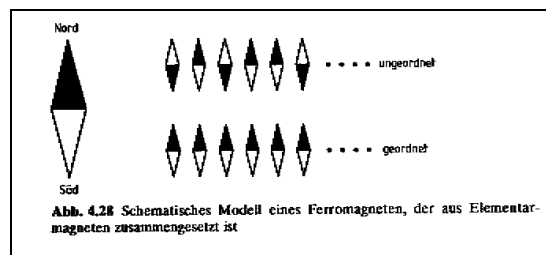


Abbildung 5.12: A ferro magnet, ordered and disordered (From [1]).

Order parameter (German: Ordner)

The order parameter is the answer of a question: What is the demon that tells the subsystems to behave in such a well organized manner? So the order parameter governs the behavior of the system's state. It seems to be a magic force, which suddenly appears and can disappears quickly.

Slaving principle (German: Versklavungsprinzip)

So we can say, that the subsystems are a slave of the order parameter. Haken calls this the slaving principle. Later on, we will see a mathematical definition and the power of the slaving principle.

Self-organization

It means, that the order parameter is a part of the system, it do not come from outside. The order parameter rules the elements, but he only exists because of the elements. So we can say, that these systems are self-organized.

Symmetry breaking, fluctuations

On the basis of the figure 5.13 I will explain what symmetry breaking means. The white ball is exactly in the middle on the top of the hill. At this position it has to decide either it falls down on the left side or the right side. This decision is made by fluctuations which are not predictable. There the microscpic chaos has an enormous effect on the macroscopic appearance.

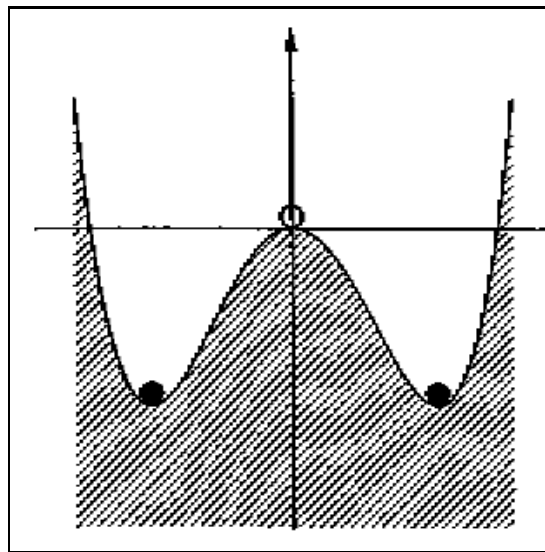


Abbildung 5.13: Symmetry breaking (From [1]).

Control parameters

The control parameters are given parameters, for e.g. the viscosity, capability of heat conduction, the borders of the environment, ... Do not mix up the control parameters with the order parameter. They are quite different. Normally we can measure the control parameters.

Mode (German: Mode)

The modes are the possible kind of appearances of the system. For e.g. in the first example the state of forming moving patterns is one mode, the state of chaos is another one.

5.5.2 The concept

We have possibilities to approach our problem. We use the first one, if we know the subsystems very well, and the other, if we do not know much about the subsystems, what are the elements in our system. I will concentrate on the first way. So we know the

subsystems, and therefore we can form the so called evolution equations. These equation represents the total state of the system.

$$\frac{dU}{dt} = N(U, \nabla, \sigma) + F$$

U is the state of the system, ∇ is a vector of place, because our system is not homogeneous. And the σ is the variable for the control parameters, because we have an open system. The variable F stays for the Fluctuations. To form this equation is the task of the special branch of science. But to represent the hole state of even one simple system, we get very much equations. E.g. the laser, we get around 10^{18} differential equations. This is really too much to solve, even if we use computers. So we have to reduce our point of view. Now we are only looking at a special part of the space state, namely the attractors (saddle points, focus, limit cycles,..). We also leave out fluctuations and the inhomogeneous space. And we are generating a linear equation by looking at little deviations $U = U_0 + q$. So we get a d-dimensional equation

$$\frac{dq}{dt} = N_L(q)$$

. To simplify this equation once more, we make a coordinate transformation. After that, we get Eigenvectors λ_i and Eigenvalues O_j .

$$\frac{d\xi}{dt} = \lambda_j \xi_j$$

So we get more less equations, but still too much to solve. Note that the Eigenvalue is a function of the control parameters ($\lambda_i = \lambda_i(\sigma)$), that means, if we change the conditions of the environment, we can influence the control parameters. So an instability we find if we solve the equation $\lambda_i = 0$. At this point, we can divide our Eigenvectors into stable s_l and unstable u_m modes (Notice, that we now uses the word modes for the Eigenvectors). So here are now our nonlinear equations.

$$\frac{du}{dt} = \lambda_l u_l + Q_l(u, s); \frac{ds}{dt} = \lambda_m s_m + P_m(u, s) \quad (5.1)$$

The Q_l and the P_m are the nonlinear terms. It important to see, that the system generates two time scales! In the two linear version of the equations, the inverse of λ (τ) is , as we know, the decrease rate of the state of the modes. There for, the decrease rate of the unstable mode are much greater than the ones of the stable modes (logical: 1 divided by nearly zero is very big). So it is:

$$\tau_u > \tau_s$$

5.5.3 The slaving principle

Because of the two time scales, the unstable modes are changing very quickly and can achieve very soon. But the stable modes are changing very slow. So we can say that the stable modes are a function of the unstable modes.

$$s = s(u) \quad (5.2)$$

This is the mathematical definition of the Slaving-Principle, and the vector u plays the role of the order parameter. And we have the advantage that often have only a few unstable

modes or order parameters, because it is very seldom, that more than one Eigenvalue of a mode becomes zero at the same time. So we get only one equation.

$$\frac{du}{dt} = \lambda_l u_l + Q_l(u, s(u))$$

We just substituted in equation (5.1) the equation (5.2). And our order parameter is the macroscopic behavior of our system.

5.5.4 Conclusion

An instability does not depend on the detailed subsystems. So we can use the concept of the order parameter directly if we do not know much about the subsystems, e.g. society. Here it sometimes more easily to find the order parameter. This is by the way the second possible approach to a problem of self-organized systems. Because of the powerful concept, the Synergetics has grown to an own science. And many scientist are investigating such systems, which plays an important role in our life and our understanding of our environment.

5.6 Add-on: Chaotic behavior

As mentioned in the first example, synergetic systems lead to behave chaotic by increasing the input energy. There exist many ways to chaos, for more information please refer to other literature.

Literaturverzeichnis

- [1] Hermann Haken, Arne Wunderlin (1991). *Die Selbststrukturierung der Materie*. Vieweg, Braunschweig und Wiesbaden.
- [2] Hermann Haken (1977). *Synergetics: An introduction* Springer Verlag, Berlin

Kapitel 6

A model of the laser

Andreas Manz

24. Mai 2000

6.1 Basic structure of the laser

The main components of a laser are the so called “resonator” and an energy source. The resonator consists of a laser medium which can be a crystal, liquid or gas. Very common is the ruby laser, with a cylindric ruby as the laser medium. On both ends of the medium mirrors are attached. One of them is permeable, and lets a certain amount ($\sim 2\%$) of the light in the resonator pass. The mirrors “resonate” the light waves, that spread parallel to the axis of the crystal. Waves with other directions leave the resonator immediately, so the mirrors cause a dominance of the axis parallel waves. As we will see later, this fact is very important.

Laser activity is only possible far from thermodynamic equilibrium, so the resonator has to be connected to an energy source. The source is often a bright lamp but also electrical or chemical energy supply is possible. The energy supplied to the resonator is called “pumping energy”.

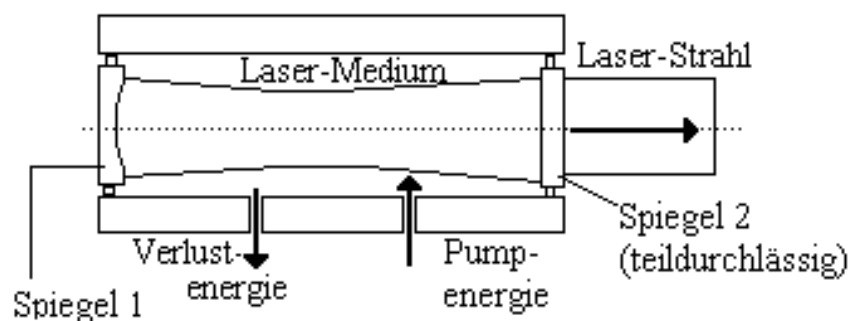


Abbildung 6.1: Schematic structure of a laser.

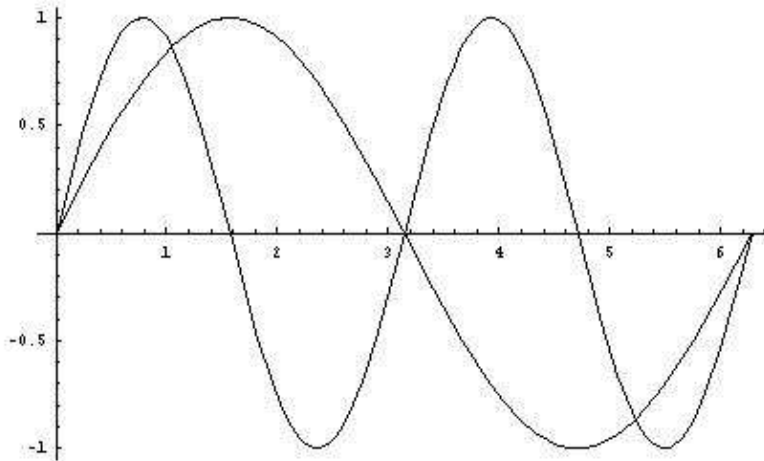


Abbildung 6.2: Possible waves in the resonator.

6.2 About light

6.2.1 Light in general

Light in general is an electromagnetic wave like microwaves or the radio frequencies but with a wavelength between 100nm (ultraviolet) and 1mm (infrared). It is generated by swinging dipoles, in this case the negative electrons surrounding the positive protons. According to the *Bohrian* atom model the electrons move around the core in discrete orbits. An electromagnetic wave is emitted, when an electron leaves the current orbit and jumps down to an orbit nearer to the core. Simplified we can say, that the wavelength of the emitted wave depends on the height of the jump. So only certain orbit changes cause the emission of visible light.

6.2.2 Light in the resonator

Light in the resonator is a standing wave with nodes at both mirrors. Of course this fact is not accidentally. Waves that behave different neutralize themselves, so just the standing waves are selected (See Fig. 6.2).

The electromagnetic field of such a wave is described by:

$$E(x, t) = E(t) \cdot \sin \frac{n\pi}{L} \cdot x \quad (6.1)$$

where $E(t)$ is the amplitude as a function of time, n the number of periods in the resonator and L the length of the resonator. The next question of interest is how the amplitude $E(t)$ behaves in time.

It is of course obvious and the “*nature*” of a wave, that it also behaves like a sinus curve in time. So the amplitude would be:

$$E(t) = A(t) \cdot \sin(\omega t + \phi) \quad (6.2)$$

where ω is the frequency, and ϕ the phase of the wave. Bringing Eq. (6.1) and (6.2) together, leads us to

$$E(x, t) = A(t) \cdot \sin(\omega t + \phi) \cdot \sin \frac{n\pi}{L} \cdot x \quad (6.3)$$

Now we have an equation for the behavior of the electromagnetic field in space and time. But we still don't know, how the amplitude of the wave in time dimension $A(t)$ will change. To answer this question we have to look at the processes taking part in the laser.

6.3 Processes in the laser

The behavior of the amplitude in time is influenced by four processes:

- energy loss at the permeable mirror
- absorption of photons
- spontaneous emission of photons
- stimulated emission of photons

For these processes, except the spontaneous emission, we will find equations, describing their influence on the amplitude.

6.3.1 Loss at the mirror

For this process the equation is quite clear. Every time the permeable mirror is reached by some photons a certain amount of them leave the resonator.

$$\begin{aligned} \dot{A}(t)_1 &= -\gamma A(t) \\ A(t)_1 &= \exp(-\gamma t) \end{aligned} \quad (6.4)$$

If just this process would take place, the electromagnetic field would behave like in Fig.6.3. Every initial wave dies out and zero is the only stable value for the amplitude.

To understand the following three interactions between photons and electrons it is useful to look a little bit closer at the *Bohrian* atom model. Like mentioned before, in this model every electron is assumed to move in an orbit of a definite radius. To each orbit a certain energy level, corresponding to the radius, is dedicated. When energy is supplied to the electrons, they can jump to an orbit with an larger radius and higher energy level. The energy needed to come into this "excited" state may be caused by collision of atoms with each other (thermal animation), collision of electrons with atoms (electrical animation) and collision of photons with atoms (photon animation). The last takes place in case of the ruby laser. The higher energy state is instable, so the electron leaves it after a certain time.

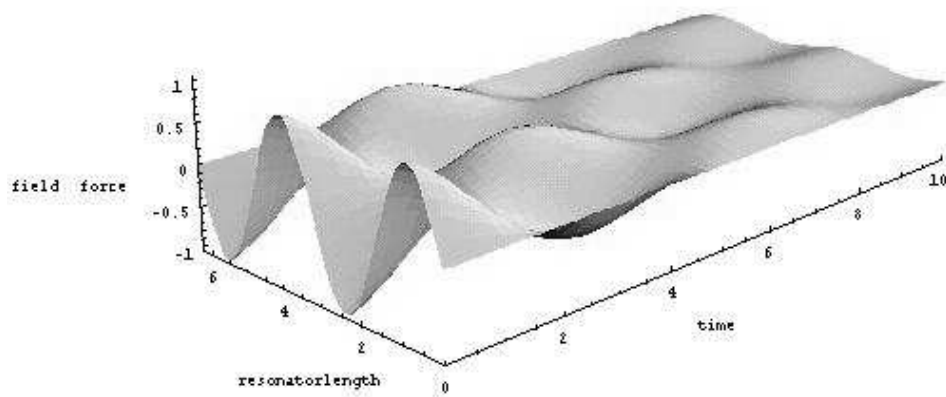


Abbildung 6.3: Wave in resonator without further power supply.

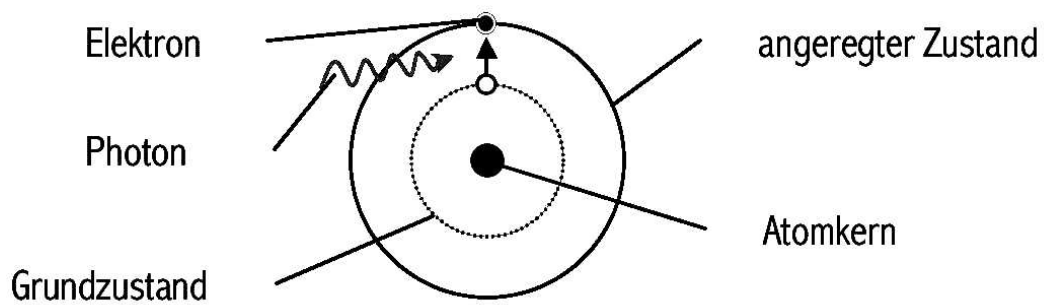


Abbildung 6.4: Absorption of an photon.

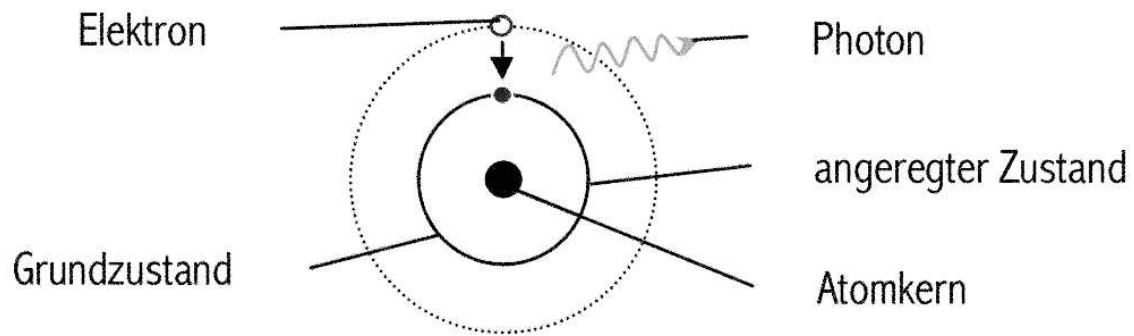


Abbildung 6.5: Spontaneous emission.

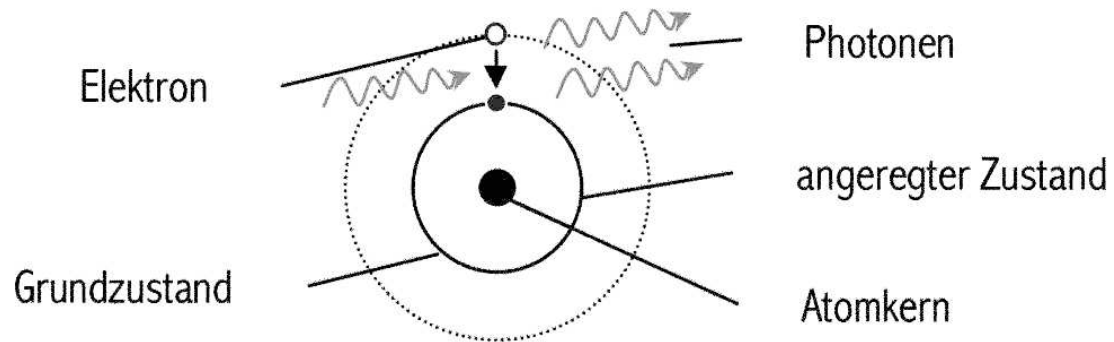


Abbildung 6.6: Stimulated emission.

6.3.2 Absorption

Absorption happens, when a photon of a certain wavelength hits an atom in low energy state. The electron jumps into an orbit with an higher energy level (See Fig.6.4). So the energy of the photon is lost for the electromagnetic field. The process reduces the amplitude of the field. It can be described by

$$\dot{A}(t)_2 = -\alpha A(t)N_1 \quad (6.5)$$

α is called the absorption constant. Of course the reduction of the field is proportional to the number of atoms in the lower energy state, which is represented by N_1 . Only they can absorb photons.

6.3.3 Spontaneous emission

The higher energy state is instable for the electron. After a few nanoseconds it falls back to its normal orbit and emits an photon with an accidental direction and phase. Because of this fortuitousness this process on the hole doesn't amplify or weaken the existing field (Fig.6.5).

6.3.4 Stimulated emission

The stimulated emission is the most important process for the laser (laser means "light amplification by stimulated emission of radiation"). When an electron is in the high energy

state it can be forced by an arriving photon to leave this state. The photon that is emitted during this process leaves the atom in exactly the same direction and with same phase like the photon that caused the emission (Fig.6.6). Now the fact becomes important, that, caused by the mirrors, the photons moving parallel to the axis of the laser stay in the resonator for a longer time. So this direction dominates all the others and there is a high probability that photons with this direction cause the stimulated emission of an excited atom. So they force more and more photons to move in same direction and same phase. When enough power is supplied this causes a chain reaction.

So we can suppose, that the “new” photon amplifies the existing wave. The process is described by the equation:

$$\dot{A}(t)_3 = \alpha A(t) N_2 \quad (6.6)$$

α is the same absorption constant as in Eq.(6.5). N_2 is the number of atoms in high energy state.

6.4 How does the amplitude behave?

6.4.1 The laser condition

We now come back to the question how the amplitude of the electromagnetic field behaves in time. We can add the equations (6.4),(6.5), (6.6) for the single processes

$$\begin{aligned} \dot{A}(t) &= \dot{A}(t)_1 + \dot{A}(t)_2 + \dot{A}(t)_3 \\ &= (\alpha(N_2 - N_1) - \gamma)A(t) \end{aligned} \quad (6.7)$$

$$G := \alpha(N_2 - N_1) - \gamma \quad (6.8)$$

$$A(t) = a \cdot \exp(Gt) \quad (6.9)$$

So the amplitude grows exponentially for $G > 0 \iff N_2 - N_1 > \frac{\gamma}{\alpha}$. This important condition is called **Laser condition** of Schawlow and Townes. $N_2 - N_1$ is called **Inversion**. When the laser condition is reached the amplitude behaves like shown in Fig. 6.7.

6.4.2 What limits the growth of the amplitude?

It is obvious, that the amplitude of the field can not grow exponentially over a long time. The power supply limits the growth. This fact appears in our calculation for the amplitude $A(t) = a \cdot \exp(Gt)$ in the way that G isn't a constant but a function of time $G(t)$ and more precise the inversion $N_2 - N_1$ is changing in time. It is quite clear, that the difference between the number of electrons in high and in low energy state decreases when light is emitted. To specify this change in detail we have to state an so called “basic inversion”.

$$D_0 := N_{2,0} - N_{1,0} \quad (6.10)$$

This is the maximum inversion that can be reached for a certain amount of pumping energy. All possible atoms are in the high energy state, but non of them already left this

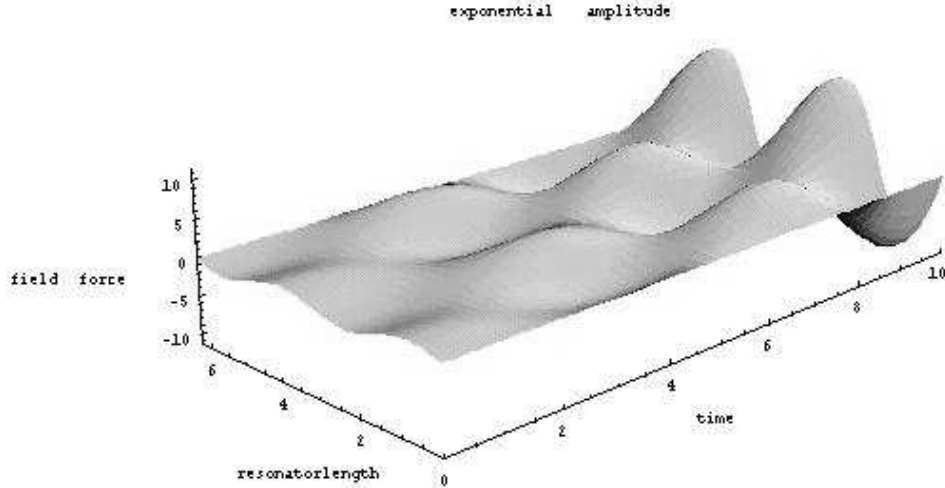


Abbildung 6.7: Exponential growth of the amplitude.

state because of spontaneous or stimulated emission. In this state $\Delta W \cdot D_0$ energy is stored in the laser medium, where ΔW is the energetic difference between the “excited” and the normal state. The energy stored during the “normal” inversion $\Delta W \cdot D$ with $D := N_2 - N_1$ is this energy during the maximum inversion minus the energy supplied to the field.

$$\Delta W \cdot D = \Delta W \cdot D_0 - bA(t)^2 \quad (6.11)$$

$$D = D_0 - cA(t)^2 \quad \text{with} \quad c := \frac{b}{W_2 - W_1} \quad (6.12)$$

Now we can substitute $N_2 - N_1$ in Eq. (6.8) by $D_0 - cA(t)^2$ from Eq.(6.12).

$$\begin{aligned} \dot{A}(t) &= (\alpha \cdot (D_0 - cA(t)^2) - \gamma)A(t) \\ &= (\alpha D_0 - \gamma) \cdot A(t) - \alpha c A(t)^3 \\ &= B \cdot A(t) - C \cdot A(t)^3 \quad \text{with} \quad B := \alpha D_0 - \gamma, \quad C := \alpha c \end{aligned} \quad (6.13)$$

6.4.3 Stability of the amplitude

Now we want to know, for which values the amplitude doesn't change, and which of these equilibrium states are stable.

$$\begin{aligned} \dot{A}(t) &= 0 \\ \iff 0 &= B \cdot A(t) - C \cdot A(t)^3 =: f(A) \end{aligned}$$

$$A_1(t) = 0, \quad A_{2,3}(t) = \pm \sqrt{\frac{B}{C}} \quad (6.14)$$

We may assume, that C is positive, so the solutions $A_{2,3}(t)$ just have real values when $B \geq 0$. That means $D_0 = N_{2,0} - N_{1,0} \geq \frac{\gamma}{\alpha}$, which is again the laser condition (which

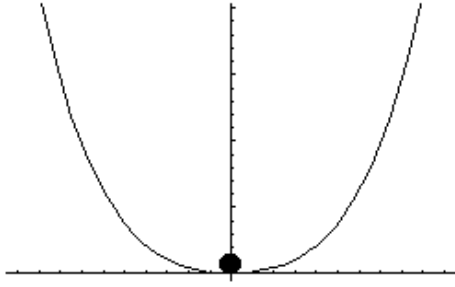


Abbildung 6.8: Potential for $G < 0$.

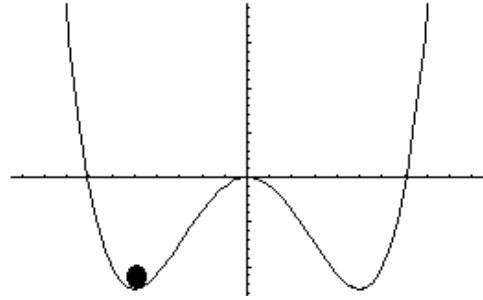


Abbildung 6.9: Potential for $G > 0$.

we also designated as $G > 0$ in Eq.(6.9)). To examine the stability of these equilibrium states we calculate the potential of $f(A)$. The “landscape” of the potential illustrates the behavior of the amplitude (Fig. 6.8,6.9).

Fig.6.8 shows the potential, when the laser condition is not yet reached. The amplitude behaves like a ball in this landscape, it will always return to the only stable value: Zero. When the pumping-energy is increased the value zero will become instable and the two new stable solutions $A_{2,3}(t) = \pm\sqrt{\frac{B}{C}}$ appear (Fig .6.9). Which of them is selected is accidental. But in fact the difference between them is only a phase transition of the wave by 180° .

Literaturverzeichnis

- [1] Haken, H., Graham, R. (1971). Synergetik - die Lehre vom Zusammenwirken. *Umschau* 6, 191-195.
- [2] Haken, H. (1981). *Erfolgsgeheimnisse der Natur*. Deutsche Verlags-Anstalt, Stuttgart.
- [3] Haken, H., Wunderlin, A. (1991). *Die Selbststrukturierung der Materie*. Vieweg, Braunschweig und Wiesbaden.

Kapitel 7

Droplet impacts - more than just a few nice pictures

Tobias Niemann

31. Mai 2000

Abstract

This seminar-paper is about the mathematical and experimental background of *droplet impacts*. There is much more about droplet collisions than just nice and fascinating pictures. This paper introduces to the fascinating theories about *droplet impacts*, sets examples of the wide range of other disciplines suffering from its results and shows different systems describing the process of impact. But, of course, from time to time there will be some pictures of very nice collisions, too.

7.1 Droplet impacts - what is that?

A droplet is a little pice of water stuck together by gravitation and the Van der Waals' force. There are many situations when such droplets collide with other droplets or with surfaces. Those situations for example come to happen in clouds, during the process of fuel injection or in medical asthma-sprays.

7.1.1 droplet collision with surface

Collisions with surfaces are most important in technical environments. Depending on the kind of surface, the droplet-speed, the art and amount of fluid and the angle of impact the collisions are varying a lot. Those influences on the impacts needed to be researched to better understand the act of collision. These impacts are easier to simulate and calculate because there is only one moving part.

7.1.2 droplet collision with droplet

Collisions with other droplets are at first natural collisions. But those collisions are also frequent in technical environments. Because both parts are moving it is more difficult to

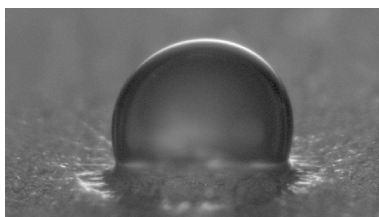


Abbildung 7.1: nice picture of a splashing droplet (taken from [2])

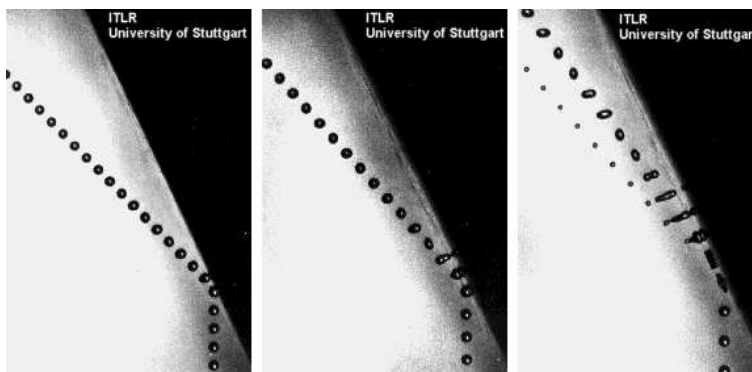


Abbildung 7.2: photo series of impacting droplets with three different impact-angles (taken from [4])

describe mathematically. You need to know both droplet-speeds, the angle of impact and the art and amount of both fluids.

7.2 Why research on it?

In science results from specific research projects are often useful in other parts or disciplines, too. All together there are two good reasons why to research it. Those reasons will be described now.

7.2.1 reference to reality

Colliding droplets is a natural process. While researching the formation of clouds you will come to the point when little droplets collide and merge to a thicker one. This is where knowledge about droplet collisions is needed. An other process is the process of spraying in medical sprays, fuel injection or fire-fighting apparatus. The haze should be dispersed so it can act as good as possible. And when thinking about it you can find a lot more other processes where droplets collide.

7.2.2 theory works in atom and star modeling

What happens when stars hit each other? It is very difficult to simulate because they are huge pieces of matter. But physicists tend to simplify all complex objects and a star simplified as a point or droplet can be described with theories from droplet collisions. The same is in atom modeling. Atoms are colliding very often. With the mathematical tools developed during simulation of *droplet impacts* many processes can suddenly be explained or simulated easily.

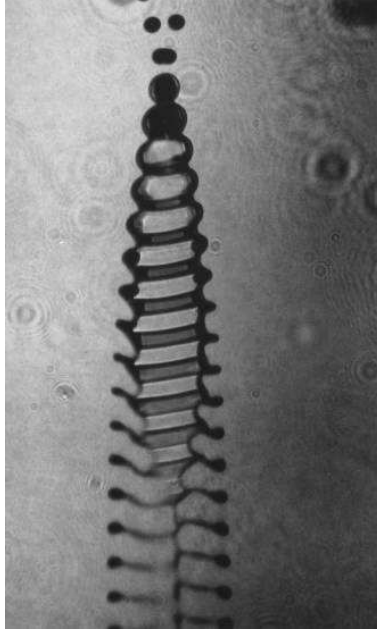


Abbildung 7.3: droplet-droplet collision (taken from [1])

7.3 Experimental approach

As computers were not that fast and therefore could not help solving large numbers of calculations the only way to research *droplet impacts* was to set up experiments and try to analyze its results.

Typical experiment arrangements (figure 7.4) for droplet-droplet and droplet-surface impacts produce continuous drops out of a jet of water.

With taking one photo scientist are able to see many stadiums of the impact (see figure 7.5).

As droplet-surface outcomes two different behaviours could be found [2]. A drop can *splash* (figure 7.6 c-f) or it can *deposit* (figure 7.6 a). The *rebound* (figure 7.6 b) phenomenon is a part of the two mentioned impact-methods.

Investigation showed 6 kind of outcomes for splashes. In reality the “splash” can be divided into 4 kinds of physical phenomenon.

But experiments can't describe other collisions. For a better understanding and the ability to simulate collisions a theoretical approach is needed.

7.4 Theoretical approach

7.4.1 classical mechanics

The first idea is to use the tools of the classical mechanics particularly the sentences of conservation of mass, momentum and energy. It is quite easy to get the equations that describe the behaviour of the two phase system of the droplet-fluids and the surrounding gas. But it is normally not that trivial to solve those equations to get a mathematical formula. The difficulty is the extreme nonlinearity of the equations because the effect

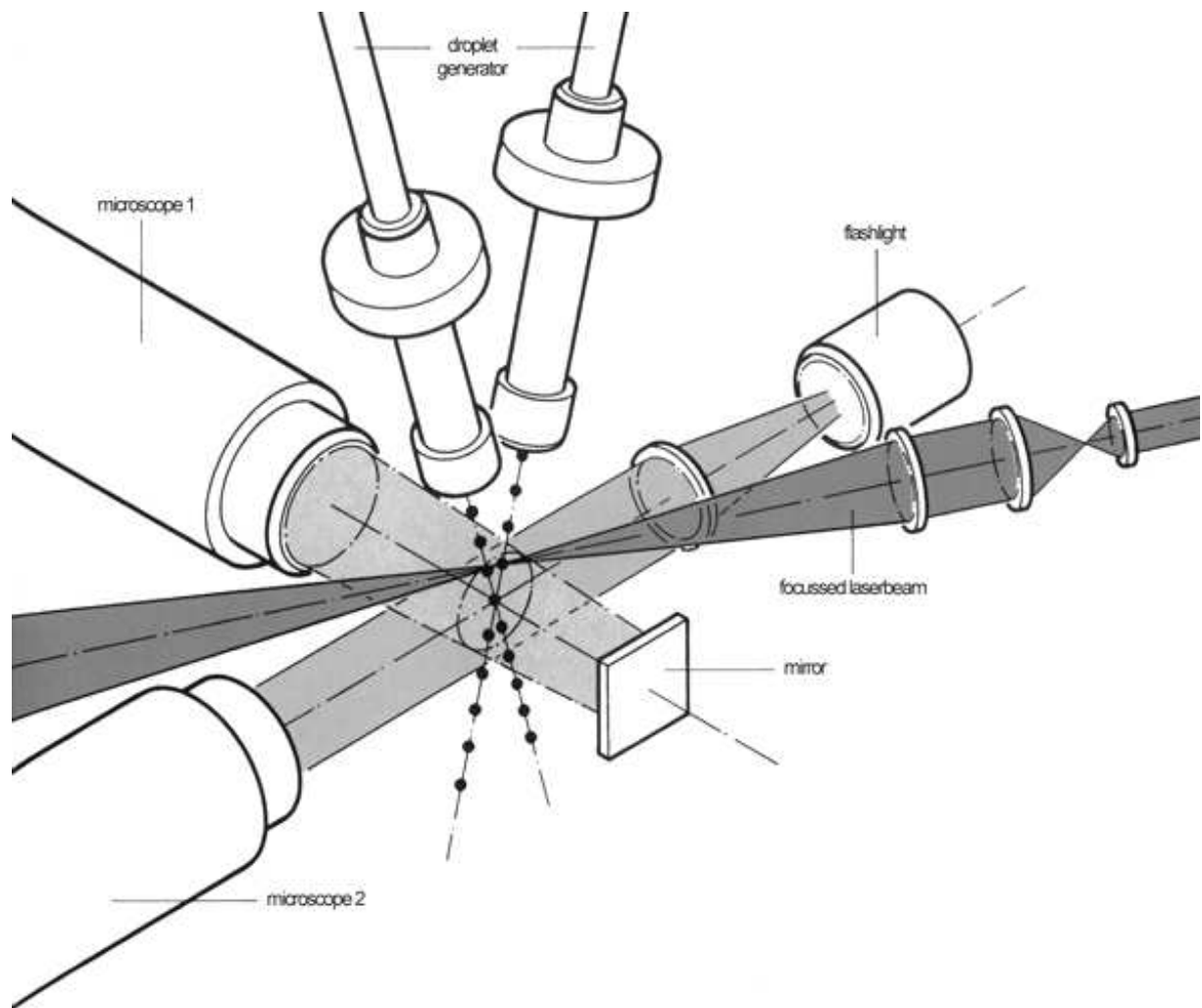


Abbildung 7.4: experimental setup for droplet-droplet collision (taken from [1])

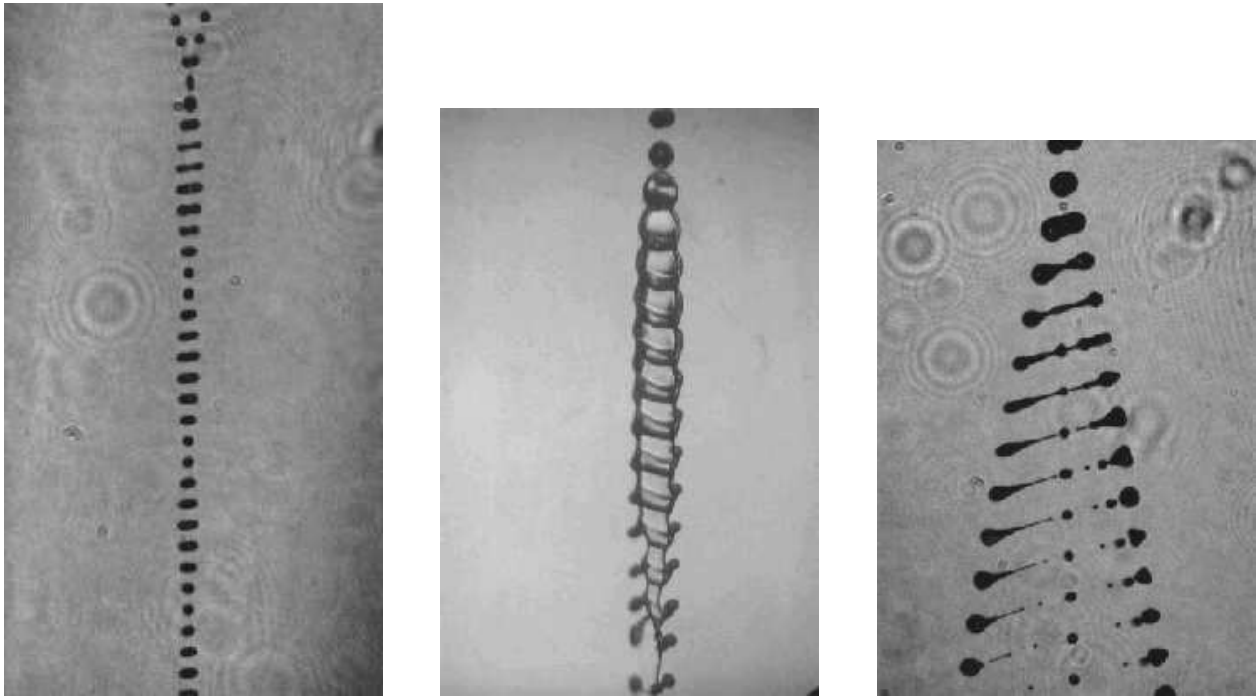
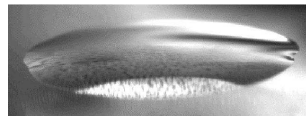
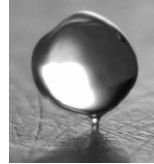


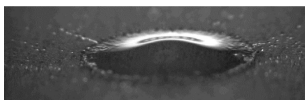
Abbildung 7.5: snapshots of droplet-droplet collisions (taken from [1])



(a) deposition



(b) rebound



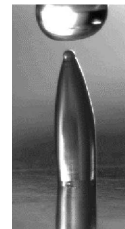
(c) prompt splash



(d) corona splash



(e) receding
break-up



(f) par-
tial
re-
bound

Abbildung 7.6: drop impact outcomes (taken from [2])

is roughly proportional to the cause. The supposition for using the usual techniques (for solving linear equations) is not given. And, of course, there are no general analytic solutions for this kind of nonlinear equations.

So how to solve such a problem? The answer is *with numerical modeling*.

7.4.2 numerical modeling

With the help of computers numerical modeling grew to a powerful tool within the last years. In general for this problem two different basic approaches can be used. One is the Navier-Stokes and the other one is the Lattice Boltzmann method.

Navier-Stokes

- Claude Navier (french engineer (1785-1838))
- Sir George Gabriel Stokes (english physicist and mathematician (1819-1903))

Lattice-Boltzmann

- lattice = Gitter
- Ludwig Boltzmann (one of the founder of thermodynamics (1844-1906))

But before using those methods some simplifications are still needed. Otherwise the number of calculations could exceed the capacity of even very fast computers. One simplification is a constant droplet temperature. Another one is a constant drift speed far below the velocity of sound so air is able to move around coming droplets. That's why the density of gas and fluid are constants like if both are incompressible medium.

Navier-Stokes

The Navier-Stokes technique is a system of nonlinear partial differential equations. The streaming medium is continuous so the system can be described with functions that have in every point at any time within the area of calculation a certain value (e.g. pressure and streaming velocity). If the system consists of two very different media another function can be introduced. It is called the *volume of fluid*. The function is 1 if it is gaseous and 0 if fluid. At the borders there is a jump between 0 and 1. The same is with the density and viscosity because they are variables from volume of fluid.

The unknown values in this special variation of the Navier-Stokes equation are:

- pressure
- streaming velocity
- volume of fluid

The equations themselves describe how the quantities change in time depending on their actual value. But the surface-tension also needs to be noticed. This value is directed to the middle and can be calculated from the volume of fluid. But for the final numerical calculation two more simplifications have to be done.

It's impossible to calculate every point. They are way too much. Instead of that it is possible to calculate the mean of volume elements. Figure 7.7 shows a surface and the grid containing the value of the volume of fluid function. The jump between 0 and 1 now is softer, too. The other simplification is the assumption of non-continuous time. Instead

| | | | | |
|---|------|------|------|------|
| 0 | 0 | 0 | 0 | 0 |
| 0 | 0 | 0,05 | 0,45 | 0,76 |
| 0 | 0,05 | 0,81 | 1 | 1 |
| 0 | 0,45 | 1 | 1 | 1 |
| 0 | 0,76 | 1 | 1 | 1 |

Abbildung 7.7: mean of volume of fluids (taken from [3])

of that several time-points are chosen. This has to be done because the equations can't be solved at once but need several iterations.

The simulation pictures of droplet-droplet collision (figure 7.9) and droplet-surface collision (figure 7.8), both calculated with the Navier-Stokes technique, show some results.

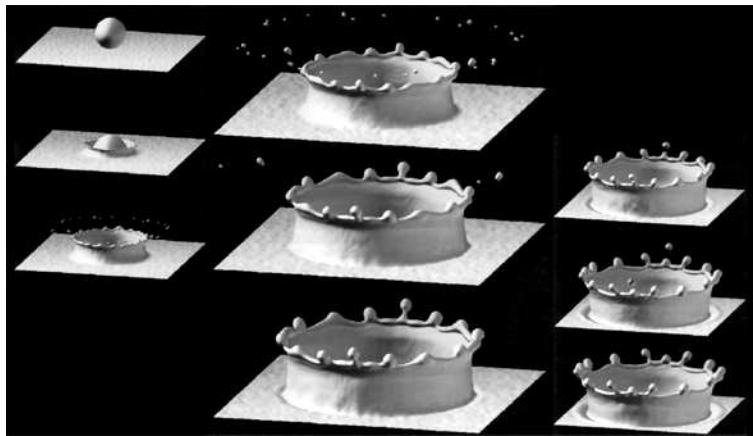


Abbildung 7.8: a nice crown produced by a small droplet splashing into a liquid (taken from [3])

Lattice-Boltzmann

Again some simplifications are needed. The streaming velocity in a point is the mean of all molecules in short distance. What you get is a function $f(\vec{x}, t, \vec{c})$ describing the probability to meet at position x at time t a molecule with the speed c . The function f can be described with a differential equation called Boltzmann equation.

The advantage of this is a more detailed description. The disadvantage is also a more detailed description because it is difficult to calculate because it is dependent on velocity. That's why another simplification is needed. The original term, describing the effects of collision between droplets, has to be substituted with a term (a force) that moves everything to thermal equilibrium. Now the continuous variables place, time and molecular velocity can be substituted with a lattice of discrete values (figure 7.10). The Lattice-Boltzmann equation can now explicitly be solved in time (without iterations) for every knot in the lattice.

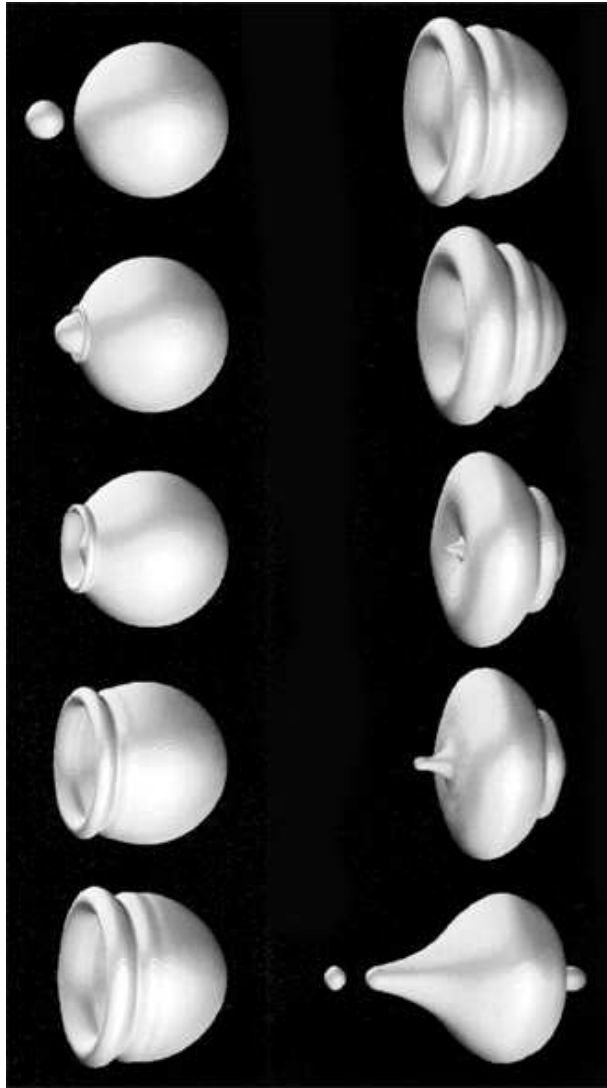


Abbildung 7.9: collision between a small and a big droplet (taken from [3])

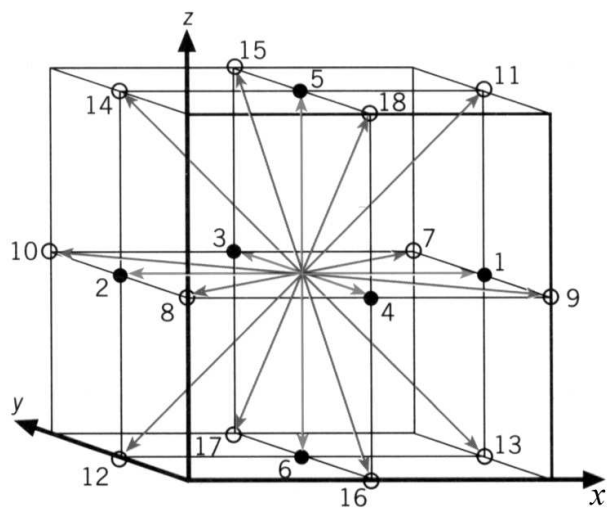


Abbildung 7.10: the lattice with the to be calculated object in the middle (taken from [3])

For modeling a two-phase-system the power of attraction between particles has to be added like in reality. It is a function of density at two neighbour-knots in the lattice. A sample pictures of droplet-droplet collisions (figure 7.11):

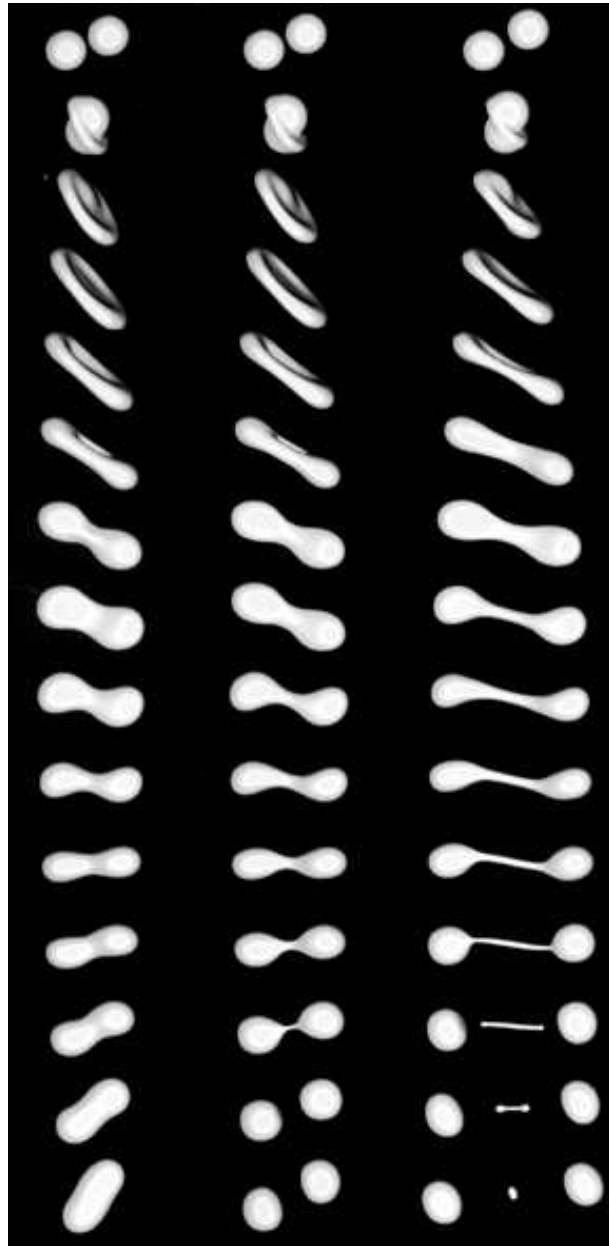


Abbildung 7.11: different droplet-droplet collisions with different angles (taken from [3])

7.5 reality vs modelling

Research showed very good conformity between both numerical solutions and experiments [3] (see figure 7.12):



Abbildung 7.12: series of pictures taken from an animated gif (taken from [3]); above the simulated collision, below the real collision

7.6 Conclusions

There are many nice pictures:

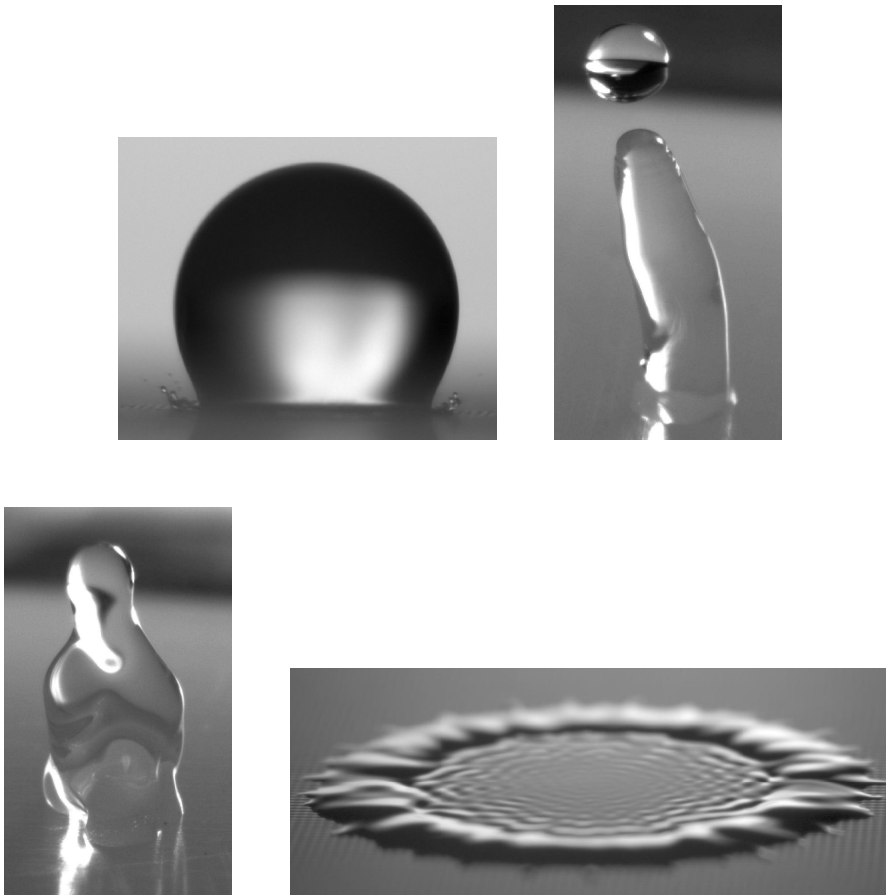


Abbildung 7.13: some nice pictures (taken from [2])

It is important for “everyday-physics” like mentioned at the beginning. The mathematical techniques can help to solve problems in other sciences. And last but not least I have to say it can be quite good modeled. The simulations are very realistic.

Literaturverzeichnis

- [1] G. Brenn, A. Frohn, *Tropfenkollisionen*, Bild der Wissenschaft, 12/1990
- [2] <http://www.sla.mschinenbau.tu-darmstadt.de/rioboo/drpgall.html>
- [3] M. Schelkle, M. Rieber, A. Frohn, *Numerische Simulation von Tropfenkollisionen*, Spektrum der Wissenschaft DOSSIER "Software", 2/1999
- [4] ITLR, Universität Stuttgart, 7/1999

Kapitel 8

Die Entstehung des Lebens

Ba Kien Tran

.ps 7. Juni 2000

Zusammenfassung

Das Thema meines Referates ist die Entstehung des Lebens. Zunächst werde ich erläutern was Leben ist und einen kurzen Überblick über die Hypothesen zur Lebensentstehung geben. Dann werde ich näher auf die „Urerde“ eingehen und die zugehörige Simulationsversuche zur präbiotischen Stoffsynthese darstelle. Danach erzähle ich etwas über die Theorie der RNA-Welt und das Organisationsprinzip Hyperzyklus.

8.1 Was ist Leben?

Um über die Entstehung des Lebens diskutieren zu können müssen wir zunächst klären, was Leben ist.

Dazu gibt es sieben notwendige und hinreichende Eigenschaften lebender Systeme (nach De Duve [1]):

1. Aus dem Material, das in seiner Umgebung zur Verfügung steht, seine Bestandteile herstellen, das heißt synthetisieren.
Der Aufbau der wichtigen Bestandteile lebender Organismen läuft im wesentlichen auf zwei Grundvorgänge hinaus: Reduktion und dehydrierende Kondensation.
2. Energie aus seiner Umgebung aufnehmen und in verschiedenen Formen von lebenswichtiger Arbeit umwandeln.
Dabei betreibt die Hydrolyse von ATP zu ADP und P_i praktisch alle Formen von Arbeit.
3. Die vielen chemischen Reaktionen katalysieren, die für seine Aktivitäten nötig sind. Die biologischen Katalysatoren sind die Enzyme. Man ging davon aus, daß alle Enzyme Proteine sind. Aber man entdeckte, daß einige RNAs auch katalytische Aktivitäten aufweisen.

4. Seine biosynthetischen und andere Vorgänge durch Informationen steuern, so daß eine getreue Reproduktion gesichert ist.
Information ist nötig für jede spezifische Bindung zwischen zwei Molekülen. Man spricht davon, daß sich zwei Moleküle erkennen. Die Grundlage solcher Erkennungsphänomene ist genaue Komplementarität zwischen den beiden Molekülen. Das sogenannte Schlüssel-Schloß-Prinzip.
5. Sich so isolieren, daß der Austausch mit der äußeren Umgebung genau kontrolliert ist.
Alle lebenden Zellen sind durch Membranen umgeben und manchmal auch weiter unterteilt. Diese Membranen bestehen hauptsächlich aus Lipiddoppelschichten. Lipiddoppelschichten sind sehr flexibel und organisieren sich automatisch zu geschlossenen Strukturen. Somit können sie sich durch Fusion miteinander verbinden und durch Spaltung trennen.
6. Seine Tätigkeit so regulieren, daß seine dynamische Organisation trotz Änderung der Umweltbedingungen erhalten bleibt.
Dies geschieht meist durch negative Rückkopplung. Dabei korrigiert die Wirkung die Störung, die sie verursacht. Entscheidend bei der Regulation sind Enzyme, deren katalytische Aktivität sich durch sogenannte Aktivatoren erhöhen und Inhibitoren erniedrigen lassen.
7. Sich vermehren.
Genaugenommen ist dies keine Voraussetzung für das Leben. Aber jede Spezies, die sich nicht vermehren kann stirbt aus. Deshalb gehört es zum Grundbauplan des Lebens.

Diese Eigenschaften findet man schon in einer einzigen Zelle. Viele Lebewesen bestehen auch nur aus einer Zelle. Somit ist die Zelle die Grundeinheit des Lebens.

8.2 Hypothesen zur Lebensentstehung

Zu der Frage nach dem Ursprung des Lebens hatten die Menschen zunächst zwei mögliche Antworten. Einmal die Schöpfung, wie in der biblischen Schöpfungsgeschichte und die spontane Entstehung „Urzeugung“. 1668 führte der italienische Biologe Francesco Redi Experimente durch, die diese Theorie widerlegen sollte. Aber erst Louis Pasteur machte durch seine sterilen Experimente die spontane Entstehung endgültig unglaubhaft. Der sowjetische Biochemiker Alexander Oparin veröffentlichte 1924 ein kleines Band zur Entstehung des Lebens. Er ist der Auffassung, daß Leben ein natürliches Phänomen innerhalb der Evolution der Erde ist. Theorien über die Entstehung des Lebens auf der Erde waren also immer stark durch die gültige Religion und Philosophie geprägt. Auch Oparins Hypothese ist von einer Ideologie geprägt, da er überzeugter Marxist war und in seinen Veröffentlichungen oft Engels zitiert wurde.

Heute geht man davon aus, daß die Evolution der Biosphäre in drei Abschnitten erfolgte:

1. Chemische Evolution:
In der chemischen Evolution entstanden aus den kleinen Molekülen der Uratmosphäre unter Nutzung geeigneter Energiequellen Biomonomere, die zu Makromolekülen (Proteine und Nukleinsäuren) zusammentraten.

2. Präbiotische Evolution:

In der präbiotischen Evolution fand die Selbstorganisation eines Gemisches von Makromolekülen zu sich selbst reproduzierenden Einheiten statt, die schließlich zur Entstehung der „Urzellen“ führte.

3. Biologische Evolution:

Die biologische Evolution führte dann zu den Formen des heutigen Lebens.

8.3 Die „Urerde“

Unsere Erde entstand etwa vor 3,7 Milliarden Jahre. Durch die unterschiedliche Gewichte der Elemente und der Temperaturzonierung reicherten sich die schweren Elemente (Eisen und Nickel) im geschmolzenen Erdkern an. Atome mittleren Gewichtes (Silizium und Aluminium) formierten sich um der Erdkern zum festflüssigen Erdmantel. Die Erdkuste oder Lithosphäre besteht aus leichten, hydratisierten Mineralien, wie Basalt und Granit. So entstand eine in Schalen differenzierte Erde.

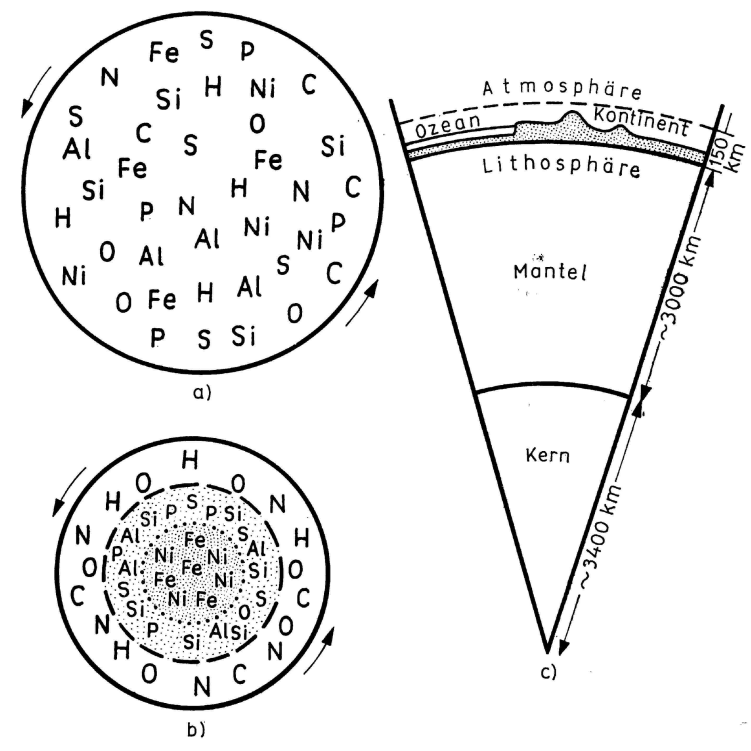


Abbildung 8.1: Entstehung des Erdaufbaus. a) diffuse Verteilung der Elemente in der glühenden Erdkugel b) zunehmende Verdichtung der Erdkugel und Anordnung der Elemente nach ihrem Atomgewicht c) Schalenbau unserer heutigen Erde (aus [5])

Da die Erde eine relativ geringe Gravitation besitzt und eine hohe Temperatur herrschte (auch durch die Sonnenentwicklung) kam es zu Entweichen von Gasen wie Wasserstoff, Helium, Methan, Ammoniak usw. in den Weltraum.

Nachdem sich die Erde unter 100°C abgekühlt hatte fand eine zunehmende Kondensation von Wasserdampf statt. Was zu der Entstehung von den Urozeanen (Hydrosphäre) führte.

8.3.1 Die Uratmosphäre

Wichtig bei der Entstehung des Lebens war noch die Uratmosphäre. Nachdem die primäre Atmosphäre entwichen war und die Erde abkühlte entstand durch Ausströmen vulkanischer Gase eine sekundäre Atmosphäre. Dies enthielt wahrscheinlich Wasserstoff (H_2), Wasser (H_2O), Ammoniak (NH_3), Stickstoff (N_2), Methan (CH_4), Kohlenmonoxid (CO), Kohlendioxid (CO_2) und Cyanwasserstoff (HCN). Wobei viele Autoren der Ansicht sind, daß die Erde nach dem Ausgasen im Zuge der Abkühlung ohne Wasserstoff war.

8.3.2 Die Energiequellen der „Urerde“

Für die abiotischen Reduktions- und Synthesereaktionen war natürlich Energie nötig. Die Energiequelle der frühen Erde waren Sonnenlicht und, aufgrund der fehlenden Ozonschicht, auch das UV-Licht. Elektrische Energie stand durch Blitze und sog. Glimmerentladungen bei Gewittern zu Verfügung. Die Gewitter und Meteoriteneinschläge setzten durch Druckwellen auch Energie frei. Der Vulkanismus war wahrscheinlich sehr ausgeprägt. Auf der Oberfläche von Lava war also thermische Energie verfügbar. Radioaktive Elemente waren eine weitere Energiequelle.

8.4 Simulationsversuche zur präbiotischen Stoffsynthese

Der erste erfolgreiche Versuch einer präbiotischen Synthese gelang 1953 den Chemiker Stanly Miller. Er versuchte im Labor des Kosmologen und Pysiker Harold Urey, die Bedingungen zu reproduzieren, die auf der Oberfläche der primitiven Erde geherrscht hatten. Miller setzte eine aus relativ wenigen Komponenten bestehende, simulierte Uratmosphäre (H_2 , NH_3 , CH_4 und Wasserdampf) elektrischer Entladung aus.

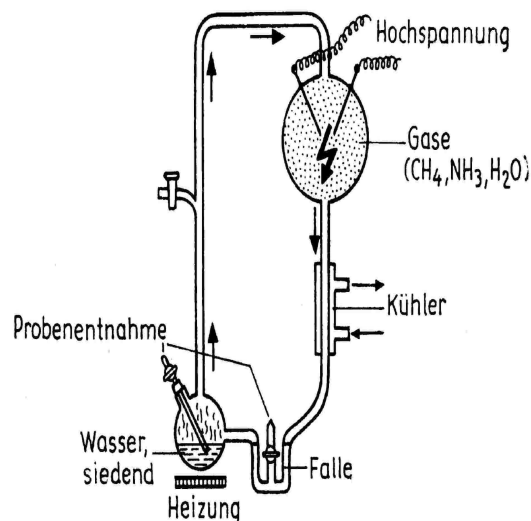


Abbildung 8.2: Apparatur für Simulationsversuche zur abiogenen Stoffsynthese (aus [5]).

Es entstanden Stoffe, die wir aus Lebewesen kennen wie z.B. Glycin, Alanin, Glutaminsäure und Milchsäure. Danach wurden noch zahlreiche Experimente dieser Art durchgeführt.

Unter anderem konnte man Aminosäuren, Purin, Pyrimidine und Zucker synthetisieren. Aminosäuren erhielt man z.B. wenn man H_2 , N_2 , CH_4 und Wasserdampf elektrische Entladungen aussetzt. Zweifelhaft an diesen Versuchen ist die Annahme von Miller und Urey, daß die präbiotische Atmosphäre stark Wasserstoffs haltig war. Denn, bei der Kondensation durch die Abkühlung der Erde, ist wahrscheinlich der größte Teil des Wasserstoffs verloren gegangen. Damit wahr als Kohlenstoffquelle eher CO_2 als CH_4 übrig. Wodurch sich die Ausbeute an Aminosäuren erheblich verringert wären.

| Verbindung | % Kohlenstoff aus CH_4 |
|------------------------------|--------------------------|
| Ameisensäure | 4,0 |
| Glycin | 2,1 |
| Glycolsäure | 1,9 |
| Alanin | 1,7 |
| Milchsäure | 1,6 |
| β -Alanin | 0,76 |
| Propionsäure | 0,66 |
| Essigsäure | 0,51 |
| Iminodiessigsäure | 0,37 |
| α -Aminobuttersäure | 0,34 |
| α -Hydroxybuttersäure | 0,34 |
| Bernsteinsäure | 0,27 |
| andere | 0,62 |

Abbildung 8.3: Präbiotische Bausteine, die entstehen, wenn man $CH_4 + NH_3 + H_2O + H_2$ entzündet (aus [1]).

Neuere Forschungen haben ergeben, daß das Weltall eine Reihe potentiell biogene Bausteine enthält. Somit könnten die ersten biogenen Bausteine auch ein Produkt interstellarer Chemie gewesen sein. Auf jeden Fall lagen in der Erdatmosphäre durch zufällige chemische Prozesse biogene Bausteine vor. Vermutlich sammelten sich die Moleküle in den Ozean und Seen an. Dort reicherten sie sich in der seichten Gebieten zu hohen Konzentrationen an (günstiger Verteilungskoeffizient) und waren durch Wasser vor den zerstörerischen Strahlungen geschützt. Es entstand die sog. „Ursuppe“. Man kann davon ausgehen, daß sich auch komplexe Kombinationen dieser Bausteine bildeten.

8.5 RNA-Welt

Es wurde angenommen, daß Nukleinsäuren als Erbspeicher sich nur mit Hilfe von Proteinen vermehrt werden können und Proteine nur exakt wieder hergestellt werden können, wenn ein informationserteilender Nukleinsäure-Code besteht. Stellt sich also die Frage was zuerst da war. 1981 wurde ein Beispiel für RNA-Katalyse gefunden (Cech) und führte zu den Konzept „RNA als Enzym“. Am Anfang war wahrscheinlich eine proteinlose RNA-Welt. Fraglich ist wie die RNA entstanden ist. Diese RNA konnten sich nur ungenau replizieren, so daß sich nach Eigen „Quasi-Spezies“ bildeten, also Populationen sehr

ähnlicher, aber nicht identischer Moleküle (Mutanten). Es fand dann eine darwinistische Evolution auf molekularer Ebene statt.

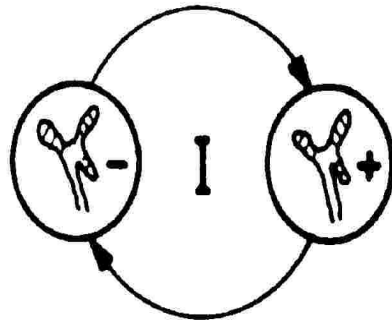
8.6 Der Hyperzyklus

Nach der Meinung von Eigen muß das „Urgen“ zwischen 50 und 100 Nukleotide lang gewesen sein. Ein kürzeres wäre zu instabil. Ein längeres enthielte zu viele Replikationsfehler. Um mehr Informationen exakt zu Reproduzieren sind Katalysatoren notwendig. Somit verbanden sich Nukleinsäuren und Proteine als Informationsträger und Funktionsträger zu einem autokatalytischen System. Eigen und Schuster sprechen von einer hyperzyklischen Organisation. Ich versuche nun die Entstehung des Hyperzyklus zu erklären.

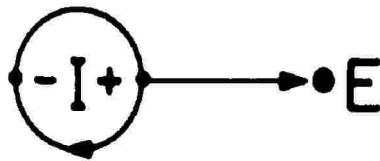
1. Zunächst war wie ich erklärt habe wahrscheinlich eine selbstreplikative RNA.



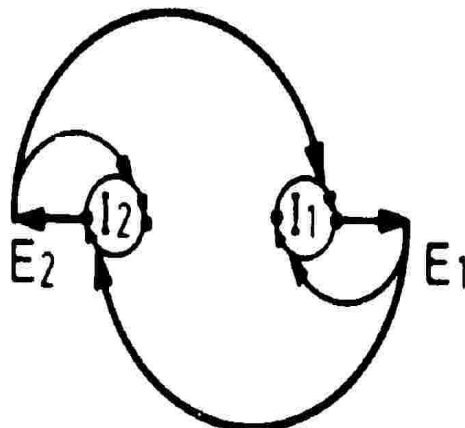
2. Diese reproduzierte sich durch Komplementarität und bildet ein Informationsträger. Diese konnten sich vermehren und mutieren. Bildeten also eine Quasi-Spezies.



3. Es kam durch Mutation dazu, daß ein Informationsträger Instruktionen zur Bildung eines Enzyms enthält. Ein passendes Enzym konnte die Bildung des Informationsträgers katalysieren.

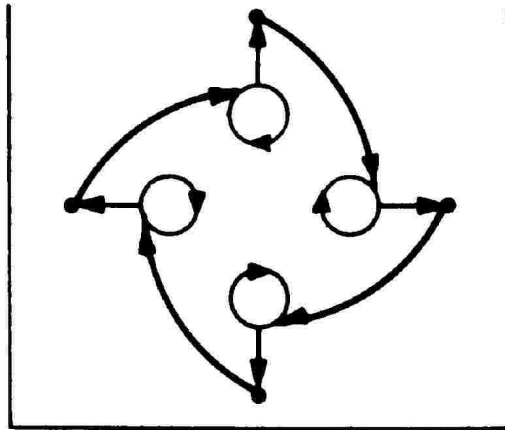


4. Dann bildet sich eine Mutation des Informationsträgers. Durch ihre Ähnlichkeit können die Enzyme beide katalysieren.

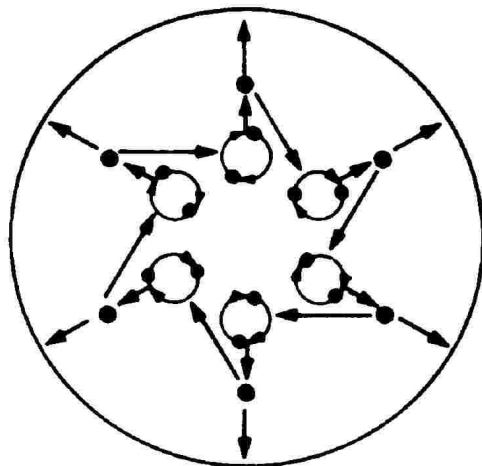


5. Das System entwickelt sich weiter und weitere Mutationen werden integriert. Die

Enzyme katalysieren dann nur noch den nächsten Informationsträger. Durch den Hyperzyklus kommt es zur Erweiterung der Informationsmenge trotz begrenzter Replikationsgenauigkeit der einzelnen Informationsträger. Die Konkurrenz wird durch die zyklische Kopplung ausgeschaltet und in eine Kooperation umgewandelt. Ein Hyperzyklus kann sich durch Mutation verbessern, ausweiten oder ökonomisieren. Er läßt sich aber nicht durch ein anders System ablösen. Dies führt zu einer „Ein-für-allemal“-Entscheidung, wodurch sich z.B. der universelle genetische Code erklären läßt. Außerdem könnte es sein, daß der Hyperzyklus z.B. durch Poren im Gestein teilweise vom Milieu abgegrenzt wurde.



6. Der nächste Schritt besteht darin, daß sich der Hyperzyklus mit Replikations- und Translationsapparat vollständig vom Milieu abgrenzt. Dies geschieht durch Membranen. Zum ersten Mal kommt es auch zu einer Kopplung zwischen Geno- und Phäno-Typ. Der kompartimentierte Hyperzyklus bildet eine Protozelle.



Literaturverzeichnis

- [1] DUVE, CH. DE (1994). Ursprung des Lebens - Präbiotische Evolution und die Entstehung der Zelle. Spektrum Akademischer Verlag, Heidelberg.
- [2] EIGEN, M., SCHUSTER, P. (1979). THE Hypercycle - A Principle of Natural Self-Organisation. Springer Verlag, Berlin.
- [3] EIGEN, M., (1987). Stufen zum Leben. R. Piper GmbH & Co. KG, München.
- [4] KÜPPERS, B.O., (1986). Der Ursprung biologischer Information - Zur Naturphilosophie der Lebensentstehung. R. Piper GmbH & Co. KG, München.
- [5] REINBOTHE, H., KRAUß, G.J., (1982). Entstehung und molekulare Evolution des Lebens. VEB Gustav Fischer Verlag Jena, Halle.

Kapitel 9

Die Erde lebt - James Lovelocks Gaia-Theorie

Alexander Meyer
figure 21. Juni 2000

9.1 Die Gaia - Theorie

9.1.1 Was ist Gaia ?

Die Gaia - Theorie beschäftigt sich mit unserem Planeten - der Erde. Vielmehr handelt es sich um eine spezielle Sichtweise die Erde und die auf ihr lebenden Organismen zu betrachten. Die Gaia - Theorie liefert folgende Definition für Gaia:

Gaia ist die Erde als ein durchgängiges physiologisches System, eine Entität, die zumindest in dem Sinne lebendig ist, als sie wie jeder Organismus ihren Stoffwechsel und ihre Temperatur selbst regelt und in den mehr oder weniger engen Grenzen hält, in den das Leben bestehen kann. ([2], S.10)

Die Erde wird somit als eine Art Lebewesen verstanden, womit natürlich nicht gemeint ist, dass Gaia z.B. einen eigenen Willen hat, sondern den Begriff eher eine metaphorische Bedeutung zukommt. Später werden wir noch sehen, dass Gaia eigentlich alle Kennzeichen für ein Lebewesen erfüllt.

Die erste Definition ist noch sehr allgemein gehalten. In der nächsten wird genauer auf die Bestandteile von Gaia eingegangen, die zur nächsten zentralen Annahme der Gaia - Theorie führt:

Gaia ist ein evolvierendes System, bestehend aus allem Lebendigen und seiner Oberflächenumwelt, den Meeren, der Atmosphäre, dem Krustengestein, wobei diese beiden Komponenten fest verkoppelt und nicht voneinander zu trennen sind. ([2],S.11)

Die Organismen der Erde sind also Teil Gaias - ja spielen so sogar eine zentrale Rolle - und stehen in Beziehung zur Umwelt. Im weiteren Verlauf werden einige zentralen, für

die Existenz des Lebens unerlässliche, Vorgänge beschrieben, bei denen Organismen mit deren Umwelt koagieren.

Die Verbindungen sind so stark, dass sie Einfluss auf die Entwicklung des Lebensumfeld der Erde ausgeübt haben und ausüben werden.

Entscheidend bei der Gaia - Theorie ist die globale Perspektive. Demnach ist die Gesundheit des gesamten Organismus entscheidend und nicht irgendwelcher Organismen - wie z.B. der Mensch. Damit weicht die Theorie von den klassischen Umweltbewegungen ab, für die die Gesundheit des Menschen im Mittelpunkt steht.

9.1.2 Begründer

Begründer der Gaia - Theorie ist James Lovelock (1919). Er promovierte in Medizin und war Dozent in Yale und Harvard. Heute arbeitet er als freier Wissenschaftler in England. Doch es gab auch schon vor Lovelock Wissenschaftler, die ähnliche Sichtweisen vertraten. Vor allem ist dort James Hutton zu nennen, auf den sich auch Lovelock bezieht. Er verglich bereits 1788 den Wasserkreislauf der Erde mit dem Blutkreislauf des Menschen. Weiterhin bezeichnete er die Erde als einen über-Organismus.

9.1.3 Entstehung

Anfang der 60er Jahre wurde James Lovelock von der NASA gebeten bei einer Mission zur Erforschung von Leben teilzunehmen. Schon damals war er der Ansicht, dass man das chemische Gefüge des Mars oder eines Planeten allgemein betrachten sollte, um Aussagen über Leben machen zu können, anstatt an einer lokalen (Lande-)Stelle nach Leben zu suchen. Der Grund dieser Annahme war, dass Leben die Atmosphäre und die Ozeane als Stofflieferant und Aufnahmespeicher seiner Stoffwechselprodukte benutzt, was zu einer Veränderung der Atmosphäre führen würde. Man müsste also nur die Atmosphäre des Planeten untersuchen. Bei einem toten Planeten würde man demnach eine Zusammensetzung der Luft erwarten, die nahe dem chemischen Gleichgewichtes ist. Folgende Tabelle zeigt die Zusammensetzung der Atmosphären der Venus, des Mars, der Erde und die fiktive Atmosphäre der unbelebten Erde.

Tab. 1: Atmosphären der Planeten und ihre Zusammensetzung

| Planet | | | | |
|-----------------|---------|-----------------|---------|---------|
| Gas | Venus | Erde ohne Leben | Mars | Erde |
| Kohlendioxid | 96,5 | 98 | 95 | 0,03 |
| Stickstoff | 3,5 | 1,9 | 2,7 | 79 |
| Sauerstoff | Gering | 0,0 | 0,13 | 21 |
| Argon | 70 ppm | 0,1 | 1,6 | 1 |
| Methan | 0,0 ppm | 0,0 ppm | 0,0 ppm | 1,7 ppm |
| Temperatur in C | 459 | 240-340 | -53 | 13 |
| Druck in Bar | 90 | 60 | 0,0064 | 1,0 |

Das war der Beginn der Gaia - Theorie. Leben scheint das Klima und den chemischen Zustand des Planeten ändern zu können. Man kann in der oberen Tabellen erkennen, dass die Zusammensetzung der Erdatmosphäre weit von denen des Mars und der Venus unterscheiden, deren wiederum fast mit dem chemischen Gleichgewicht übereinstimmt.

9.1.4 Leben

In den vorherigen Teilabschnitten kam häufiger der Begriff Leben vor. Doch was genau ist Leben ? Eine genaue Definition des so alltäglichen Begriffes ist bei genauerer Betrachtung nicht so leicht zu geben. Je nach Fachrichtung des Wissenschaftlers würden unterschiedliche Definitionsversuche zu Tage treten.

Ein neodarwinistischer Biologe definiert ein Lebewesen als einen zur Fortpflanzung fähigen Organismus, der eventuelle Reproduktionsfehler - sprich Mutationen - durch natürliche Auslese beheben kann.

Ein Biochemiker würde einen Organismus als lebendig definieren, der Sonnenlicht oder potentielle chemische Energie als Nahrung aufnimmt und diese Energien für das von den Genen gesteuerte Wachstum verwendet.

Der Physiker wählt als Ansatzpunkt die Hauptsätze der Thermodynamik. Der erste besagt, dass bei einer Umwandlung von einer Energieform in eine andere weder Energie verloren noch gewonnen wird. Die Energie bleibt also stets erhalten. Der zweite besagt, dass bei einer Umwandlung von einer Energieform in eine andere immer ein Anteil von Wärme abstrahlt.

Daraus folgt, dass letztlich alle Energie zu Wärme wird. Damit streben alle Systeme einem Gleichgewichtszustand an, in dem aus dem System keine Energie mehr entnommen werden kann. In diesem Zusammenhang kommt auch der Begriff der Entropie ins Spiel. Entropie (Unordnung) ist eine physikalische Größe, welches ein Maß der Annäherung zum Gleichgewicht darstellt. Je höher die Entropie desto größer die Annäherung zum Gleichgewicht. Ein Lebewesen hat eine niedrige Entropie. Es benötigt also ständig Energie, um dem zweiten Hauptsatz der Hauptsatz entgegenzuwirken.

Die Definition eines Geophysiologen (die Systemwissenschaft der Erde) baut auf den anderen auf.

Der Geophysiologe definiert einen lebendigen Organismus als ein umgrenztes, den Fluss von Materie und Energie erlaubendes System, das bei seinem inneren Medium die Konstanz der Zusammensetzung aufrechterhält und seinen physikalischen Zustand unter wechselnden Umweltbedingungen wahren kann; er ist fähig seine Homöostase zu wahren. ([2],S.29)

Nach dieser Definition wäre Gaia ein Lebewesen. Aber auch nach denen vom Physiker und vom Biochemiker wäre Gaia lebendig.

Nur der neodarwinistische Biologe widerspricht dieser Ansicht. Da Gaia sich nicht fortpflanzt, ist kein Lebewesen. Aus diesen Reihen gibt es auch die meiste Kritik an der Gaia - Theorie.

Zum Abschluss dieses Abschnittes folgt noch ein kleiner Vergleich zwischen Gaia und anderen Lebewesen:

Tab. 2: Vergleich von Kennzeichen biologischer Organismen mit Ökosystemen

| Kennzeichen | Lebensform bzw. Ökosystem | | | | |
|---------------|---------------------------|-----------|------|-------------|------|
| | Bakterie | Säugetier | Baum | Bienenstaat | Gaia |
| Vermehrung | + | + | + | - | - |
| Stoffwechsel | + | + | + | + | + |
| Evolution | + | + | + | + | + |
| Thermostase | - | + | - | + | + |
| Chemostase | + | + | + | - | + |
| Selbstheilung | + | + | + | + | + |

9.1.5 Kritik

Es gibt zwei Hauptkritikpunkte gegen James Lovelocks Gaia - Theorie. Als erstes wäre der Vorwurf der Teleologie, was soviel bedeutet, dass es eine Vorrausplanung bzw. Intelligenz geben müsste, damit z.B. der chemische Haushalt reguliert werden kann.

Der zweite Kritikpunkt bemängelt, dass die Gaia - Theorie die biologische Regulierung nur von einer Seite betrachtet. In der realen Welt hingegen gibt es eine Co-Evolution von Lebenden und Nichtlebenden.

Im weiteren möchte ich auf den ersten Kritikpunkt eingehen. Lovelock hält den Vorwurf, dass das Gaia - Prinzip teleologisch für falsch. Und er hat es sogar bewiesen. Dabei bediente er sich des Daisyworld - Modells, auf welches ich jetzt zu sprechen kommen möchte.

9.2 Das Daisyworld - Modell

9.2.1 Modellannahmen

Das Daisyworld - Modell soll ein vereinfachtes Modell der Erde darstellen. Auf Daisyworld gibt es jedoch nur Gänseblümchen - daher der Name. Die genauen Modellannahmen sind folgende:

- Größe wie die Erde
- Identische Achsendrehung und Umlaufbahn um die Sonne wie die Erde
- Die Sonne besitzt die gleiche Leuchtkraft und Masse wie unsere Sonne
- Die Wärmeentwicklung der Sonne wird mit zunehmendem Alter größer (wie bei unserer Sonne)
- Es gibt es etwas mehr Landfläche als auf der Erde
- Die Wasserversorgung ist jedoch überall gut
- Es können überall auf dem Land Gänseblümchen wachsen, wenn das entsprechende Klima herrscht
- Unterhalb von 5 C können keine Gänseblümchen wachsen; über 40 C gehen sie ein; Optimaltemperatur ist 20 C

- Mittlere Durchschnittstemperatur wird allein von Albedo bestimmt
- Der unbewachsene Boden hat ein Albedo von 0,4

Albedo ist der Fachausdruck für durchschnittliche Farbschattierung eines Planeten. Von einem niedrigen Albedo sprechen wir, wenn die Tönung der Planetenoberfläche dunkel ist. Dann absorbiert der Planet mehr Wärme aus dem Sonnenlicht, so dass sich die Oberfläche erhitzt. Bei heller Planetenoberfläche wird hingegen ein Grossteil des Sonnenlichtes reflektiert. Diese Oberfläche wäre kälter als die dunkle. Der Wertebereich des Albedos erstreckt sich von 0 (schwarz) bis 1 (weiß).

9.2.2 Modell mit zwei Arten

Zu Beginn betrachten wir das Modell mit nur zwei Gänseblümchenarten. Es gibt eine helle Art (Albedo 0,7) und eine dunkle Art (Albedo 0,2). Das Modell lieferte folgendes Ergebnis:

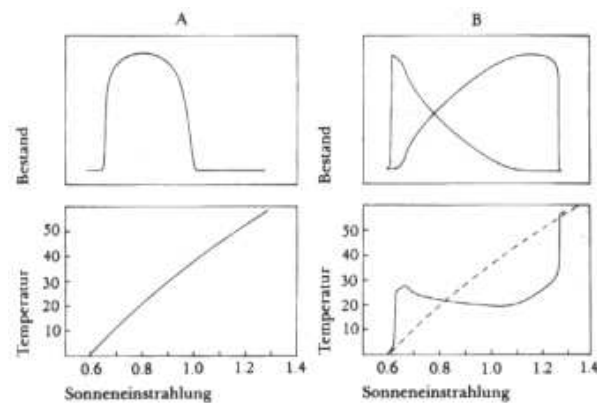


Abbildung 9.1: Entwicklung des Daisyworld-Modells nach konventioneller Wissenschaft (A) und Geophysiologie (B)

Die Entwicklung zeigt eindeutig, dass das Klima durch die Gänseblümchen reguliert worden ist. Wie ist das geschehen ?

Zu Beginn war die Sonneneinstrahlung noch sehr schwach. Die ersten Gänseblümchen konnten sich daher nur am Äquator ansiedeln, da nur dort die mittlere Temperatur über 5°C lag. Zu Beginn waren noch beide Arten gleich oft vertreten, doch sehr schnell wuchs die Population der Dunklen. Dadurch erhöhte sich die Sonnenlichtabsorption an ihren Standorten - es wurde also wärmer. Die hellen Gänseblümchen jedoch gingen ein, da ihre hohe Reflektionsrate die Temperaturen unter 5°C trieb.

In den folgenden Jahren breiteten sich die dunklen Gänseblümchen immer weiter aus, bis schließlich fast der ganze Planet mit ihnen bedeckt war. Die Temperatur war inzwischen über die obere Wachstumsgrenze gestiegen. Dadurch vermehrten sich dann die hellen Gänseblümchen, die zur Abkühlung beitrugen.

Im folgenden wird der Sonneneinstrahlung immer stärker, so dass sich immer mehr Helle entwickeln. Am Ende wird die Einstrahlung so stark, dass die Vegetation die Oberflächentemperatur nicht mehr unter 40°C halten kann und alle Gänseblümchen gehen ein. Damit kann man den im vorherigen Kapitel Vorwurf der Teleologie schon als widerlegt betrachten.

Aus diesem Modell hat Lovelock weitere Erkenntnisse gezogen. Er fasst die Bestandteile von Gaia in vier Punkten zusammen:

1. Lebende Organismen, die alle Ressourcen zum Wachstum benutzen
2. Organismen, die der natürlichen Selektion unterworfen sind
3. Organismen, die ihre physikalische und chemische Umgebung beeinflussen
4. Grenzen des Lebens

9.2.3 Modelle mit mehreren Arten

Das Modell im ersten Abschnitt war von Lovelock entworfen worden, um den Vorwurf der Teleologie zu entkräften, was ihm ja auch gelungen ist. Doch die Kritiker gaben nicht so schnell auf. Sie behaupteten, dass die Gänseblümchen einen Teil ihrer Energie für die Pigmentbildung verbrauchen würden und damit einen Nachteil gegenüber unpigmentierten Gänseblümchen hätten. Damit würde dann auch keine Temperaturregulierung geben. Daher erweiterte Lovelock sein Modell um eine Gänseblümchenart ohne Pigmente und gab dieser Art eine etwas höhere Entwicklungsrate. Wie die Graphik unten zeigt, änderte sich überhaupt nichts an der Fähigkeit der Temperaturregulierung.

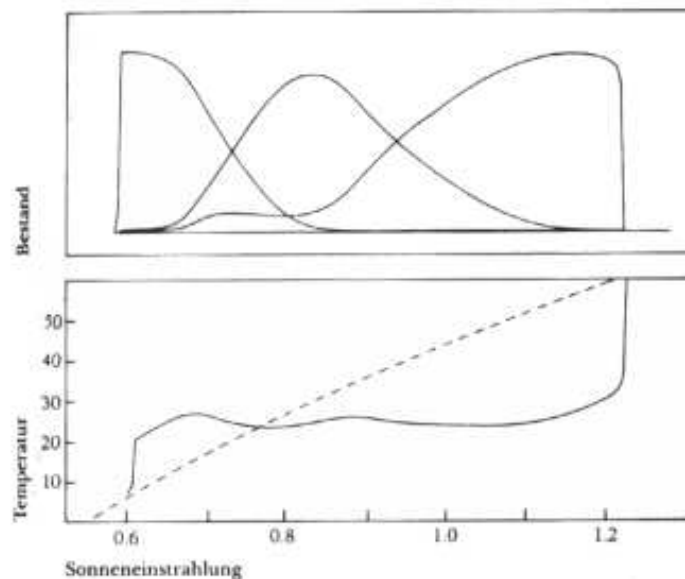


Abbildung 9.2: Entwicklung des Daisyworld-Klimas bei drei Gänseblümchenarten

Die dunklen Gänseblümchen gedeihen wieder bei Kälte, die hellen bei Hitze und die grauen, wenn keine Regulierung notwendig ist.

Damit war auch dieser Vorwurf entkräftet. Es bot sich jedoch an dieser Stelle an das Modell auf beliebig viele Gänseblümchenarten zu erweitern.

Im folgenden sind die Ergebnisse des Daisyworld - Modells mit 20 verschiedenen Gänseblümchenarten aufgeführt:

Dabei ist zu erkennen, dass die Artenvielfalt am größten ist, wenn die Belastung am geringsten ist, d.h. die Gänseblümchen sich nicht um die Temperaturregulierung bemühen müssen. Das ist zugleich der ideale Wachstumszeitpunkt. Bei starker Belastung ist die Vielfalt am geringsten. Es gibt fast nur helle bzw. dunkle Gänseblümchen.

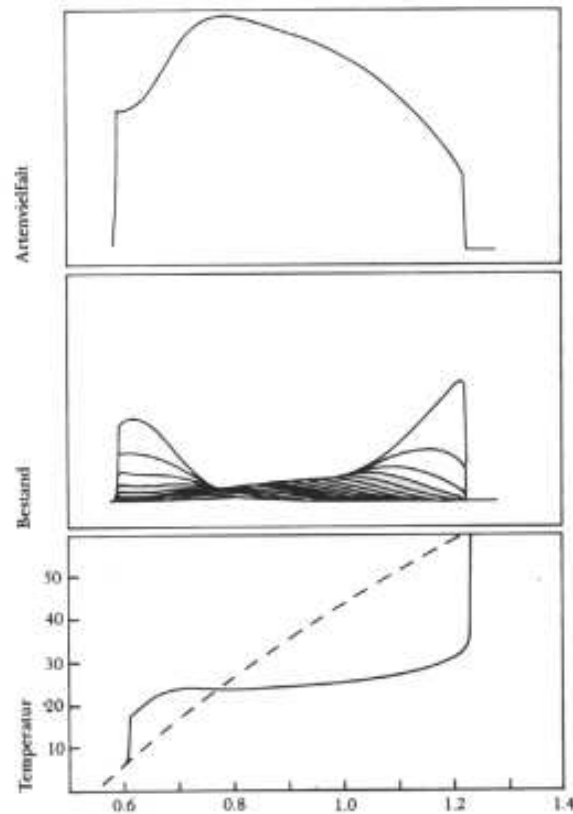


Abbildung 9.3: Entwicklung des Daisyworld-Klimas bei zwanzig Gänseblümchenarten

9.2.4 Stabilität des Daisyworlds - Modells

Lovelock war von seinem Daisyworld - Modell sehr beeindruckt, da es sich durch eine große Stabilität auszeichnete. Der Grund dafür war das Koagieren von Umwelt und Vegetation. Es kann niemals zu einem unkontrollierten Wachstum der Gänseblümchen kommen, da durch eine weite Verbreitung sich das Lebensumfeld verschlechtert und das Wachstum bremst. Umgedreht führt die Existenz der Vegetation zu konstanten Umgebungsbedingungen, durch das Wachsen der Gänseblümchen in der passenden Farbe. Um die Stabilität des Systems weiter zu testen, führte Lovelock noch eine Reihe von weiteren Versuchen durch.

Zuerst kombinierte er das Daisyworld - Modells mit dem bekannten Lotka - Volterra - Modell durch. Die Gänseblümchen wurden also von Kaninchen gefressen, die wiederum Beute von Füchse waren. Zusätzlich führte er vier Katastrophen ein, in dem jeweils 30% der Population vernichtet werden.

Auch diese neuen Störungen konnten die Stabilität des Modells beeinflussen.

In den bisherigen Beispielen stieg die Sonnenenergie kontinuierlich an. In folgenden überprüfte Lovelock sein Modell bei konstanter Wärmerate auf Stabilität. Er verwendete dabei 10 Gänseblümchenarten, die durch eine einmalig auftretende Katastrophe zu 60% dezimiert werden.

Die gestrichelte Linie gibt dabei die Temperatur ohne die Gänseblümchen an. Nach Ende der Störung kehrt das System schnell wieder in den Gleichgewichtszustand wieder zurück, da die Gänseblümchen sich nach der Katastrophe durch rasantes Wachstum die Temperaturschwankung auffangen.

In einem weiteren Modell kombinierte Lovelock die Störungsexperimente mit den den Ver-

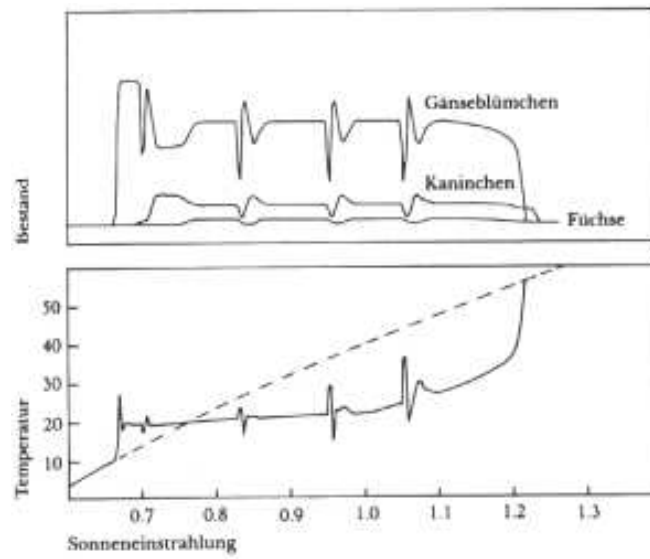


Abbildung 9.4: Daisyworld mit Füchsen, Kaninchen und vier Krankheiten, die 30% der Gäseblümchen vernichten

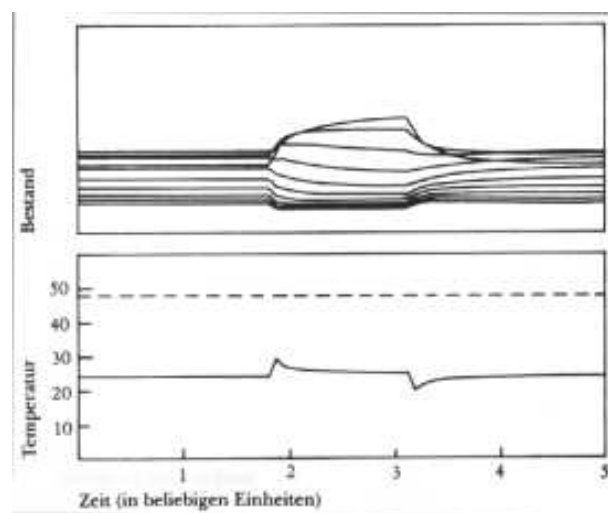


Abbildung 9.5: Klimaauswirkung einer Krankheit, an der 60% der Gäseblümchen zu Grunde gehen

suchen der stetigen Veränderung. Er setzte 10 Gänseblümchenarten einer wiederkehrenden Krankheit aus. 10% der Gänseblümchenpopulation sterben dabei jeweils. Die Sonnenenergie steigt wiederum im Verlaufe des Experimentes.

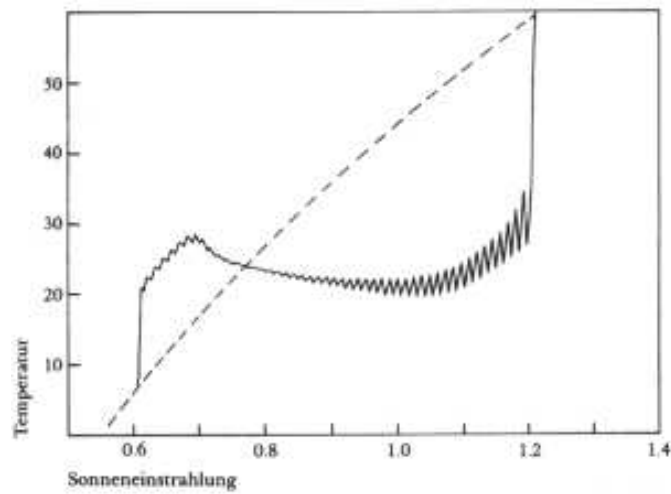


Abbildung 9.6: Auswirkungen einer periodisch auftretenden Krankheit auf das Klima

Man kann erkennen, dass die Temperaturschwankungen genau dann am stärksten sind, wenn die Anzahl der Arten am geringsten ist.

9.2.5 Gesamtergebnis

Die wichtigsten Gesichtspunkte vom Daisyworld - Modell sind in folgender Darstellung noch einmal zusammengefasst.

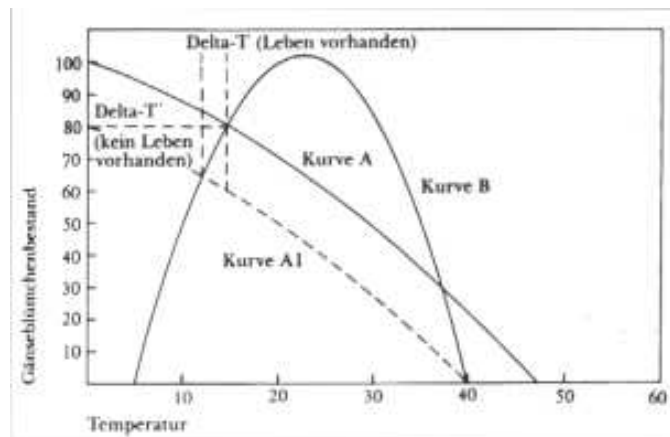


Abbildung 9.7: Regulierung der Klimas durch weiße Gänseblümchen

Grundlage hierfür ist das Daisyworld - Modell mit nur einer Gänseblümchenart - und zwar weißen. In ihren Vegetationsgebieten ist es also kühler als auf unbewachsenen, da sie das Albedo des Planeten dort erhöhen .

Kurve A zeigt das Verhältnis zwischen dem von den Gänseblümchen bewachsenen Gebiet und der Oberflächentemperatur. A1 zeigt das dasselbe Verhältnis bei Veränderung eines Parameters , wie z.B. die Verringerung der Sonneneinstrahlung.

Kurve B drückt das Verhältnis zwischen der Oberflächentemperatur und dem Gänseblümenbestand aus.

Die Schnittpunkte sind die möglichen stabilen Zustände. Es stellt sich aber heraus, dass nur der linke stabil ist, da dort Veränderung - sprich die Ableitung - unterschiedliche Vorzeichen haben. Bei jeder im Wachstumsspielraum (5 - 40 C) liegende Anfangstemperatur strebt die Planetentemperatur auf diesen Schnittpunkt zu und bleibt dort.

Aus den Daisyworld - Modellen entwickelte Lovelock 4 Kernpunkte über die Sichtweise der Gaia - Theorie der Erde gegenüber:

1. *Leben ist eine Erscheinung planetarer Größenordnung. In dieser Größenordnung ist es nahezu unsterblich und braucht sich daher nicht fortzupflanzen.*
2. *Eine partielle Besetzung eines Planeten durch lebende Organismen kann es nicht geben. Ein solches Phänomen wäre genauso kurzlebig wie ein halbes Lebewesen. Zur Regulierung der Lebensumwelt braucht ein Planet eine genügende Anzahl lebender Organismen. Bei einer unvollständigen Verteilung der Organismen würden ihn die unvermeidbaren Kräfte physikalischer und chemischer Entwicklung bald unbewohnbar machen.*
3. *Die gängige Auslegung von Darwins großartigen Thesen muss verändert werden. Gaia weist auf die Fehlerhaftigkeit der Anpassungstheorie hin. Die Aussage, dass sich Organismen, die besser als andere angepasst sind, mit größerer Wahrscheinlichkeit fortpflanzen, genügt nicht mehr. Man muss hinzufügen, dass das Wachstum eines Organismus seine physikalische und chemische Umgebung beeinflusst. Die Evolution der Arten und die Evolution der Gesteine sind daher in einem gemeinsamen und unteilbaren Prozess eng miteinander verbunden.*
4. *Die theoretische Ökologie wird vergrößert. Wenn man die Arten und ihre physikalische Umgebung zu einem einzigen System zusammenfasst, kann man erstmals ökologische Modelle erstellen, die mathematisch stabil und dennoch eine große Anzahl untereinander konkurrierender Arten umfassen. Eine größere Artenvielfalt in diesen Modellen führt zu einer besseren Regulierung.*
([1],S.95f)

9.3 Die Epigenese aus der Sicht der Gaia - Theorie

9.3.1 Archaikum

Im weiteren befasst wird die Entwicklung der Erde in der Frühzeit dargestellt. Da diese Zeit schon Milliarden von Jahren zurückliegt, wissen wir nur sehr wenig über diese Zeit. Viele Wissenschaftler stellten Theorien über den Verlauf der Entwicklung der Erde aus. Dabei interessierten sie sich vor allem auf die chemische Zusammensetzung der Atmosphäre, die Temperaturverhältnisse und ähnliche Dinge.

Auch James Lovelock hat sich mit dieser Fragestellung auseinandergesetzt. Er betrachtete die Entwicklung der Erde aus der Sicht der Gaia - Theorie. Weiterhin suchte er weitere Beweise für seine Theorie.

Die Erde entstand vor etwa 4,55 äonen durch eine Supernova. Das gilt als gesicherte These, da die Erde auch heute noch radioaktive Strahlung aufweist und außerdem aus Stoffen wie

Eisen, Silizium und Sauerstoff besteht, die nicht aus normalen Prozessen hätten entstehen können.

Das Leben entstand etwa eine Äone später, was durch den unterschiedlichen Anteil von Atomen des stabilen Kohlenstoffs in den Gesteinen gezeigt werden kann. Doch darüber wie es zu dieser Zeit auf der Erde aussah, wissen wir nur sehr wenig. Doch aus den spärlichen Informationen kann man doch einige Erkenntnisse ziehen.

Wie auch schon in vorherigen Abschnitten ist die Atmosphäre ein entscheidendes Charakterisierungsmerkmal eines Planeten. Beginnen wir also mit der Atmosphäre.

Wie im ersten Abschnitt schon ausgezeigt worden ist, befindet sich die Atmosphäre eines toten Planeten in der Nähe des chemischen Gleichgewichtes. So sah es auch auf der Erde aus, bevor das Leben begann. Man vermutet, dass alle chemischen Komponenten, die für Leben benötigt werden, bereits von Beginn an auf der Erde waren. Das waren die Stoffe, die wir heute als organisch bezeichnen, wie z.B. Aminosäuren und Zucker. Doch nur das Vorhandensein dieser Stoffe führt nicht automatisch zu Leben. Auch die chemischen Voraussetzungen bzw. das Klima muss stimmen, damit die Entwicklung beginnen kann.

Aus Gesteinsproben dieser Zeit kann man schließen, dass damals Wasser vorhanden war. Damit aber Leben entstehen konnte musste weiterhin eine Temperatur zwischen 0 - 50 Celsius geherrscht haben, da sonst das Wasser entweder gefroren wäre oder gekocht hätte. Ein Klimatologe hat errechnet, dass die Temperatur in etwa so hoch wie heute war. Das ist sehr erstaunlich, da die Sonne weit weniger Energie abstrahlte als heute. In der Astrophysik hat man festgestellt, dass die Sonne seit Beginn des Lebens ihren Energieausstoß um 25 Prozent vermehrt hat, ein typisches Phänomen alternder Sterne.

Man vermutet, dass die Erde sich durch eine Art Treibhauseffekt warmgehalten hat. Atmosphärgase, die mehr aus mehr als zwei Atomen bestehen, zu denen Wasserdampf, Kohlendioxid und Ammoniak gehören, haben die Eigenschaft, dass sie die langwellige und zugleich die energiereiche Strahlungsenergie der Sonne durchlassen. Diese erwärmen die Erdoberfläche. Hingegen können die langwelligen Infrarotstrahlung die von der Erdoberfläche ausgeht diese Gasschicht nicht durchdringen. Dadurch wird eine Erwärmung der Erdoberfläche erreicht.

Allerdings sollte man den Faktor Vulkane bei der Wärmeerzeugung nicht unterschätzen. Zur damaligen Zeit war die Radioaktivität der Erde noch wesentlich höher als heute, was zu einer großen Wärmeentwicklung im Erdinneren führte, die durch die Vulkane nach außen drang.

Nach soviel Vorrede über die Entstehung des Lebens und die Umweltbedingungen im Archaikum, möchte ich jetzt zu Lovelocks Modell des Archaikums kommen. Wie nicht anders zu erwarten, zeigt es eine enge Wechselwirkung zwischen Leben und der Umwelt auf.

Die ersten Lebewesen waren Einzeller, die sich vor allem von chemischen Stoffen ernährten, die im Überfluss vorhanden waren. Doch relativ schnell erkannten diese Lebewesen die Sonnenenergie als Nahrungsquelle zu nutzen, also Photosynthese zu betreiben. Die ersten Bakterien, die dieses als erstes vollbrachten, waren die Cyanobacteria (Blaualgen). Sie bildeten eine zentrale Rolle im planetarischen System des Archaikums. Im Sonnenlicht an der Erdoberfläche lebten die Cyanobacteria, die Photosynthese betrieben und sich vermehrten. Ihr Abfallprodukt - der Sauerstoff - konnte sich zu diesem Zeitpunkt noch nicht weiter verbreiten, da er sofort mit nichtorganischen Stoffen reagierte. Weiterhin gab es die Methanogene, die die Produkte der Cyanobacteria bzw. abgestorbene Exemplare dieser Spezies beseitigten. Sie sorgten dadurch indirekt dafür, dass der Kohlenstoff als Methan und Kohlenstoffdioxid wieder in die Umgebung abgegeben wurde.

Diese Organismen entwickelten ein stabiles Selbstregulierungssystem mit ihrer Umwelt - was auch dringend notwendig war. Die Cyanobacteria - als Photosyntheseanwender - verbrauchten den Kohlenstoffdioxid aus der Atmosphäre, der wie oben erwähnt, die Erde vor der Abkühlung bewahrt. Hätten die Methanogene diesem nicht entgegengewirkt, wäre die Temperatur schnell unter den Gefrierpunkt gesunken, womit dann das Leben ausgestorben wäre. Aufgrund des Sedimentgesteine lässt sich beweisen, dass niemals eine solche Katastrophe eingetreten ist.

Wie oben erwähnt produzierten die Methanogene Methan, welches aufgrund der großen Population in Mengen produziert wurde. Doch was geschieht mit dem Methan? Lovelock hat auch hierzu eine Hypothese. Das Methan in der Atmosphäre ist sehr instabil und wird durch UV-Licht gespalten. Die entstehenden Stoffe verbinden sich mit anderen Molekülen zu komplexen chemischen Verbindungen, die eine Art Smog hätten bilden können, welcher die Möglichkeit der Absorption von UV-Strahlung und langwelligem Licht hätte - eine Ozonschicht aus Methan also.

Dieses System hat Lovelock aufbauend auf dem Daisyworld - Modell modelliert. Die Photosynthese - Organismen lassen sich mit den weißen Gänseblümchen vergleichen, da durch den Verbrauch von Kohlenstoffdioxid die Erde abkühlt. Die Methanogene entsprechen den dunklen Gänseblümchen, da durch den Ausstoß von Kohlenstoffdioxid die Temperatur steigt.

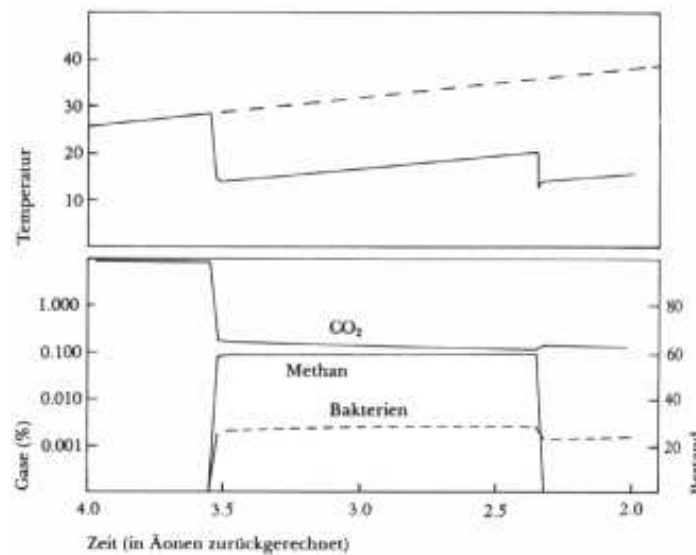


Abbildung 9.8: Temperaturverlauf, Konzentration der Atmosphärgase und Bakterienbestand im Archaikum-Modell

Das obere Diagramm zeigt den Temperaturverlauf mit und ohne Leben (gestrichelte Linie) an. Im zweiten ist die Konzentration an Kohlenstoffdioxid und Methan sowie der Bestand an Bakterien dargestellt.

Auffällig ist der rapide Temperatursturz als das Leben begann, was auf den Rückgang der Kohlenstoffdioxidkonzentration zurückzuführen ist, welcher durch die Entstehung der Bakterien herbeigeführt worden war, da sie es für den Aufbau des Körpers benötigten. Das Gefrieren wird durch das Treibhausgas Methan abgefangen.

Zum Vergleich hier noch eine Tabelle zum Vergleich der Atmosphäre vor und nach der Entstehung des Lebens:

Tab. 3: Eine Schätzung zur Zusammensetzung der Atmosphäre im Archaikum vor und nach der Entstehung des Lebens

| Gas | ohne Leben | mit Leben |
|--------------|------------|-------------|
| Kohlendioxid | dominant | 0,3 Prozent |
| Stickstoff | unbekannt | 99 Prozent |
| Sauerstoff | 0 | 1 ppm |
| Methan | 0 | 100 ppm |
| Wasserstoff | etwas | 1 ppm |

Zurück zu den Diagrammen. Bei ca. 2.3 Äonen ist ein weiterer Temperatursturz zu verzeichnen. An diesem Zeitpunkt endet das Archaikum, das Proterozoikum beginnt. Man kann nachweisen, dass etwa um diesen Zeitpunkt herum eine Eiszeit stattgefunden hat. Worauf ist dieser Temperatursturz zurückzuführen? Es war das Auftreten von Sauerstoff. Dadurch dass die Bakterien im Archaikum einen kleinen Teil ihres Kohlenstoffumsatzes immer ablagerten, kam es zu einem stetigen überschuss an Sauerstoff. Zunächst wurde er, wie oben bereits erwähnt, bei der Oxidation der Reduktionsmittel aufgebraucht. Doch irgendwann war mehr Sauerstoff da, so dass er sich in der Atmosphäre anreichern konnte. Irgendwann war der Sauerstoff in einer solchen Menge vorhanden, dass das reduzierende Methan nicht mehr ins Gewicht fiel.

Ein neues Zeitalter begann.

9.3.2 Proterozoikum

Das Proterozoikum stellte einen neuen Zeitabschnitt im Verlauf der Erde dar. Da die Atmosphärenzusammensetzung sich in Richtung des Sauerstoffs verschoben hatte, konnten sich neue Formen von Leben entwickeln. Im Archaikum herrschten die Prokaryoten vor. Im Proterozoikum kam es zur Entwicklung von Eukaryoten, die im Gegensatz zu Prokaryoten Organellen und Zellkerne besaßen. Die neuen sauerstoffliebenden Einzeller lebten zunächst in Koexistenz mit den photosynthesebetreibenden Lebewesen. Doch sie lernten schnell, dass das Fressen von Prokaryoten eine gute Nahrungsquelle darstellte. Da die Eukaryoten aber keinen Mund besaßen, verwandten sie die sogenannte Phagozytose (Beutezelle mit eigener Membran umschließen). Doch anstatt der jetzt folgenden Verdauung, kann die eingeschlossene Zelle seinen Wirt angreifen (z.B. Leprabakterien). Dieser Rückschlag kann aber auch zur Symbiose führen. Somit kann man die Cyanobacteria als Vorläufer unserer heutigen Chloroplasten bezeichnen.

Diese komplexeren Organismen waren sehr viel widerstandsfähiger als die Prokaryoten. Doch durch die Komplexität ergaben sich auch Nachteile. Vor allem ist das Alterungsproblem zu nennen. Da die Informationen der Gene nicht verloren gehen sollte, musste sich ein neues Verfahren entwickeln, um dieses zu Verhindern, welches in der Geschlechtertrennung bzw. die daraus resultierende Sexualität gefunden wurde.

Jetzt möchte noch einmal auf den Sauerstoff und der daraus entstehenden Entwicklung zu sprechen kommen. Wie hoch war der Sauerstoffgehalt im Proterozoikum? Eine durchaus wichtige Frage. Im Archaikum, wie auch in obiger Tabelle zu sehen, betrug der Sauerstoffgehalt 1ppm. Man vermutet, dass der Gehalt in relativ kurzer Zeit auf etwa 1% gestiegen ist. Auf jeden Fall wissen wir, dass Sauerstoff im Proterozoikum dominant war,

was bedeutet, dass die Sauerstoffkonzentration mindestens im Verhältnis 2:1 zur Methankonzentration steht.

Die Sauerstoffkonzentration steht im Zusammenhang mit dem Wachstum der Organismen (durchgezogene Linie). Umgedreht hat aber auch die Anzahl der Organismen Auswirkung auf den Sauerstoffgehalt (gestrichelte Linie). Dieses zeigt auch folgendes Diagramm.

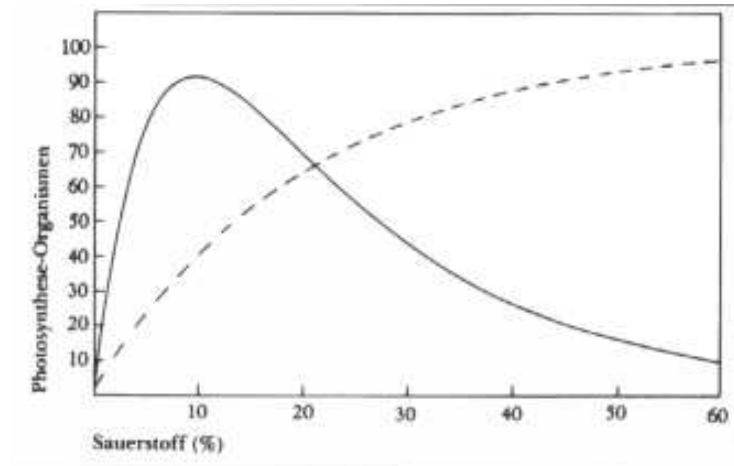


Abbildung 9.9: Auswirkung des Sauerstoffs auf Organismen und umgekehrt

Auch für das Proterozoikum hat Modell erstellt, welches die Verhältnisse dieses Zeitabschnittes aufzeigt.

Im untersten Diagramm ist die Temperaturentwicklung des belebten Planeten mit der eines unbelebten dargestellt, wobei die gestrichelte Linie letzteren charakterisiert. Man kann eine Temperaturregulation erkennen.

Im mittleren Diagramm sind die Konzentration der wichtigsten Atmosphäregase aufgeführt. Man kann das sprunghafte Ansteigen der Sauerstoffkonzentration erkennen.

Im obersten schließlich sind die Populationen der drei großen Klassen - Photosynthese-Organismen (Sauerstoff-Erzeuger), Sauerstoffverbraucher und Methanogene - aufgeführt. Durch das Auftreten des Sauerstoffs konnten sich Sauerstoffverbraucher, aber auch Sauerstoffproduzenten stark vermehren.

9.3.3 Neuzeit

Mit der Neuzeit bezeichnet man den Zeitabschnitt, der vom Ende des Proterozoikum und Beginn des Kambriums vor 600 Mill. Jahren bis heute reicht und wird auch Phanerozoikum genannt.

In diesem Kapitel möchte ich mich mit den Eiszeiten befassen, da James Lovelock eine sehr interessante Sichtweise zu diesem Thema entwickelt hat. Angestoßen wurde er dazu von dem Klimatologen Stephen Schneider, der behauptete, dass Gaia, wenn sie wirklich existiert, Eiszeiten niemals zugelassen hätte.

Lovelock betrachtet die Eiszeiten als bevorzugten Zustand der Erde und die zwischenzeitlichen Wärmeperioden, in so einer wir gerade leben, als eine Art Fieberanfall, der auf zunehmendes Alter Gaias und damit der nicht mehr hundertprozentig funktionierenden Temperaturregulierung hindeutet. Auch das Daisyworldmodell reagiert bei zunehmender Erwärmung mit stärkeren Temperaturschwankungen.

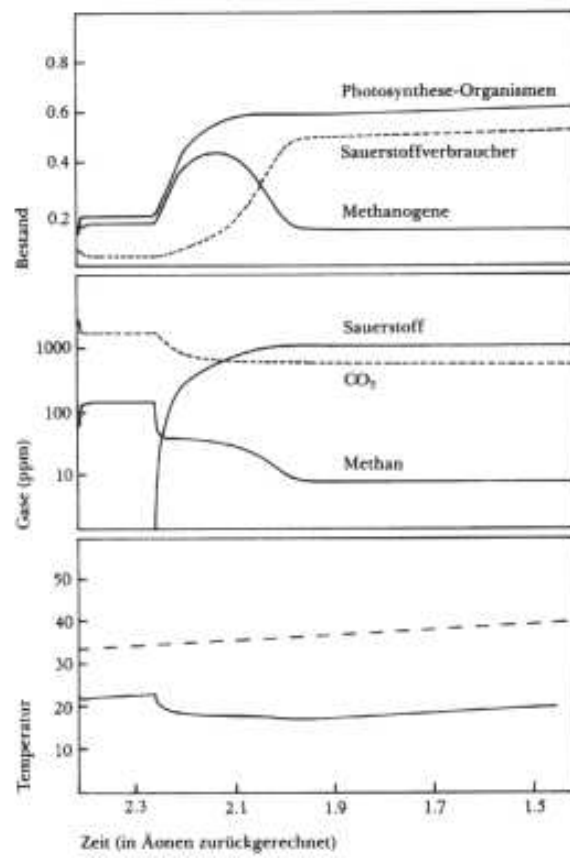


Abbildung 9.10: Ergebnisse des Archaikum/Proterozoikums-Modells

Wir wissen nicht genau wie die Eiszeiten entstanden sind. Es ist allerdings bekannt, dass sie periodisch auftreten. Auslöser könnte der nach seinem Entdecker benannte Milankovic - Effekt sein. Der Jugoslawe entdeckte, dass die Veränderungen der Erdumlaufbahn und die Lage der Erdachse eine unterschiedliche Wärmezufuhr der Sonne zur Erde bewirken. Dieser Milankovic - Zyklus ist identisch mit dem Auftreten der Eiszeiten. Man meint allerdings, dass das alleine nicht ausreichte, um eine Eiszeit hervorzurufen.

Wir wissen, dass während der Eiszeit ein sehr geringer Kohlendioxidgehalt in der Atmosphäre war. Lovelock sieht die Existenz von mehr Organismen als während der Zwischenzeiten als einzige logische Erklärung dafür ! Das widerspricht unserem bisherigen Bild der Eiszeiten, da die Eisdecken Teile des Landes bedeckten. Es ist aber der Gegenteil der Fall. Durch die Entstehung der Gletscher dürften der Meeresspiegel aber um einige hundert Meter gefallen seien, so dass die großen Flächen des Kontinentalschelfs Land waren, die mit tropischen Wäldern bewachsen waren.

9.4 Physiologie aus Sicht der Gaia - Theorie

9.4.1 Schwefel-Kreislauf

In diesem Kapitel wollen wir uns mit essentiellen Stoffkreisläufen der gegenwärtigen Erde betrachten. Beginnen wir mit dem Schwefel. Lange Zeit war nicht klar wie er funktionierte. Das Festland verliert ständig Schwefel, da durch die Flüsse Sulfat - Ionen ins Meer geschwemmt werden. Damit ist auf dem Land nicht zu einem Schwefel- Defizit kommt, muss es einen Mechanismus geben, der Schwefel vom Meer zurück zum Festland befördert. Die Theorien besagten bis dorthin, dass der Rücktransport in Form von Schwefelwasserstoff erfolgte.

James Lovelock schien das unwahrscheinlich, da Schwefelwasserstoff im Wasser schnell oxidiert und dann (nach faulen Eiern) stinkt. Er baute sich eine eigene Theorie unter den Gesichtspunkten von Gaia auf.

Der Chemiker Frederik Challenger hatte bereits in den fünfziger Jahren bewiesen, dass viele Meeresorganismen Dimethylsulfid - im weiteren DMS - abgeben. Lovelock konnte dieses bestätigen.

Er fand heraus, dass vor allem die Alge *Polysiphonia fastigata* viel DMS emittierte. In den achtziger Jahren bewies der Chemiker M.O. Andreae , dass die Menge an DMS, die von den Algen abgeben wurde, reichte um den Schwefelkreislauf zu schließen.

Doch man fand noch viel erstaunlichere Sachen heraus: Durch die Emission von DMS entstehen Wolken ! In der Luft wird das DMS schnell zu Schwefelsulfonsäure oxidiert. Diese bildet die Kondensationskerne für den Wasserdampf zur Wolkenbildung.

Dadurch erhalten die Algen ihren Lohn. Die Wolkenbildung führt nämlich zu einer stärkeren Reflexion der Sonnenstrahlen, was eine Absinken der Temperatur und Erhöhung der Windgeschwindigkeit zur Folge hat, welches es stärkeres Aufwühlen des Meeres fördert, wodurch Nährstoffe aus tieferen Regionen des Meeres an die Oberfläche befördert werden und den Algen als Nahrung dienen.

9.4.2 Kalzium

Kalzium ist eines der wichtigsten Elemente, die es gibt. Es wird z.B. zum Aufbau der Knochen und Zähne benötigt. In Form von freien Ionen ist es allerdings so giftig wie

Zyanid. Dieser Stoff führt zu einer weiteren interessanten Theorie, die wieder die enge Verbundenheit von Biosphäre und Umwelt zeigt.

Wie oben erwähnt sind Kalzium - Ionen giftig. Jedoch ist es nicht so unwahrscheinlich, dass es in der Urzeit eine Bakterienart gegeben hat, die in der Lage war Kalzium - Ionen in unlösliches Kalziumkarbonat zu verwandeln. Dieses Kalziumkarbonat hätte sich dann in großen Mengen am Meeresboden abgelagert. Der Geologe Don Anderson stellt in diesem Zusammenhang eine Theorie auf. Er behauptet, dass sich irgendwann eine so große Menge abgelagert hätte, dass sich die chemische Zusammensetzung der Erdkrustengesteine auf dem Meeresboden in der Nähe der Kontinentalränder verändert hätte (Basalt - Eklogit - Phasenübergang). Dadurch änderte sich die physikalischen Eigenschaften so sehr, dass die Plattentektonik einsetzte. Er schrieb in einem Artikel im *Science* 1984:

Die Erde bildet offensichtlich auch darin eine Ausnahme, als sie eine aktive tektonische Plattenverschiebung aufweist. Wenn sich das Kohlendioxid in der Atmosphäre der Venus in Kalkstein umwandeln könnte, würden die Temperaturen an der Oberfläche und im oberen Mantel sinken. Der Basalt - Eklogit - Phasenübergang würde sich an tieferen Stellen fortsetzen, woraufhin der untere Teil der Planetenkruste instabil würde. Aus ebendiesem Grund gibt es Plattenverschiebung auf der Erde: weil sich hier ein Leben entwickelt hat, das Kalkstein erzeugt.

([1],S.146)

Die meisten Geologen halten diese Theorie jedoch für sehr unwahrscheinlich.

9.4.3 Salz

Der Salzgehalt der Meere soll uns im folgenden beschäftigen. Wie hat sich der Salzgehalt in Laufe der Millionen von Jahren entwickelt ? Nur sehr wenige Organismen können einen Salzgehalt von mehr als 6% aushalten. ähnlich wie beim Schwefel - Kreislauf gelangt über Flüsse oder vulkanische Aktivität ständig Salz bzw. Natriumchlorid ins Meerwasser. Man sollte an dieser Stelle erwähnen, dass das Meerwasser nicht nur aus Natriumchlorid, sondern Natrium und Chlorid treten neben anderen Ionen auch separat auf. Die positiven Ionen lagern sich als Sediment wieder ab. Das Problem ist es die Chlorid - Ionen abzubauen. Damit wurde jetzt schon implizit angedeutet, dass der Salzgehalt seit Millionen von Jahren konstant ist. Würde kein Abbau vom Salz erfolgen, wäre bereits nach 800 Millionen Jahren, die kritische Grenze von 5% erreicht worden und in der heutigen Zeit könnte kein Organismus mehr im Meer leben.

Wie werden also die Chlorid - Ionen abgebaut ? Dies geschieht vornehmlich in Lagunen in Küstennähe. In diesen seichten Gewässern erwärmt sich das Wasser schneller und verdunstet. Wasserdampf steigt auf und kondensiert, so dass Regen fällt, welcher ins Meer zurückfließt und zur Verdünnung beiträgt.

Doch was hat das alles mit Gaia zu tun? James Lovelock hat eine Theorie aufgestellt, wodurch Lebewesen aktiv beim Aufbau dieser Lagunen beteiligt sind. Es soll wie folgt vonstatten gehen: Grundlage für den Aufbau der Lagunenbarrieren ist Kalziumkarbonat. Im vorherigen Abschnitt wurde schon gezeigt, wie Kalziumkarbonat entstehen könnte. Diese Ablagerung treten verstärkt in küstennahen Gebieten auf, da dort große Kolonien von Mikroorganismen. Ohne Leben entstehen die Kalksteinablagerungen nur durch die Reaktion von Kalziumsilikatgestein und Kohlendioxid und nachfolgender Spontankristallisation der Kalzium - und Bikarbonat - Ionen. Die Anordnung der Ablagerungen wären

in diesem Fall willkürlich.

Die verstärkte Ablagerung an der Küste führt dann ganz automatisch zur Auftürmung von Kalkstein und zur Abtrennung der Lagunen vom Rest des Meeres. Zu Beginn hätte diese Bildung nur lokalen Einfluss, doch mit der Zeit wird durch die Masse des Kalksteins die Erdkruste verändert, so dass Vertiefungen entstehen und damit die Lagune vergrößern. Wenn man noch weitergeht und Don Andersons Theorie (s. vorheriger Abschnitt) hinzuzieht, könnten die Kalksteinablagerungen die Bildung von Gebirgen und die Auffaltung von Felswänden an den Kontinentalrändern bewirkt haben, wodurch die Küstenlänge wiederum vergrößert hätte, so dass sich noch mehr Verdunstungslagunen bilden können.

9.5 Der Mensch und Gaia

9.5.1 Heutige Probleme

Kohlendioxidanstieg

Ein schwerwiegendes Problem, welches durch den Menschen - genauer gesagt durch die Verbrennung fossiler Brennstoffe - entstanden ist, ist die der Anstieg des Kohlendioxides. Auch der Zeitpunkt ist nicht günstig. Die Erde steht gerade an einem instabilen Punkt im Kohlendioxidkreislauf der Atmosphäre. Durch die zunehmende Sonnenstrahlung musste im Laufe der Millionen Jahre der Kohlendioxid - Gehalt der Atmosphäre immer mehr absinken, um eine Temperaturerhöhung aufzufangen. Bald ist die Untergrenze aber erreicht. Zunehmend treten C4 - Pflanzen auf, die auch bei niedriger Kohlendioxid-Konzentration gedeihen können.

Doch dank des Menschen steigt die Konzentration wieder. Sie wird sich etwa zwischen 2050 und 2100 verdoppelt haben. Die Folgen sind unabsehbar. Wahrscheinlich dürfte eine starke Temperaturerhöhung folgen, so dass eventuell die Eisberge an den Polen abschmelzen und einen großen Teil von Nordeuropa überfluten. Gaia - als das übersystem - wird zwar auch dann wahrscheinlich wieder einen homöostatischen Zustand erreichen können - doch der muss nicht den bevorzugten Lebensbedingungen der Menschen entsprechen.

Saurer Regen

Ebenfalls wird durch die Verbrennung fossiler Brennstoffe das Auftreten von saurem Regen gefördert. Dadurch werden viele Bäume und Fische vor allem in Skandinavien geschädigt. Doch nur durch Entschwefelungsanlagen in den Industriebetrieben - und die sind in Osteuropa noch weit entfernt - würde der saure Regen nur wenig zurückgehen. Wie im oberen Kapitel über den Schwefel - Kreislauf schon gesagt produzieren auch die Algen Schwefel in Form des Dimethylsulfid. Durch die verstärkte Nährstoffanreicherung in den Meeren, kommt es zur stärkeren Bildung der schwefelproduzierenden Algen und dadurch zu einer Verstärkung der Belastung.

Landwirtschaft und Entwaldung

James Lovelock betrachtet die Landwirtschaft als größte Bedrohung. Da bald zwei Drittel der natürlichen Ökosysteme durch teilweise sogar unproduktive Landwirtschaft ersetzt worden ist, geht die Fähigkeit der Klimaregulierung immer mehr verloren. Besonders schlimm ist die Abholzung der tropischen Regenwälder. Sie sind die Hauptauslöser des

Regens, in dem sie große Menge an Wasser verdunsten (Evapotranspiration) . Die Auswirkungen sind nicht lokal, sondern global, da die Tropen Teil des Kühlsystems der ganzen Erde sind. Eine weitere Abholzung würde eine völlige Austrocknung zur Folge haben - wie die Sahara - nur hundertmal so groß.

Lösung ?

Es ist nicht leicht eine Lösung für Gaias Probleme zu finden. Dazu wissen wir immer noch zu wenig über die Mechanismen und Kreisläufe von Gaia. Doch man muss etwas tun, sonst ist es bald zu spät. Die Rodung sollte natürlich gestoppt werden und weniger fossile Brennstoffe verpulvern, doch Lovelock warnt auch auf zu drastische Eingriffe in das System Erde. Die Menschen müssten lernen Gaia als selbstregulierendes System zu verstehen. Gaia befindet sich in einem instabilen Zustand, betrachtet man z.B. die Kohlendioxid - Konzentration. Eingriffe könnten das System in einen nicht vorherzusagenden chaotischen Zustand befördern.

Vermieden werden sollte vor allem - laut Lovelock - der Ausstoß von Treibhausgasen und Kohlendioxid. Er unterstützt den Umstieg auf alternative und umweltfreundliche Energiequellen, wie Methanol und Wasserstoff.

Allgemein gesagt prangert Lovelock die Ignoranz der Politiker oder der Gesellschaft insgesamt an. Es muss sich noch viel ändern, sonst steuern wir auf eine Katastrophe zu. Wir haben nicht mehr viel Zeit.

9.6 Literaturverzeichnis

- [1] Lovelock, J. (1993). Das Gaia - Prinzip. Insel taschenbuch 1542, Insel Verlag, Frankfurt am Main und Leipzig
- [2] Lovelock, J. (1991). Gaia - Die Erde ist ein Lebewesen. Sachbuch 19/441, Heyne Bücher München

Kapitel 10

Einführung in die Thermodynamik

Christian Fiedler

28. Juni 2000

10.1 Die Klassische Thermodynamik

10.1.1 Einleitung

Die klassische Thermodynamik ist eine Theorie, die sich auf der Basis von vier Hauptsätzen mit den allgemeinen Gesetzmäßigkeiten beschäftigt, denen das Verhalten eines thermodynamischen Systems unterworfen ist. Die Theorie wurde in der Mitte des letzten Jahrhunderts formuliert, also zu einer Zeit, als die mikroskopische Struktur von Materie noch nicht richtig erkannt wurde, bzw. der mikroskopische Lösungsansatz noch nicht bekannt war. Die Erfahrungen wurden auf empirische Art gewonnen, so daß man bei der klassischen auch häufig von der phänomenologischen Thermodynamik spricht.

Die Objekte, die untersucht werden, sind physikalische makroskopische Systeme (alles was mit diesen Objekten wechselwirkt ihre Umgebung). Das System muß von seiner Umgebung deutlich isolierbar sein, das heißt seine inneren Wechselwirkungen müssen viel stärker sein, als seine Wechselwirkungen mit der Umgebung.

Der thermodynamische Zustand eines Systems wird durch einen Satz von thermodynamischen Zustandsgrößen beschrieben, also durch messbare makroskopischen Größen wie z.B. Temperatur, Druck usw...

Da die Zustandfunktionen und ihre partiellen Ableitungen nach den Zustandsvariablen die Zusammenhänge zwischen den makroskopischen Eigenschaften des Systems wiedergeben, reicht es einen Teil der Eigenschaften experimentell zu ermitteln, die übrigen lassen sich dann mit Hilfe der thermodynamischen Beziehungen berechnen.

Da wie oben schon erwähnt die mikroskopische Struktur von Materie noch nicht bekannt war, nimmt die klassische Thermodynamik keine Kenntnis vom atomaren, diskontinuierlichen Aufbau der Materie.

Thermodynamische Systeme können sich im Zustand der Stabilität, der Instabilität, sowie der Metastabilität befinden. Die Erfahrung der phänomenologischen Thermodynamik zeigt aber, dass jedes von der Umgebung isolierte System nach hinreichend langer Zeit in einen Gleichgewichtszustand übergeht, den es spontan dann nicht wieder verlässt.

Dabei durchläuft das System Zustandsänderungen die entweder von selbst ablaufen oder unter dem Einfluß äußerer Eingriffe stattfinden. Kann der Ausgangszustand des Systems nicht ohne bleibende Änderungen in der Umgebung wiederhergestellt werden (im Prinzip also fast alle realisierbaren Prozesse), so heißt der Prozeß irreversibel.

Der reversible Prozeß läuft nur durch Gleichgewichtszustände und kann deshalb ohne Änderungen in der Umgebung zurückzulassen zum Ausgangszustand oder einem früheren Zustand zurückgeführt werden.

Näherungsweise lässt sich ein solcher Prozeß durch eine sehr langsame Prozeßführung realisieren, die Zustandsänderung verläuft dann quasistatisch. Die Zustandsvariablen hängen bei dieser Prozeßführung nicht von der Zeit ab.

Jeder quasistatische Prozess, der von einem Zustand 1 zu einem Zustand 2 führt, lässt sich im Zustandsraum der unabhängigen Zustandsvariablen als Kurve zwischen den Punkten 1 und 2 darstellen.

Man unterscheidet bei der Thermodynamik zwischen abgeschlossenen (kein Materie- und Energieaustausch), geschlossenen Systemen (Energie-, aber kein Stoffaustausch), adiabatisch isolierten Systemen (kein Wärmeaustausch) und offenen Systemen (Energieaustausch durch Arbeit oder Wärme und Stoffaustausch).

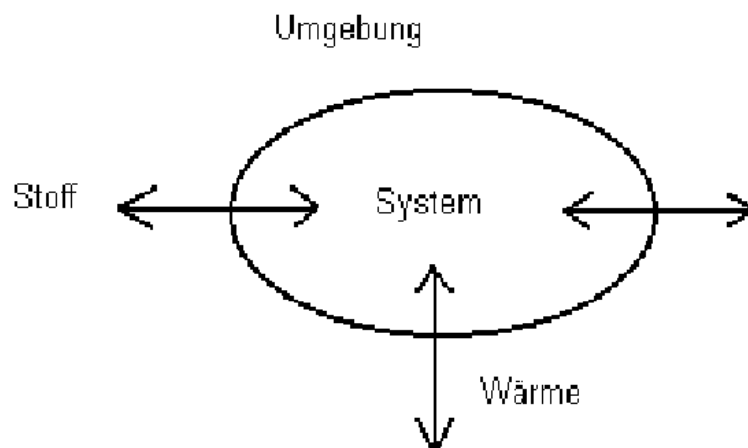


Abbildung 10.1: Wechselwirkungen

Ein System kann ein- oder mehrphasig sein. Unter einer Phase versteht man dabei einen homogenen Bereich, innerhalb dessen keine sprunghafte Änderung irgendeiner physikalischen oder chemischen Eigenschaft auftritt.

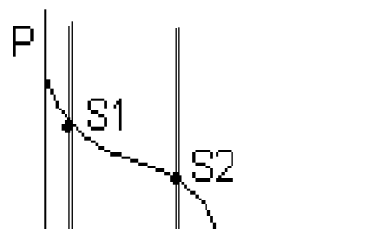


Abbildung 10.2: Phasen und Phasenübergang

Als Beispiel sei das System Wasser und Wasserdampf genannt. Zwischen den Phasen liegen schmale Übergangszonen, in denen sich die Zustandsvariablen sehr schnell mit Zeit

und Ort (meist sprunghaft) ändern (es sind also inhomogene Zonen).

Die Zustandsgrößen lassen sich in zwei Klassen aufteilen: extensive und intensive Größen. Extensive Größen sind zu der Masse der Phase, der sie zugeordnet sind, proportional (z.B. Volumen, Masse, Energie und Entropie).

Extensive Größen haben die wichtige Eigenschaft, dass man für sie Bilanzgleichungen aufstellen kann.

$$\frac{dA}{dt} = \frac{d_a A}{dt} + \frac{d_i A}{dt}$$

Kann eine extensive Größe weder produziert noch vernichtet werden, dann ist $d_i A = 0$. Man spricht dann von einer Erhaltungsgröße, denn für abgeschlossene Systeme ($d_a A = 0$) gilt dann der Erhaltungssatz $\frac{dA}{dt} = 0$ oder $A = \text{const.}$

Beispiele für Erhaltungsgrößen sind die Masse, die Ladung, die Energie, der Impuls und der Drehimpuls. Einige der extensive Größen lassen sich ineinander umwandeln (z.B. kinetische, potentielle oder elektromagnetische Energie). Ebenfalls können extensive Größen von einem System auf ein anderes übertragen werden.

Intensive Größen hingegen sind unabhängig von der Größe der Masse (und damit der Ausdehnung) der Phase der sie zugeordnet sind. Sie haben also in einer Phase, die durch zusammensetzen zweier (in allen Eigenschaften gleicher) Phasen entsteht, denselben Wert wie in den Einzelphasen.

Zu den intensiven Größen gehören z.B. die Temperatur, der Druck, sowie jeder Quotient zweier extensiver Größen, wie etwa die Massendichte $\rho = M/V$.

Zur Beschreibung von Nichtgleichgewichtszuständen kann man keine extensiven Größen nehmen. Das liegt daran, dass die klassische Thermodynamik zur Untersuchung eines Nichtgleichgewichtszustandes das System in infinitesimale Teile zerlegt, um zu erreichen, dass in diesen Teilen wiederum Gleichgewichtszustände herrschen. Dies hat aber zur Folge, dass auch die Größen der Volumenelemente nur infinitesimale kleine Werte haben, mit denen keine Berechnungen durchgeführt werden können.

Man erhält aber aus diesen Werten endliche Größen, wenn man sie durch die ebenfalls infinitesimal kleinen Werte der Volumen oder der Masse teilt. Diese Quotienten sind dann aber wieder intensive Größen, die man dann als Zustandsvariablen eines Volumenelementes, dem Raumpunkt dieses Volumenelementes zuordnen kann.

10.1.2 Der nullte Hauptsatz

Für jedes thermodynamische System existiert eine Zustandsgröße, die Temperatur genannt wird. Ihre Gleichheit ist notwendige Voraussetzung für das thermische Gleichgewicht zweier Systeme oder zweier Teile des gleichen Systems. Sie wird durch eine Zahl charakterisiert, ist also ein Skalar.

Die Temperatur ist eine in der Mechanik und Elektrodynamik unbekannt, speziell für die Thermodynamik eingeführte Zustandsgröße, deren Definition eng mit dem Begriff des thermischen Gleichgewichts verbunden ist.

Als Beispiel gehe man aus von zwei Systemen, A und B, von denen sich jedes in einem Gleichgewichtszustand befindet.

A könnte z.B. ein mit Wasser gefülltes Gefäß sein, B ein erhitztes Stück Metall. Vereint man beide Systeme zu einem Gesamtsystem, in dem die beiden Teilsysteme über diathermische Wände miteinander in Kontakt stehen (man bräuchte das Metall nur in das Wasser fallen zu lassen), so laufen in beiden Teilsystemen messbare Prozesse ab, bis sich ein neuer Gleichgewichtszustand eingestellt hat. Das Gesamtsystem befindet sich nun im thermischen Gleichgewicht. Hätte man bei der Vereinigung von A und B in den Teilsystemen keine Veränderung feststellen können, dann wären sie bereits vor der Vereinigung miteinander im thermischen Gleichgewichtszustand gewesen. Diese Aussage läßt sich auf mehr als zwei Systeme ausweiten.

Damit die Temperatur als eine Größe zur Beschreibung des Zustandes eines Systems verwendet werden kann, benötigt man natürlich eine Form der Temperaturmessung. Dabei gehen wir von zwei Systemen S_1 und S_2 aus, deren Zustände durch jeweils zwei Zustandsvariablen (X, Y) beschrieben werden können. Die Temperatur V_1 in S_1 ist als Zustandsgröße eine Funktion von X_1 und Y_1 , V_2 als Temperatur von S_2 entsprechend eine Funktion von X_2 und Y_2 :

$$V_1 = f_1(X_1, Y_1), V_2 = f_2(X_2, Y_2)$$

Wir betrachten den Fall, daß sich beide Systeme in thermischen Gleichgewicht befinden, also gleiche Temperatur besitzen:

$$V_1 = f_1(X_1, Y_1) = V_2 = f_2(X_2, Y_2)$$

Durch diese Bedingung sind die Zustände der beiden Systeme offensichtlich noch nicht festgelegt. Man kann vielmehr zu einem Zustand des Systems S_1 unter Wahrung des thermischen Gleichgewichts eine ganze Satz von Zuständen (X_1', Y_1') , (X_2', Y_2') , ... finden, die alle zur gleichen Temperatur V_2 gehören. Verbindet man in der X_2, Y_2 Ebene alle diese Punkte, so erhält man eine Isotherme, und für verschiedene Werte der Konstanten erhält man sogar eine ganze Schar von Isothermen.

Um nun eine empirische Temperaturskala festzulegen muß man jeder Isothermen eine Temperatur zuordnen. Da die Isotherme ja überall den selben Wert hat, genügt es nur einem Punkt auf der Isotherme eine Temperatur zuzuordnen. Wir wählen den Schnittpunkt der Isothermen mit der Geraden $Y = Y_0$.

Die Temperatur ist dann eine Funktion die nur noch von X abhängt.

$$v = v(X)$$

Jetzt legt man willkürlich eine lineare Skala fest:

$$v(X) = aX$$

Die Konstante a muß dabei aus Dimensionsgründen eingeführt werden. Ihr genauer Wert wird durch einen Fixpunkt festgelegt der nach int. Konvention bei dem Tripelpunkt des Wassers liegt und ordnet ihm willkürlich die Temperatur von 273,16 K zu.

Vor der Abhandlung der weiteren Hauptsätze sei nochmals darauf hingewiesen, dass

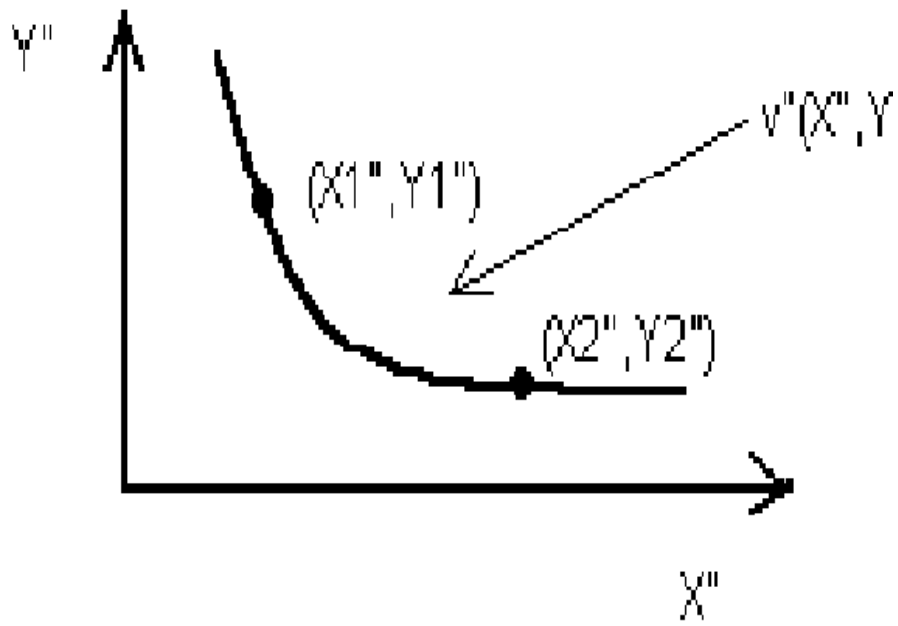


Abbildung 10.3: Isotherme

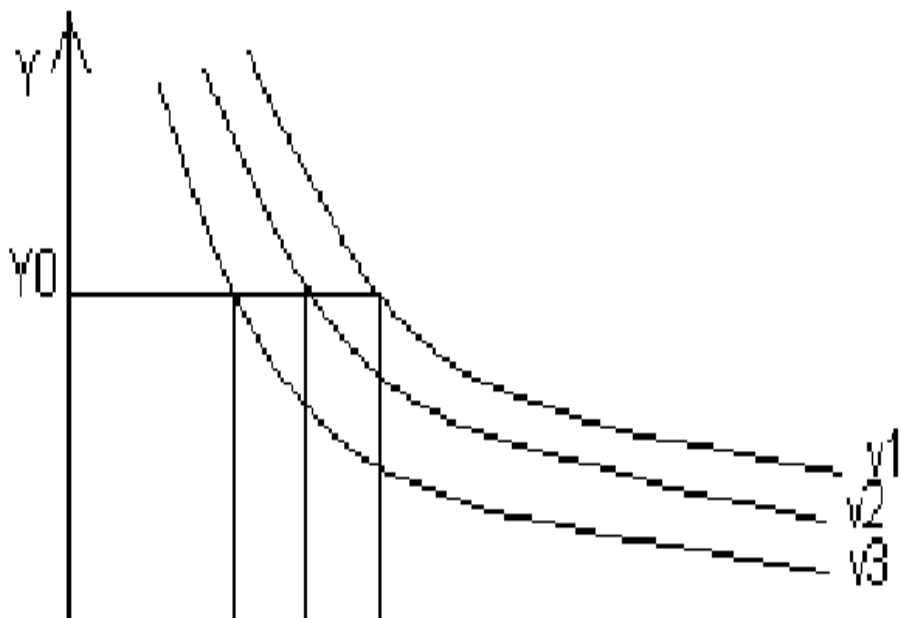


Abbildung 10.4: Festlegung von Temperatur

die Hauptsätze weder abgeleitet noch bewiesen werden. Sie geben lediglich Erfahrungen wieder von denen aber keine Ausnahmen bekannt sind!

10.1.3 Der 1. Hauptsatz

Für jedes System ist die Energie E eine (extensive) Zustandsgröße, welche bei abgeschlossenen Systemen erhalten bleibt. Tauscht ein System Energie mit seiner Umgebung aus, dann gilt für das totale Differential $dE = \hat{d}Q + \hat{d}W$ hierbei ist $\hat{d}Q$ die dem System zugeführte Wärmemenge und $\hat{d}W$ die dem System zugeführte mechanische Arbeit.

Wie man sieht wird zwischen zwei verschiedenen Arten der Energiezufuhr unterschieden. Zum einen gibt es die mechanische Arbeit $\hat{d}W$. Beispiele hierfür sind etwa eine Volumenarbeit oder die Änderung der Teilchenzahl.

Zum anderen kann dem System auch Energie in Form von Wärme $\hat{d}Q$ zugeführt werden und zwar ohne das dabei Arbeit verrichtet werden muß. Zu beachten hierbei ist das $\hat{d}W$, sowie $\hat{d}Q$ keine Ableitungen im mathematischen Sinne sind, da Q und W keine Zustandsgrößen im System sind und somit keine wohldefinierten Werte in Gleichgewichtszuständen besitzen. Die Erklärung dieses Sachverhalts kann man leicht an einem Kreisprozess klar machen. Das System kehrt von einem bestimmten Zustand über verschiedene Zwischenzustände in den Ausgangszustand zurück, wobei die Gesamtänderung für jede Zustandsgröße logischerweise verschwindet. Z.B. für die Energie: $\int dE = 0$

Für die während des Kreisprozesses zugeführte Wärme und Arbeit gilt dagegen i. a.:

$$\int \hat{d}Q = -\int \hat{d}W \neq 0$$

Das heißt aber das die Änderung der Wärme und der mechanischen Arbeit i.a. vom Austauschprozess selbst abhängt und nicht, wie bei der Energie, lediglich vom Anfangs- und Endzustand des Systems. Anders ausgedrückt : Im Gegensatz zu dE sind $\hat{d}Q$ und $\hat{d}W$ keine totalen Differentiale .

Die Erkenntnis des ersten Hauptsatzes ist im wesentlichen J.p. Joule zu verdanken, der in präzisen Experimenten nachgewiesen hat, dass Wärme nur eine spezielle Form der Energie ist.

Aus dem ersten Hauptsatz ergibt sich auch noch der Widerspruch zum perpetuum mobile 1. Art, da ein perpetuum mobile eine Maschine ist, die immer wieder Kreisprozesse durchläuft und dabei mehr Energie abgeben soll, als sie aufnimmt. Da aber nach jedem Kreisdurchlauf die Maschine (das System) wieder im Anfangszustand ist, muß die Veränderung $dE = 0$ sein. Damit muß dann aber auch gelten $\hat{d}W + \hat{d}Q = 0$, was wiederum zu der Aussage führt, dass keine periodische Maschine mehr Energie abgeben kann, als die ihr von außen zugeführte Wärmemenge.

10.1.4 Der 2. Hauptsatz

1. Teil: Es gibt eine extensive Zustandsgröße, die Entropie S , und eine intensive Größe, die absolute Temperatur T , mit der folgenden Eigenschaft: Für ein nicht isoliertes System, welches in einem quasistatischen Prozess die Wärmemenge $\hat{d}Q$ absorbiert gilt:

$$dS = \frac{\hat{d}Q}{T} \text{ (reversibler Proze\ss)}$$

$$dS > \frac{\hat{d}Q}{T} \text{ (irreversibler Proze\ss)}$$

2. Teil: Die Entropie eines abgeschlossenen Systems kann mit der Zeit nur gr\o\sser werden und ist im Gleichgewicht maximal: $\Delta S \geq 0$

(Die Zustandsgr\o\ss e Entropie wurde 1850 von R. Clausius eingef\uhrt.)

In abgeschlossenen Systemen kann die Entropie nur zunehmen oder h\o\chstens gleichbleiben. Solange im abgeschlossenen System noch Prozesse von allein ablaufen, wird Entropie produziert, d.h. die Entropie des Systems w\acchet an. Erst wenn der Gleichgewichtszustand erreicht ist, h\o\rt die Entropieproduktion auf; die Entropie selbst hat dann einen Maximalwert erreicht.

Im Zusammenhang mit dem ersten Hauptsatz haben wir gesehen, da\ss Energie nicht erzeugt werden kann, wohl aber die verschiedenen Energieformen ohne jegliche Form der Einschr\ankung ineinander umformbar sind. So w\are es beispielsweise m\o\glich die riesigen Vorr\ate an innerer Energie, die in den Weltmeeren stecken, zur Arbeitsleistung nutzbar zu machen, indem man eine Maschine konstruiert, die ohne weitere Energiequellen einfach durch Abk\uhlen des Meeres Arbeit leistet und dadurch etwa ein Schiff antreibt. Alle Versuche, eine solche Maschine zu bauen, sind aber gescheitert. Man fa\ss t diese Erfahrung in dem Satz der Unm\o\glichkeit eines perpetuum mobile 2. Art zusammen:

Es ist unm\o\glich, ein perpetuum mobile 2. Art, d.h. eine periodisch funktionierende Maschine zu konstruieren, die weiter nichts bewirkt als das Leisten einer Arbeit und Abk\uhlung eines W\armereservoirs.

10.1.5 Die Gibbsche Fundamentalform

Unter Verwendung des 1. und 2. Hauptsatzes ergibt sich die Gibbsche Fundamentalform: Bei einer quasistatischen Zustands\anderung ist:

$$dE = \hat{d}Q + \hat{d}W \leq T dS - \sum X_i dx_i$$

Das Gleichheitszeichen gilt f\ur reversible Prozesse.

L\o\sen wir diese Gleichung (reversibler Prozesse) nach dS auf

$$dS = \frac{dE}{T} + \sum \frac{X_i}{T} dx_i$$

und bilden andererseits das totale Differential der Entropie,

$$dS = \frac{\delta S(E, x)}{\delta E} dE + \sum \left(\frac{\delta S(E, x)}{\delta x_i} \right) dx_i$$

dann folgt durch Vergleich der letzten beiden Gleichungen

$$\frac{1}{T} = \frac{\delta S(E, x)}{\delta E}, \quad X_i = T \frac{\delta S(E, x)}{\delta x_i}$$

Diese Ausdr\ucke entsprechen gerade den Definitionen der Temperatur und der generali-

sierten Kräfte in der statistischen Physik. Hieran sieht man das die Begriffe „Entropie“ , „Temperatur“ und „generalisierten Kräfte“ in der statistischen und phenomenologischen Thermodynamik äquivalent sind und das die statistische und phenomenologische Thermodynamik gerade im zweiten Hauptsatz sehr eng miteinander verbunden sind.

10.1.6 Der 3. Hauptsatz

Bei absolutem Temperatur-Nullpunkt $T=0$ nähert sich die Entropie eines Gleichgewichtssystems dem Entropie-Nullpunkt:

$$S \rightarrow 0 \quad (T \rightarrow 0)$$

Bei der Untersuchung von Reaktionen zwischen reinen, kristallisierten Festkörpern stellt man fest, dass die diesen Reaktionen zuzuordnende Entropieänderung, die Reaktionsänderung δS bei Annäherung an den absoluten Nullpunkt gegen null strebt. Dies hat auch zur Bedeutung, dass es unmöglich ist, den absoluten (Temperatur-, sowie Entropie-) Nullpunkt zu erreichen.

Der Reiz der Thermodynamik liegt in ihrer allgemeinen Gültigkeit. Die Beschreibung von komplexen Systemen kann durch nur wenige Hauptsätze erreicht werden. Allerdings hat dies auch seine Schattenseiten. So können z.B. keine absolute Berechnungen von Zustandsgrößen erfolgen.

10.2 Der Stirlingmotor als Beispiel für eine Wärmekraftmaschine

10.2.1 Einleitung

Der bereits im Jahre 1816 von Reverend Robert Stirling in Schottland entwickelte Heissgasmotor stellt eine Wärmekraftmaschine mit "Außenverbrennung" dar. Zunächst mit dem Ziel entwickelt, die Zahl der schweren Unfälle beim Betreiben der damals noch sehr unvollkommenen Dampfmaschinen zu senken, spielen Stirling-Maschinen heute eine wichtige Rolle bei der umweltfreundlichen Bereitstellung von Arbeit und elektrischer Energie.

Das Wirkprinzip ist einfach: der Temperaturunterschied zwischen zwei Wärmereservoirien wird zur Erzeugung mechanischer Arbeit ausgenutzt. Das dazu verwendete Arbeitsmedium befindet sich in einem geschlossenen Kreislauf und wird somit nicht verbraucht; die Art der Wärmequelle ist beliebig, so dass sich die Maschine mit umweltfreundlichen Energieträgern und insbesondere auch mit Sonnenenergie betreiben lässt. Moderne Stirling-Maschinen arbeiten teilweise mit Wasserstoff oder Helium als Arbeitsmedium und mit höheren Betriebsdrücken. Maschinen in der Größenordnung von 10 KW bis 200 KW, ja sogar bis 5000 Ps, sind kommerziell verfügbar. Sie werden u.a. als Autoantrieb, Blockheizkraftwerk und bei der Umwandlung von Sonnenenergie in mechanische Bewegungen zum Antrieb eines Stromgenerators eingesetzt.

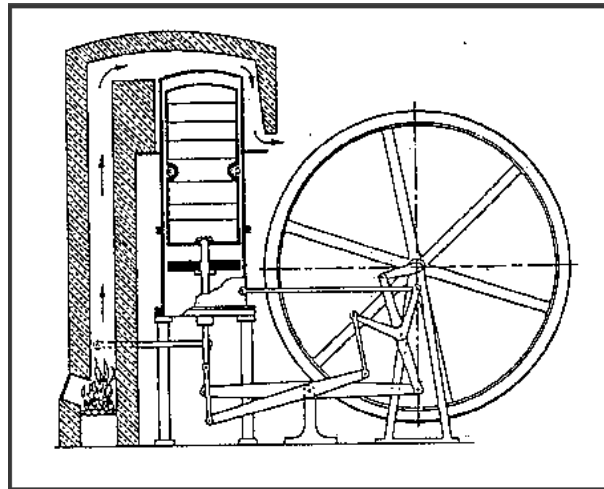


Abbildung 10.5: „Erster“ Stirlingmotor

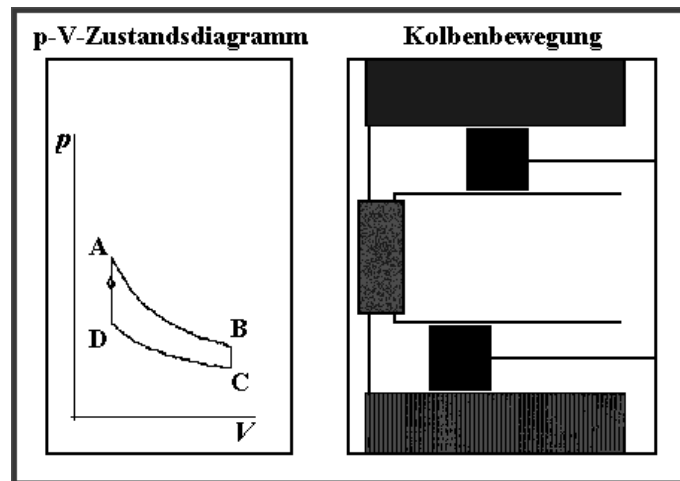


Abbildung 10.6: Schema eines Stirlingmotors

10.2.2 Theoretische Grundlagen

Die Menge aller möglichen Zustände lässt sich in einem p - V -Zustandsdiagramm darstellen. Das entsprechende Diagramm für einen idealen Stirlingmotor ist links zu sehen.

Auf der rechten Seite sieht man die schematische Darstellung eines Stirling Motors. Oben befindet sich das heiße Wärmereservoir, das kalte unten. Die Temperatur des in der linken Fenstermitte befindlichen Regenerators ändert sich in Abhängigkeit vom Betriebszustand der Maschine.

Arbeitszyklus einer 2-Kolben-Stirling-Maschine

Das Arbeitsmedium des Motors bewegt sich zwischen heißen und kaltem Wärmereservoir hin und her und durchströmt dabei den sogenannten Regenerator. Dem heißen Reservoir wird von außen Wärme zugeführt. Die beiden Kolben sind über eine Mechanik (z.B. Schwungrad) miteinander verbunden. Im Arbeitstakt der Maschine (Übergang vom Zustand A im Zustandsdiagramm zum Zustand B) dehnt sich das Arbeitsgas aus.

Dies geschieht isotherm, denn aus dem heißen Reservoir kann Wärme (Q_{AB}) aufgenommen werden.

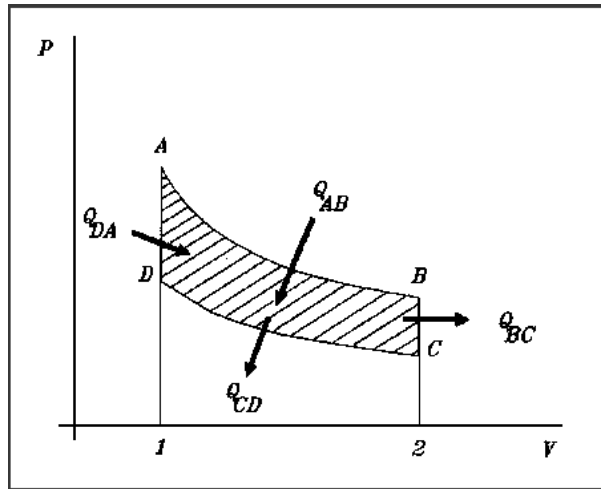


Abbildung 10.7: Zustandsdiagramm

Der obere, heiße Kolben bewegt sich nach rechts, während der untere, kalte Kolben stillsteht. Ab Punkt B kehren sich die Bewegungsrichtungen der Kolben (über die Kolbenmechanik vermittelt) um. Unter Beibehaltung des Gesamtvolumens wird heißes Gas in den kalten unteren Zylinder verfrachtet. Da die mittransportierte Wärme (Q_{BC}) an den Regenerator abgegeben wird, kühlt sich das Arbeitsmedium ab und der Druck fällt: Punkt C wird erreicht. Von C nach D wird das Arbeitsgas isotherm komprimiert. Der kalte Kolben bewegt sich nach links, während der heiße Kolben festgehalten wird.

Die Temperatur steigt im Unterschied zum Druck durch die Verdichtung nicht, da Wärme (Q_{CD}) an das kalte Reservoir abgegeben werden kann - ein notwendiger Verlust. Der vierte Schritt von D nach A beschließt den Arbeitszyklus. Kaltes Gas strömt in den heißen, oberen Zylinder, nimmt auf den Weg dorthin Wärme (Q_{DA}) aus dem Regenerator auf: Druck und Temperatur steigen, während das Gesamtvolumen konstant bleibt. Die vom Kurvenverlauf eingeschlossene Fläche entspricht der geleisteten Arbeit.

Berechnung der Arbeit, die eine ideale Stirling-Maschine leistet

Bei der Expansion des Arbeitsgases von A nach B wird die (negative) Arbeit W_{AB} geliefert (siehe Abbildung Fläche 1AB2).

Um das Gas zu komprimieren, muß die (positive) Arbeit W_{CD} aufgebracht werden (Fläche 2CD1). Die Arbeit W_{id} , die eine ideale Stirling Maschine leisten kann, ergibt sich zu:

$$W_{id} = W_{CD} + W_{AB}$$

Einsetzen der Zustandsgleichung des idealen Gases und anschließende Integration liefern:

$$W_{id} = -mR_s T_C \ln \frac{V_D}{V_C} - mR_s T_A \ln \frac{V_B}{V_A}$$

Damit ergibt sich:

$$W_{id} = -m R_s (T_A - T_C) \ln (V_B/V_A)$$

mit $m :=$ Masse des Arbeitsgases

R_s := spez. Gaskonstante [$R_s = R/M$]

R := molare Gaskonstante

M := molare Masse

T_A := Temperatur am Punkt A

T_C := Temperatur am Punkt C

Außerdem gilt für Arbeit sowie zu- und abgeführte Wärme:

$$W_{id} = Q_{zu} - |Q_{ab}|$$

Hieraus kann man jetzt den Wirkungsgrad ablesen:

$$\eta = |W_{id}| / Q_{zu} = (T_A - T_C) / T_A = 1 - T_C / T_A$$

Man sieht das der Wirkungsgrad eines idealen Stirling Motors lediglich von der Temperaturdifferenz zwischen warmem und kaltem Reservoir abhängt.

10.3 Literatur

- [1] Wachter, Armin: Repetitorium Theoretische Physik. Springer-Verlag, 1998.
- [2] Enzyklopädie Naturwissenschaft und Technik. Verlag Moderne Industrie, 1981. Sonderausgabe für den Zweiburgen Verlag.
- [3] Kluge, Gerhard: Grundlagen der Thermodynamik. Spektrum Akademischer Verlag, 1994.

Kapitel 11

Introduction to Self-Organized Criticality

Harold Fellermann

July 5, 2000

Abstract

The theory of self-organized criticality (SOC) was first introduced by Per Bak, Chao Tang and Kurt Wiesenfeld in 1987. Their publication proclaimed to explain a well-known ubiquitous phenomenon called $1/f$ -noise (one-over-f-noise). Although this kind of signal has been observed for many years in a wide range of systems, up to now a theory of its origin had been lacking. The SOC-hypothesis proposed by the authors can describe various different systems, e.g. traffic jams, earthquakes or the fluctuation of stockvalues, without the need for specific physical or economical assumptions. SOC is therefore a concept in the context of complexity research. According to SOC, complex behavior emerges from the systems critical state. In equilibrium systems, this state is difficult to achieve, however, if the system is far from any equilibrated state, criticality might be self-organized.

11.1 Properties of Complex Systems

Before discussing SOC, we need to specify the boundary of this theoretical concept. What is the family of systems SOC can be applied to?

Consider an ecosystem, a stockmarket, a social society or a neuronal population. For decades an irreducible and unpredictable nature was thought to be the only common characteristic of these and comparable systems, troubling scientists of all disciplines. In the mid 80s, there was a new attempt to understand the underlying principles of those systems, leading to the development of complexity theory¹. In the usual course of a new theory emerging no bounding definition for classifying the investigated phenomena could be agreed on. Since no consensus interpretation was found for the term “complexity”, the most pragmatic and quite intuitive “*I know it when I see it*”-method was used instead for deciding on the focus of complexity theory.

¹On the historical background of complexity theory see [9]

In 1987, when Bak, Tang and Wiesenfeld published their theory, a wide range of exemplary systems was already considered to be complex. This allowed to gather several characterizing properties:

Hierarchical Structure: Most complex systems are composed of numerous similar interacting components, e.g. large numbers of individual companies and consumers in world trade, individual cars in traffic systems or computers forming the world wide web. These components can be conceived as “agents”. Observing a system, an agent is usually seen as an unstructured whole, but of course an agent can also be regarded as a system itself, composed of a number of components: e.g. billions of neurons form the human brain. Each neuron itself has a sophisticated structure of cell-organelles. Organelles are constituted by thousands of different biomolecules. More levels of decomposition could be listed. Although the hierarchy may be multi-leveled, defining entities suggest themselves on each level. That the behavior of each level is an emergent property resulting from the structure of the underlying level can be seen as a paradigm-like rule-of-thumb in complexity research. This implies, that one should not investigate all details but better take up a panoramic point of view.

Interacting agents: Typically those agents can interact or communicate with each other. Complexity theory tends to focus on systems with mainly local interaction, thus each agent is restricted to communicate with only few agents in its neighborhood. For example, a commuter involved in a rush hour’s traffic jam plans his way through the city according to the density of other cars in his vicinity: Each crossing offers a choice for his further route, where he will usually choose the one less traveled by. But he can never perceive the global situation of the traffic jam. There are no restrictions on the agents’ actions — they need not be reasonable or deterministic as long as they are well defined for every local situation the agent is confronted with.

Interaction with the environment: All complex systems are considered to be open, able to interact with its environment (And once again, there is no limit to the kind of external inputs). Why is the openness so essential? It drives the system out of its equilibrated state. In this state the system would be stationary, and when disturbed, return quickly to balance. But this is not observed in the systems described. In fact, they are far away from any equilibrated state! There is still another reason from a thermodynamic point of view. According to phenomenological thermodynamics, the entropy in all closed systems increases — which means that it gets unordered over time. In complex systems indeed, one always finds a high level of order.

Complicate behavior: The conjunction of a hierarchical interwoven structure and the absence of an equilibrium is responsible for the typically complicate behavior of those systems. Agent interaction can amount to strong inherent dynamics, leading to unpredictable responses under external influence. Therefore the outcome of two identical inputs may be completely different, even if the internal state of the system is nearly the same.

Although this list of properties might suggest that complex systems are always complicate, there are lots of exemplary systems of a very clear and simple structure. The canonical introductory example of SOC is an ordinary pile of sand. Though applying SOC to such

a miniscule example might seem like an overkill, the sandpile fits all the described properties: It has a hierarchical structure consisting of thousands of sand grains (the agents) interacting by pushing and holding each other. If the surface of the pile is disturbed by placing a single grain on its peak, the response of the sandpile will vary: in many cases nothing will happen, but often some grains will be dragged down forming avalanches of all sizes. The behavior of the system in reaction to external input can be conceived as unpredictable and of unspecified extent.

11.2 The Theory of Self-Organized Criticality

11.2.1 catastrophes and $1/f$ -Noise

In science and especially in dynamics nature has always been regarded as gradual. From its Newtonian beginnings up to the vast field of nonlinear dynamics the underlying mathematics are continuous and furthermore differentiable. Small irregularities in experimental data have been called disturbances. Where nature was rough, scientists smoothed it to fit their idealistic assumptions; where it was jagged they tried to determine trends. From this point of view greater irregularities like catastrophes cannot be handled. Due to their spontaneous and irregular appearance they were excluded from science over centuries². But obviously catastrophes are widespread phenomena: earthquakes, epidemics, stock-exchange crashes, wars and others can be found in most systems of all scientific domains. In the 20th century the study of catastrophes was taken up in several disciplines and exposed that they are only the striking top of a fundamental principle called $1/f$ -noise (“one over f ”-noise)³. It was observed that *the frequency of a specific event is inversely proportional to the strength of the event*. For example in geophysics there is the so-called Gutenberg-Richter-Law formulated in 1956. It proclaims, that if in a given interval there occur 1000 earthquakes of magnitude 4 on the Richter scale, there are approximately 100 earthquakes of magnitude 5, 10 of magnitude 6 and so on. As the Richter scale is logarithmical (If the earthquake is one unit stronger, it releases 10 times more energy) the relation between the earthquake’s strength and frequency is inversely proportional⁴. By now the number and kind of systems where $1/f$ -noise was observed is enormous. The examples range from the emitted light of a quasar or fluctuations in highway traffic or the level of the river Nile, however, in each the origin of $1/f$ -noise was unknown. It was this universal phenomenon, Bak, Tang and Wiesenfeld aimed at explaining in their theory of self-organized criticality.

Consider the noisy timeseries of a complex system’s output, the trembling fluctuations of stock values for instance. The chart of the output-signal is shown in figure 11.1. It contains peaks of all sizes. One may take this signal as a superposition of signals each one having its own fixed frequency. Using a Fourier-transformation one can extract and display the intensity of each frequency at a given time: the power spectrum of the signal. It turns out that the power spectrum of $1/f$ -noise is given by the simple equation

$$P(x) \sim x^{-\alpha} \tag{11.1}$$

²R. Thoms catastrophe theory was eventually the first attempt to regard spontaneous changes in a systematic manner.

³ $1/f$ -noise is also called pink noise or flicker noise.

⁴Do not confuse probabilistic frequency with periodicity! The fact that there was no earthquake for a long time does not implicate that an earthquake is due!

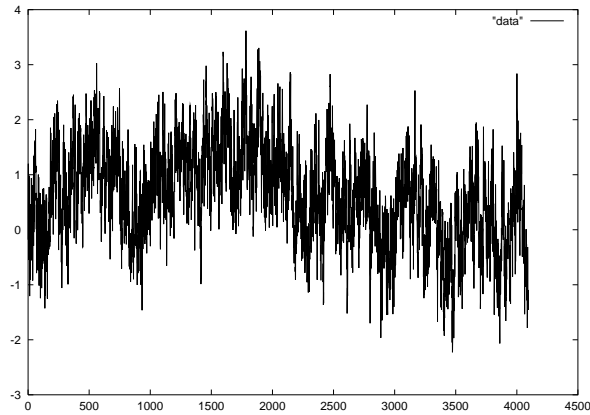


Abbildung 11.1: A typical signal showing $1/f$ -noise — a superposition of weighted frequencies.

a power law where $P(x)$ indicates the frequency of an event with magnitude x and $\alpha \approx 1$. Since

$$P(x) \sim x^{-\alpha} \Leftrightarrow \log P(x) \sim \log x^{-\alpha} = -\alpha \log x \quad (11.2)$$

this power spectrum reveals approximately a straight line in the frequency/amplitude-diagram if plotted logarithmically (See figure 11.2).

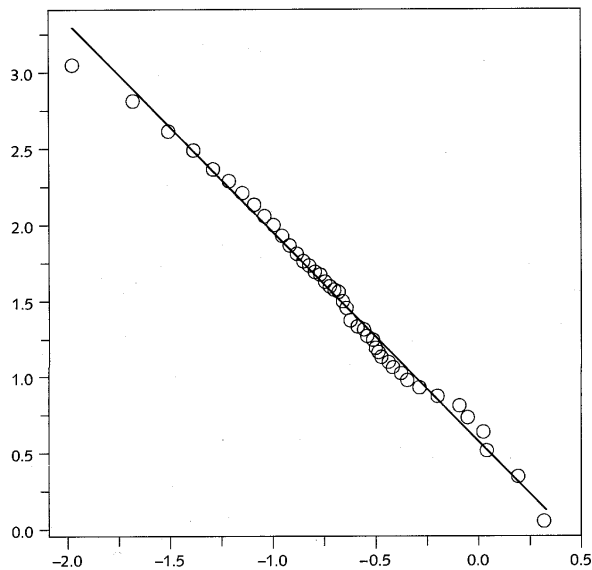


Abbildung 11.2: The power spectrum of $1/f$ -noise reveals approximately a straight line.

One might wonder, why a straight line in the power spectrum of $1/f$ -noise would be exceptional. At first it shows, that the signal is scale-independent: The relation between two changes in the timeseries, where one is k times stronger than the other is given by a function independent of x :

$$\frac{P(kx)}{P(x)} = k^{-\alpha} \quad (11.3)$$

In contrast to white noise and a strictly periodical signal, $1/f$ -noise shows a fractal, self-similar composition.

Another point is, that power spectra can be related to time correlation. If α is equal to 0 the corresponding signal is white noise that shows no time correlation: One cannot predict the next value of the timeseries, since fluctuations of all amplitudes have the same probability. Compared to $1/f$ -noise, white noise (shown in figure 11.3) is definitely less structured. If, on the other hand, α is close to 1 the time correlation reaches its maximum⁵ and the probability of great bumps (catastrophes) is by far smaller than the one of small steps. This is the reason why $1/f$ -noise is of such particular interest.

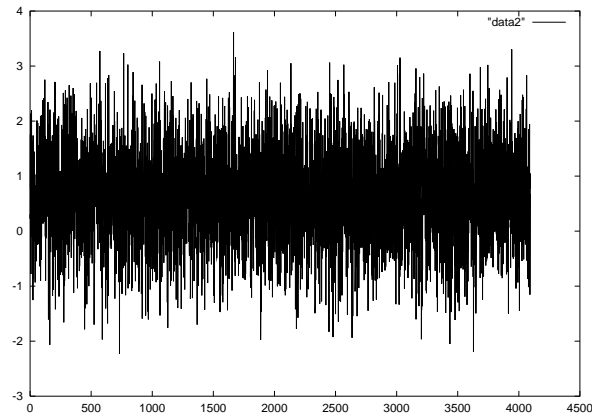


Abbildung 11.3: In white noise signals of all frequencies are superposed with equal amount.

11.2.2 The Critical State

$1/f$ -noise can also be found in thermodynamics. Consider a closed tank that is halfway filled with benzol for instance. Under standard conditions the benzol is liquid apart from a small number of molecules which are solved in the gaseous phase. Between the liquid and the gas is a sharp phase transition – the surface of the liquid. By increasing the temperature or pressure of the system the ratio of liquid to gaseous benzol can be changed. If the pressure of the system is above the so-called critical value (which is for benzol approximately 48 bar), increasing the temperature can lead to a breakdown of the sharp phase transition. By heating the tank the liquid phase gets more and more diffuse while the air over the benzol becomes cloudy: above the critical temperature (here 288°C) the surface vanishes completely, as the distinction between the phases disappears. The benzol is then in a plasmatic state where fluid and gaseous structures pervade each other. As Waldrop says, the extension of these interwoven structures accords to the distribution of $1/f$ -noise. This plasma is called the critical state of the system. The reason for this name is its instability:

“When the temperature of the system is precisely equal to the transition temperature, something extraordinary happens. For all other temperatures, one can disturb the system locally and the effect of the perturbation will influence only the local neighbourhood. However, at the transition temperature, the

⁵For a closer view to the relation between power spectra and temporal correlations see [6]

local distortion will propagate throughout the entire system [...] The system becomes *critical* in the sense that all members of the system influence each other.⁶

11.2.3 Self-Organized Criticality

In thermodynamics the system must be tune it into the critical state — one has to adjust its temperature and pressure. If you want to explain the ubiquitous nature of $1/f$ -noise by critical behavior, you must point out the tuning-processes, which drive the system in this special state. This is exactly what Bak, Tang and Wiesenfeld did: They proposed that *the critical state of complex systems is self-organized*. How can these two concepts be reconciled?

Think of earthquakes once again. They are the sudden response of the earth's crust to inner stress accumulated over years by slow tectonic movements. A similar process can be found in the sandpile-model: We drop one grain after the other, but between two drops, the slope can quickly change its configuration. Ever in self-organized critical systems, there is a slow external force driving the system towards the critical state. Whenever the state gets overcritical the force is released by a comparatively quick relaxation process of the system's components: *the critical state is an attractor of the complex system's dynamics*⁷. H. J. Jensen proposes to call these systems *slowly driven, interaction-dominated threshold systems* (SDIT systems):

“The interesting behavior arises because many degrees of freedom are *interacting*; and the dynamics of the system must be *dominated* by the mutual interactions between these degrees of freedom, rather than by the intrinsic dynamics of the individual degrees of freedom.⁸”

Remember that complex systems are far from any equilibrated state. The phasespace of those systems is typically covered with many marginally and metastable points. One also refers to them as punctuated equilibria. In the relaxation process the system slides over these thresholds until it rests in one metastable state. But due to the rough phasespace, the magnitude of this slide is unpredictable — it accords to the characteristic distribution of $1/f$ -noise.

11.2.4 Spatial fractals and SOC

Scale-independence of a signal (as described in section 11.2.1) is a famous property of fractal geometry introduced in the early 1980s by B. Mandelbrot⁹. As well as $1/f$ -noise, fractal patterns were found in a huge set of natural and man-made systems. And similarly a systematical explanation concerning their origin has not been found. Could SOC eventually give a key to the understanding of fractal pattern formation?

A typical example for fractal structures in nature are river networks. Zooming into a map, reveals the same structure of small rivers flowing into greater rivers on all scales. It

⁶[6] page 2

⁷In [3] P. Bak and K. Chen argue that normally the critical state is even insensitive against slight changes of the system's rules.

⁸[6] page 126

⁹It is interesting that one of Mandelbrot's first studies of fractals (in 1963) was on economical time-series showing $1/f$ -noise!

is absurd to define a mean for a typical rivers length — in fact a mean might not even exist! However there is a simple correlation between the length and frequency of rivers, known as Hortons Law. This law defines the order of river segments as the number of links to other river segments that has to be passed before the river reaches the ocean. Then it states *that the number of segments of each order increases as a power law in the order*. Another empirical law states that the length L of a river relates to the area A that is drained by it by a power law:

$$L = 1.4A^{0.6} \quad (11.4)$$

Maybe these correlations indicate a link between SOC and fractality¹⁰. Unfortunately, up to now there has only been little research in this field (One approach will be discussed in section 11.3.2). Nevertheless the following argument could explicate the connection:

As laid out in section 11.2.3 the critical state results from two processes: energy supply from the outside connected with energy dissipation through the system. The same processes are found in river formation: energy is supplied by an uplifting process and dissipated through erosion. What happens has been shown clearly by H.J. Jensen:

“When the system is either driven at random or started out from a random initial state, regions that are able to transmit a signal will form some sort of a network. This network will be modified, or correlated, by the action of the internal dynamics induced by the external drive. The dynamics stop every time the internal dynamics have relaxed the system, so that all local regions are below threshold [...] The result is a complicated, delicately interwoven web of regions that are coupled dynamically.”¹¹

11.3 Computer Models of SOC

In the meanwhile, the number of systems SOC was applied to is remarkable. Also, several computer models were implemented. Two of them will be discussed in this section. But first, a few introductory words need to be said concerning these models. Since they only consist of a small set of parameters and rules, one should regard them as Gedankenexperimente. Therefore they do not describe any specific system, and one must not expect specific quantities taken from empirical measurements. Both are qualitative models which aim at reducing the dynamics of natural systems to the theoretically required minimum, and at showing, that the observed behavior can be explained within this theory.

11.3.1 The Canonical Sandpile Model

The first model of SOC was made by Bak, Tang and Wiesenfeld to back up their statements. It simulates the dynamics of the sandpile which is represented by a two-dimensional cellular automaton. In the beginning each cell $c(x, y)$ of the lattice is initialized with a random number ranging from 0 to c_{crit} (which usually is 4). This value can be convinced as an approximation to the local gradient of the slope. In the iteration process a grain of

¹⁰Obviously, this power law is not the one from equation 11.1, but there a temporal compared to a spatial correlation was discussed.

¹¹[6], page 5

sand will be dropped at a random position (x, y) on the pile's surface, which is represented by two steps: The first step describes conservative perturbation¹²:

$$\begin{aligned} c(x, y) &\rightarrow c(x, y) + 2 \\ c(x - 1, y) &\rightarrow c(x - 1, y) - 1 \\ c(x, y - 1) &\rightarrow c(x, y - 1) - 1 \end{aligned} \tag{11.5}$$

After perturbation the automaton enters the relaxation process. While there is at least one cell $c(x, y)$ in the lattice, whose value exceeds the critical value c_{crit} the following rules are applied:

$$\begin{aligned} c(x, y) &\rightarrow c(x, y) - 4 \\ c(x \pm 1, y) &\rightarrow c(x \pm 1, y) + 1 \\ c(x, y \pm 1) &\rightarrow c(x, y \pm 1) + 1 \end{aligned} \tag{11.6}$$

(That corresponds to Neumann-neighborship.) Note that cells in the neighborhood of $c(x, y)$ may get out of balance themselves! This general set of rules is completed by boundary conditions. At the boundaries of the automaton's lattice (where either x or y is equal to 0 or the lattice's size L), the perturbation rule is unaffected except that accesses to cells behind the boundaries are neglected. The relaxation process is modified depending on whether the boundary is open or closed¹³. For a closed boundary the relaxation rule 11.6 is altered to $c(x, y) = 0$. If the boundary is closed, the rules remain the same except that an access behind the boundaries is neglected once again. Alternatively the boundaries may be periodic — then

$$\begin{array}{ll} c(-1, y) & \text{is mapped on } c(L, y), \\ c(x, -1) & \text{on } c(x, L), \\ c(L + 1, y) & \text{on } c(0, y), \text{ and} \\ c(x, L + 1) & \text{on } c(x, 0). \end{array}$$

When the values of all cells have dropped below c_{crit} , the cycle of iterations restarts. The sandpile-automaton accords to the assumptions of SOC in the following way:

- The system consists of many similar components, the lattice's cells, that interact locally in the way of rule 11.6.
- The system is torn away from its equilibrium, that is when all cells are below the critical value, by an external stream of grains (rule 11.5).
- Whenever a cell gets overcritical all components have enough time to relax again.

According to SOC, these assumptions are sufficient to produce the desired phenomenon. So let us have a look at the outcome of the simulation! In the beginning, most cells will be undercritical, so that no relaxation takes place. The average value will slowly increase until several cells exceed c_{crit} . This starts a short period where scattered avalanches reform the

¹²One might even simplify the perturbation to a nonconservative one, where the only rule is $c(x, y) \rightarrow c(x, y) + 2$ but then the corresponding image of the slopes gradient breaks down.

¹³For each boundary one can define individual conditions.

local neighborhood of overcritical cells. As the average gradient increases, the relaxation processes expand and drive the model into a configuration, where perturbations pass through the whole lattice (See figure 11.4). Up to now the model behaves in the claimed way. What can be said about $1/f$ -noise? As Jensen shows in all the detail¹⁴ the model exhibits this behavior in three different ways: The statistical size of avalanches, the number of cells involved in the relaxation process that follows a single perturbation, and also the total number of simultaneous updates while relaxing show $1/f$ -noise (The number of relaxing cells corresponds to the released energy, while the number of updates can be convinced to be the avalanches lifetime).

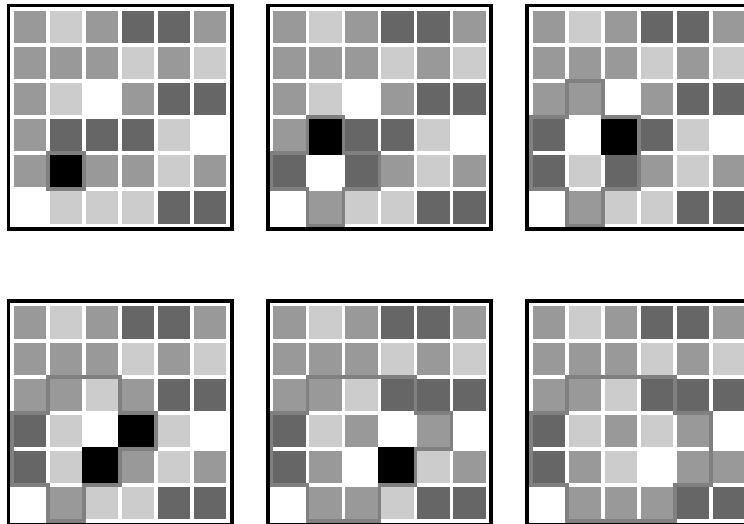


Abbildung 11.4: An avalanche passes the slope of the sandpile: the gradient of the slope rises from white to gray. Black indicates the value $c_{crit} = 4$. The boundaries are closed. After 6 relaxations the slope is in balance again.

There is another point that can be explained by the model: Rules 11.5 and 11.6 change the value of each cell by a constant amount that does not depend on c_{crit} . If one increases c_{crit} , the starting period, where perturbation drives the system into its critical state extends. But when this state is reached, there will be no difference in the systems behavior apart that the gradient of the slope is steeper.

Additionally, it should be mentioned, that several experiments were conducted to measure $1/f$ -fluctuations in physical sandpiles. The experimental conditions varied in a wide range: two- and three-dimensional piles consisting of different grains of sand (and also rice) were observed. It turned out that many systems got into a state in which they produced $1/f$ -noise.

11.3.2 A Model of River Networks

The model discussed in the previous section was a cellular automaton. The outcome we observed were timeseries. Spatial properties of the underlying lattice were ignored. However, these properties are crucial, when we analyze the formation of fractal patterns. Unfortunately there are only a few studies on the pattern-generating task of SOC. One of these studies is the work of A. Rinaldo, R. Rigon, I. Rodriguez-Iturbe and R. Brass who

¹⁴see [6] page 36ff

modelled the branching structure of river networks formed by the effects of erosion and uplifting plate tectonics.

The model of Rinaldo et al. is based on a lattice, where each cell represents the segment of a river in a hilly landscape. Again the value of the cell indicates the gradient of the slope. Starting from a random initial state (a river network with random interconnections between segments), the local stress on each cell is calculated by the formula

$$stress = \sqrt{As^3} \quad (11.7)$$

where s indicates the gradient of the slope, while A represents the area drained by the actual segment¹⁵, or inversely spoken the area that supplies the segment with water (as rainfall was assumed to be homogeneous, this is approximately proportional to the amount of carried water). Whenever the local stress surpasses a given critical value, erosion takes place, shifting one unit of material down the slope, until the whole landscape is subcritical again. In addition to this relaxation process, the system is disturbed by uplifting processes which increase the value of each cell. In contrast to the model of the sandpile, here the stress is computed in a global manner, as the area A may stretch through the whole lattice.

The simulation shows that the system tends to go into the stationary critical state where the two dynamics are balanced (see figure 11.5). Statistical analysis proofs that the outcome fits both Hortons Law and the power law of equation 11.4. Furthermore the model exhibits fractal characteristics, e.g. a fractional dimension.

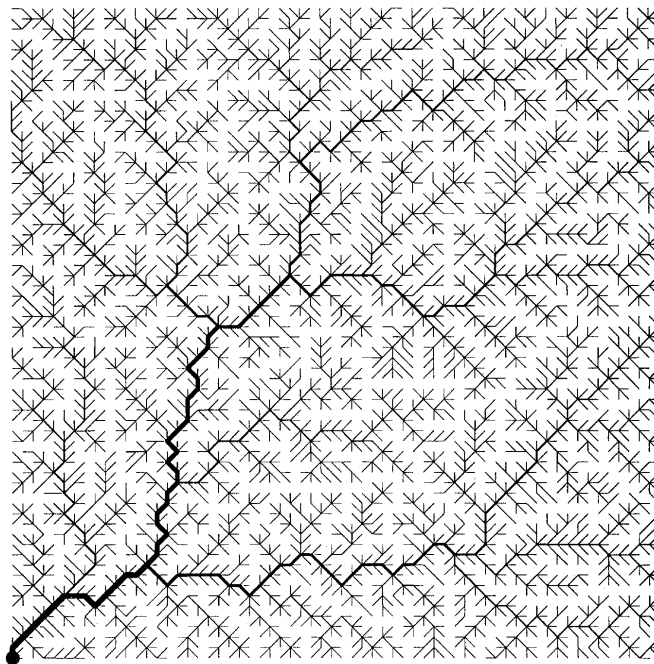


Abbildung 11.5: A computed river network by A. Rinaldo that shows similar statistical properties as natural rivers do.

A striking feature of the one mentioned and similar systems is their optimized structure in the stationary state. The contrary processes of uplifting and erosion smooth the landscape into a so-called “optimal channel network” by reducing the global energy, or potential, of

¹⁵This equation was cited from [5]. [4] mentions another formula: $stress = \sqrt{As}$

the system, comparable to systems in thermal equilibrium. This indicates that the critical state is attractive¹⁶. 1998 Marani et al. modelled an open electrical system where self-organized criticality was also supposed to form fractal patterns. In their publication they concluded:

“Whether or not dynamical rules necessarily obey a variational principle, and thus whether fractal forms in nature are or are not metastable states of a dynamical search for optimality, our results suggest that one ingredient for the emergence of fractality is the existence of a set of stationary or recursive states satisfying a global constraint.¹⁷”

Of course these models pose no sufficient proof for a general link between the two theories, but they certainly state that the thoughts presented in section 11.2.4 are worth further research.

11.3.3 Other SOC Computer Models

Up to now, a wide range of systems was corroboratively modelled by assuming self-organized criticality. Some of them are briefly discussed in this section:

Physical systems: SOC was first applied on a number of physical processes, e.g. light emission of quasars, super-conductive materials, and droplet formation on a glass surface. Computer models based on the assumptions of SOC urged that the theory can describe the gathered empirical data. For further reading, consult [4] and [6].

Geological systems: Similar to the “toy”-avalanches on sandpiles, earthquakes and volcanos were modelled as well as the expansion of forest fires. Their outcomes both corresponded to time- and space-dependent structures of real geophysicist processes.

Traffic Jams: K. Nagel and M. Schreckenberg constructed a one-dimensional cellular automaton to study traffic jams on a one-lane highway. Each cell can carry exactly one car that can accelerate, if the highway in front of it is empty, and break, if another car occupies the next cell. The attractive, stationary state of the automaton is critical and the size of traffic jams occurs in $1/f$ -noise manner. Furthermore, this state is optimized in the meaning that, inspite of jams, the achievable throughput of cars is maximized.

Biology: Nowadays SOC has found its way into evolutionary theory. It could give an outstanding explanation to the postulated saltationism proclaimed by evolutionists: Evolution of both biodiversity and the fitness of a single population is said to be not gradual but erratic — like the extinction of the dinosaurs to name one of the most momentous “catastrophes”. A computer model of Bak and Sneppen lead to the proclaimed saltational behavior¹⁸.

¹⁶see [4]. Note that the systems aim only to local optima, thus the process is a kind of heuristic hill-climbing in the rough phasespace of complex systems.

¹⁷[7]

¹⁸see [6]

The Game of Life: Last but not least, this famous automaton invented by John Conway¹⁹ shows characteristic critical behavior: A little perturbation of the automaton in stationary state leads to dynamics on all scales. It might irritate that the theoretical concept of SOC was applied on an abstract automaton that does not correspond to any natural phenomenon. What is the gain if one views the Game of Life (GoL) to be critical? In fact, GoL is the canonical example to complexity research, like the sandpile to SOC. The complex behavior of GoL together with it being critical is definitely the most powerful interconnection between the theories. This will be briefly discussed in the last section.

11.4 Further implications of SOC

11.4.1 SOC and the Edge of Chaos

The concept of phase transitions in thermodynamics was adapted to the Game of Life nearly two years before the idea of self-organized criticality had been introduced. C. Langton compared cellular automata with thermodynamical systems. There are cellular automata with rules that lead (after a short period of time) to stationary or periodic structures (Wolfram's class 1 and 2 automata) while others show chaotic fluctuations without any recognizable patterns (Wolfram's class 3). GoL lies on this edge of chaos and order. It is a representant of automata (Wolfram's class 4) which are able to produce complex structures, for instance it has been proven that one can compute boolean algebra using special configurations in GoL²⁰.

Langton set this edge of chaos in relation to the phase transitions in thermodynamic critical states. This comparison was motivated by the fact, that perturbations can in both systems extend through the whole available space. While this corresponds exactly to the theory of Bak, Tang and Wiesenfeld, the implications Langton aimed at, differed from the one of SOC in a drastic manner: Bak, Tang and Wiesenfeld claimed to explicate catastrophes and $1/f$ -noise, while Langton is interested in complexity, computation and intelligent or individual behavior. "Can a sandpile do algebra?" was the provocative question with which Waldrop commented the unresolved discrepancy between the diverging concepts.

11.4.2 Possibilities to predict

As the theory of self-organized criticality deals with statistical properties, it forgoes any possibility to predict a real system's behavior. For instance it cannot predict the next earthquake but only the probability of its magnitude. However, it urges some general implications on predictions: It was argued that catastrophes are the consequence of small momenta traveling great distances within a system. If one observes the system locally, one can certainly predict events caused by momenta within the radius of observation, but there will always be influences from regions out of the observation range:

"If this picture is correct for the real world, then we must accept instability and catastrophes as inevitable in biology, history, and economics. Because the

¹⁹For a description of Game of Life, refer to standard literature or [9]

²⁰The Game of Life is even Turing-equivalent, thus one can compute algebraic functions on natural numbers!

outcome is contingent upon specific minor events in the past, we must also abandon any idea of detailed long-term determinism or predictability.²¹”

On the other hand, the quality of prediction corresponds to this observation range in the manner of the described power law. If the dynamics were chaotic, the real value would deviate exponentially from the predictions. In spite of this, dealing with a power law ensures, that the deviation will be only polynomial.

Conclusion: Although SOC gave rise to criticism by a couple of experts in several disciplines (as most universal theories do), the concept can explain yet ununderstood phenomena. Its implications match empirical data. Furthermore, it fits with other theories, like fractality and the “Edge of Chaos”-concept in complexity research. Thus, it has firmly found its place in the field of complexity theory. However, a lot of research has still to be done on the links between the theories, as well as the mathematical foundations.

²¹[5] page 32.

Literaturverzeichnis

- [1] Adami, C. (1995). *Self-organized criticality in living systems*. Physics Letters, July.
- [2] Alstrom, P, Leao, J. (1994). *Self-organized Criticality in the Game of Life*. Physical Review E49.
- [3] Bak, P., Chen, K. (1991). *Selbstorganisierte Kritizität*. Spektrum der Wissenschaft, May.
- [4] Bak, P., Paczuski, M. (1993). *Why nature is complex*. Physics World, December.
- [5] Bak, P. (1997). *How nature works. The science of self-organized criticality*. Oxford University Press, Oxford, Melbourne, Tokyo.
- [6] Jensen, H. J. (1998). *Self-Organized Criticality. Emergent Complex Behaviour in Physical and Biological Systems*. Cambridge University Press, Cambridge, New York, Melbourne.
- [7] Marani, M., Banavar, J. R., Caldarelli, G., Maritan, A., Rinaldo, A. (1998). *Stationary self-organized fractal structures in an open, dissipative electrical system*. J. Phys. A: Math. Gen. 31, L337-L343.
- [8] Rinaldo, A., Rodriguez-Iturbe, I. (1997). *Fractal River Basins. Chance and Self-Organization*. Cambridge University Press, Cambridge, New York, Melbourne.
- [9] Waldrop, M. M. (1992). *Complexity. The Emerging Science at the Edge of Order and Chaos*. Simon & Schuster, New York.

Beiträge des Instituts für Umweltsystemforschung der Universität Osnabrück

- Nr. 1: Eberhard Umbach: Umweltverträgliches Wirtschaftssystem in den Bereichen Abfall und Emissionen. März 1997.
- Nr. 2: Stefan Trapp, Bernhard Reiter, Michael Matthies: Überprüfung und Fortentwicklung der Bodenwerte für den Boden-Pflanze-Pfad - Teilprojekt Transferfaktoren Boden-Pflanze. August 1997.
- Nr. 3: Michael Matthies (Hrsg.): Stoffstromanalyse und Bewertung. September 1997.
- Nr. 4: Dirk Melcher: Quantifizierung, Klassifizierung und Modellierung der Phytotoxizität organischer Chemikalien. Oktober 1997.
- Nr. 5: Stefan Schwartz: Organische Schadstoffe in der Nahrungskette - Vorstudie zur Validierung von Expositionsmodellen. November 1997.
- Nr. 6: Volker Berding: Private Hausbrunnen - Vergleichende Bewertung von Maßnahmen zur Verbesserung der Trinkwasserqualität. Oktober 1997.
- Nr. 7: Horst Malchow (Hrsg.): Modellbildung und -anwendung in den Wissenschaften I. Januar 1998.
- Nr. 8: Birgit Radtke: Bifurkationen in einem Modell mariner Planktondynamik. Januar 1998.
- Nr. 9: Werner Berens: Konzeption eines Umweltinformationssystems für die Universität Osnabrück. Juni 1998.
- Nr. 10: Michael Matthies (Hrsg.): Studienprojekte 1998. September 1998.
- Nr. 11: Michael Matthies (Hrsg.): Globaler Wandel. September 1998.
- Nr. 12: Klaus Brauer (Hrsg.): Institutsbericht. September 1998.
- Nr. 13: Klaus Brauer, Horst Malchow, Michael Matthies, Eberhard Umbach (Hrsg.): Materialien des Arbeitstreffens Systemwissenschaft in der Lehre, Universität Osnabrück, 29./30.9.1998. Dezember 1998.
- Nr. 14: Horst Malchow (Hrsg.): Modellbildung und -anwendung in den Wissenschaften II. Dezember 1998.
- Nr. 15: Horst Malchow (Hrsg.): Modellbildung und -anwendung in den Wissenschaften III. August 1999.
- Nr. 16: Michael Matthies (Hrsg.): Regionale Nachhaltigkeit. September 2000.
- Nr. 17: Markus Klein: Langjähriger Wasserhaushalt von Gras- und Waldbeständen. Entwicklung, Kalibrierung und Anwendung des Modells LYFE am Groß-Lysimeter St. Arnold. Juni 2000.

- Nr. 18: Markus Brune: Multimediale Umweltmodellierung mit Fuzzy-Mengen. Juli 2000.
- Nr. 19: Michael Matthies (Hrsg.): Fraktale in Hydrologie und Biologie. Oktober 2000.
- Nr. 20: Stefan Fuest (Dissertation): Regionale Grundwassergefährdung durch Nitrat. Dezember 2000.
- Nr. 21: Carsten Schulze (Dissertation): Modelling and evaluation the aquatic fate of detergents. April 2001.
- Nr. 22: Horst Malchow (Hrsg.): Modellbildung und -anwendung in den Wissenschaften IV. April 2001.

ISSN 1433-3805

Die Beiträge können gegen einen Selbstkostenpreis (ca. 10 – 20 DM pro Exemplar) beim Institut für Umweltsystemforschung, Universität Osnabrück, 49069 Osnabrück bestellt werden.