

Räumlich organisierte, biomolekulare Evolution

I n a u g u r a l - D i s s e r t a t i o n

zur

Erlangung des Doktorgrades

der Mathematisch-Naturwissenschaftlichen Fakultät

der Universität zu Köln

vorgelegt von

Stephan Wolfgang Altmeyer

aus Bottrop

19. Dezember 2001

Berichterstatter: Prof. Dr. Michael Lässig
Prof. Dr. John S. McCaskill
Prof. Dr. Dietrich Stauffer

Tag der mündlichen Prüfung: 11.12.2001

Inhaltsverzeichnis

1	Einleitung	11
1.1	Biomolekulare Evolution	11
1.1.1	Koevolution und Parasiten	12
1.1.2	Hyperzyklen und Informationskrise	12
1.1.3	Kompartimentierung, Spiralen und Parasiten	13
1.1.4	Kollaps durch Mutation	14
1.1.5	Der Inhalt dieser Arbeit	15
2	Methoden und Konzepte	17
2.1	Analytische Konzepte	17
2.1.1	Das PRESS Modell	17
2.1.2	Eigenschaften des PRESS Modells	18
2.1.3	Zustände und Raten	18
2.1.4	Master Gleichung	20
2.1.5	Stationäre Zustände	20
2.1.6	Selbstkonsistenz	21
2.2	Numerische Simulation	23
2.2.1	Gittertopologie	23
2.2.2	Simulation von Diffusion	23
2.2.3	Simulation von Reaktionen	24
2.2.4	Simulationsprogramm	24
3	Selbstreplikatoren	27
3.1	Homogene Quasispezies	27
3.1.1	Modell der Evolution	27
3.1.2	Vereinfachende Modellannahmen	30
3.2	Quasispezies ohne Zerfall im PRESS Modell	33
3.2.1	Lokal konstante Populationen	33
3.2.2	Das 1-Spezies Modell	33
3.2.3	Kritische Fehlerrate $R_c(m)$	34
3.2.4	Skalierungsverhalten	35
3.2.5	Selbstreplikatoren im Raum	37
3.3	Quasispezies mit Zerfall im PRESS Modell	39
3.3.1	Ratengleichungen mit Zerfall	39
3.3.2	Verallgemeinerung des PRESS Modells	40
3.3.3	Kritische Migrationsraten	43

3.3.4	Charakterisierung der m_1 und m_2 Übergänge	46
3.3.5	Der Einfluß expliziten Zerfalls	47
4	Kooperierende Katalysatoren	53
4.1	Katalytisches Modell ohne Erkennung	53
4.1.1	Modell altruistischer Katalysatoren	53
4.1.2	Reduzierung der Funktionen	54
4.1.3	Zustände und Raten	55
4.1.4	Nichtlineare Dynamik	56
4.1.5	Stabilisierung von Kooperation	58
4.1.6	Hopf Bifurkation und Grenzzyklen	60
4.1.7	Kritische Fehlerraten	62
4.1.8	Simulationsergebnisse und euklidischer Raum	62
4.1.9	Vergößerung der Parzellenpopulation	63
4.2	Partielle Sequenzerkennung	69
4.2.1	Reduzierung des Sequenzraums	69
4.2.2	Information und Struktur	72
4.2.3	Reaktionen bei Teilerkennung	73
4.2.4	Teilerkennung im Simplexmodell	75
4.2.5	Sequenzverlängerung	76
5	Simulation mit konfigurierbarer Hardware	83
5.1	Der Parallelrechner NGEN	83
5.1.1	Konventionelle Rechnerarchitekturen	83
5.1.2	FPGA basierende Architektur	84
5.1.3	Hardwarekomponenten	85
5.1.4	Datenflußarchitektur	85
5.2	Parallele, dreidimensionale Diffusion	88
5.2.1	Algorithmische Struktur	88
5.2.2	Erfüllung der Diffusionsgleichung	88
6	Diskussion und Ausblick	93
6.1	Diskussion der Ergebnisse	93
6.1.1	Zusammenfassung und Schlußfolgerung	93
6.1.2	Quo vadis?	95
A	Beweise, Ergänzungen und Vertiefungen	99
A.1	Beweis der Rekursionsgleichung	99
A.2	Rückmutationswahrscheinlichkeit	100
A.3	Genauigkeit von Zufallszahlen	101
A.4	Eigenschaften der Matrix \mathbf{L}	102
A.5	Erzeugung der Matrix \mathbf{L}	102
A.6	Matrix \mathbf{L} für Quasispezies mit Zerfall	103
A.7	Transformation von \mathbf{L} zu \mathbf{L}'	103
A.8	Homogene Quasispezies mit Zerfall	104
A.9	Homogenes Kooperationsmodell	104
A.10	Erzeugung der Matrix \mathbf{R}	105

A.11 Lineare Stabilitätsanalyse	106
A.11.1 Allgemeine Begriffe	106
A.11.2 Stabilitätskriterien	107
A.11.3 Anwendung auf das Simplex Modell	109
A.12 Zweidimensionaler Margolus-Algorithmus	110
A.13 Internen Zustände und externe Koordinaten	111
A.14 Implementierung in Hardware	112
Literatur	114
Curriculum Vitae	121
Deutsche Zusammenfassung	123
English abstract	125
Danksagungen	127

INHALTSVERZEICHNIS

Abbildungsverzeichnis

1.1	Hyperzyklus Systeme	14
1.2	Entstehung von Hyperzyklusspiralen	15
2.1	Simplex Topologie	19
2.2	Lineare Kette von Zuständen	21
2.3	Dimension und Nachbarschaft	24
2.4	Schaubild äußere Programmschleife	25
3.1	Homogenes Quasispeziesmodell	31
3.2	Zustände im 1-Spezies Modell	35
3.3	Einfache Quasispezies Analyse	35
3.4	Vergleich der exakten Werte von $R_c(m)$	36
3.5	Skalierungsverhalten der Quasispezies	37
3.6	Vergleich Simplex und endliche Dimensionen	38
3.7	Zustände im zwei-Spezies Modell	41
3.8	Bereiche der Quasispezies mit Zerfall	45
3.9	Die Größe $R_c(m)$ für unterschiedliche D	49
3.10	Phasendiagramm der Quasispezies	50
3.11	Hohe Zerfallsraten	50
3.12	Vergleich Analyse und Simulation	51
4.1	X - und Y -Populationszahlen	59
4.2	Migrationsschnitte der Populationszahlen	61
4.3	Stabilität in Abhängigkeit von m	65
4.4	Stabilitätsbereiche des kooperativen Systems	66
4.5	Kritische Fehlerraten im kooperativen Modell	66
4.6	Verhalten in physikalischen Raumdimensionen	67
4.7	Skalierung der physikalischen Raumdimensionen	67
4.8	κ/n für unterschiedliche n	68
4.9	ζ/n für unterschiedliche n	68
4.10	Partielle Sequenzerkennung	70
4.11	Bijektive Faltungsabbildung	74
4.12	Bindung gefalteter RNA-Sequenzen	75
4.13	Vergleich lineare und Faltungserkennung	76
4.14	Stabilitätserhöhung ohne Sequenzverlängerung	79
4.15	Stabilitätserhöhung mit Sequenzverlängerung	80
4.16	Sequenzverlängerung bei Teilerkennung	81

ABBILDUNGSVERZEICHNIS

5.1	Struktur eines CLBs	86
5.2	Datenfluß durch das dreidimensionale Volumen	87
5.3	Die vier Operationen des Diffusionsalgorithmus	89
5.4	Einbettung eines $(2 \times 2 \times 2)$ Kubus	90
5.5	Zeitlicher Verlauf der Aufenthaltswahrscheinlichkeit $P(x, t)$	91
5.6	FWHM als Funktion unterschiedlicher p	91
5.7	Diffusionskoeffizient und Rotationswahrscheinlichkeit	92
A.1	Phasenraumabbildung	108
A.2	Stabile und instabile Fixpunkte	109
A.3	Funktionsweise des 2D Margolus Algorithmus	110
A.4	Schaltplan der Diffusionsebene	113
A.5	Schaltplan eines Ausschnitts	113

Tabellenverzeichnis

3.1	Zustände im Quasispezies Modell mit explizitem Zerfall	41
4.1	Zustände im Kooperationsmodell mit explizitem Zerfall	57
4.2	Anzahl der Mutationen in Abhängigkeit von Q	72
4.3	Erkennungsregion und -wahrscheinlichkeit	76
6.1	Vergleich der Modelleigenschaften	96
A.1	Interne Zustände und Koordinaten	111
A.2	Zustandsänderung bei Neufüllung	111
A.3	Zustands- und Koordinatenwechsel	111

TABELLENVERZEICHNIS

Kapitel 1

Einleitung

1.1 Biomolekulare Evolution

Die Arbeit hat sich zum Ziel gesetzt, das Verhalten von biologischer Evolution in räumlichen Umgebungen zu untersuchen.

Der Begriff der Evolution ist auf Darwin zurückzuführen [1], der diesen 1859 definierte. Er beschreibt die Vorstellung, daß die Lebensformen veränderlich sind und dies unter dem Druck der Mutation und Selektion geschieht. Im Laufe der Zeit etablierte sich der Begriff der Gene, Informationsspeichernde Sequenzen, die vollständig die Funktion und die Gestalt der Lebensformen zu bestimmen schienen. Heutzutage ist die moderne Biologie in der Lage, die Gene direkt zu manipulieren.

Das - theoretische - Verständnis über die Genetik wäre jedoch nicht soweit, wenn es die Zusammenarbeit der Biologie mit der Physik nicht gegeben hätte. So stellte sich bereits 1944 Schrödinger die Frage „Was ist Leben?“ [2] und zeigte Verbindungen vor allem zur statistischen Physik auf, die auch nach 60 Jahren intensiver Forschung weitgehend ihre Gültigkeit behalten haben.

Einen chronologischen Abriß über den Einfluß der Physik auf die Biologie zu geben, würde den gegebenen Rahmen deutlich übersteigen. Um trotzdem einen Einstieg in die Prinzipien und Probleme evolvierbarer Systeme zu geben, werden die zwei von Eigen eingeführten Modelle der Quasispezies [3] und der Hyperzyklen [4] benutzt, da anhand ihnen gut die Eigenschaften evolvierbarer Systeme veranschaulicht werden können. Dazu geben sie eine gewisse Haupttrichtung an, an der sich viele Modellansätze der letzten Jahre orientierten.

Gerade das von Eigen eingeführte physikalische Modell der Selbstreplikatoren, das Quasispezies System, das in Kapitel 3.1.1 ausführlich erläutert wird, war der erste weitreichende physikalische Ansatz, die Prinzipien von Mutation und Selektion zu beschreiben. Es sagt voraus, wie sich eine Population von Selbstreplikatoren unter dem Druck von Mutation und Selektion entwickeln wird. Dabei zeigt sich, daß eine Größe maximiert, die als Fitneß bezeichnet wird und die - grob gesagt - die Anpassung einer Spezies an ihre Umwelt definiert. Nicht nur die nahe Verwandtschaft der Fitneß zur Entropie zeigt die direkte Beziehung zur Physik auf. In späteren Arbeiten konnte gezeigt werden, daß das Quasispezies Modell exakt auf ein zweidimensionales Ising Modell abbildbar [5, 6] und auch mit den Methoden der Feldtheorie zu bearbeiten ist [7] (bzw. im allgemeineren Kontext [8]).

Die Ergebnisse dieser Arbeit konnten viele Erscheinungen auf der Ebene von Zellen bzw. Phagen erklären [9, 10] (beispielweise die Rate der mutativen Veränderungen). Gleichzeitig

war es abstrakt genug, um auch Ergebnisse im informationstheoretischen Sinn zu liefern. Wahrscheinlich hat kein bisheriges Modell so viele Einblicke in die Mechanismen der molekularen Informationsverarbeitung geliefert, wie das Quasispezies System. Trotz aller Vorteile, eignet es sich wenig, die ersten Zeitpunkte der Evolution zu beschreiben, da die Annahme von Selbstreplikatoren bereits - im molekularen Sinne - sehr komplexe Systeme voraussetzt.

Die ersten evolutionären Systeme mußten also wechselwirkender Natur sein. Ob sie bereits eine vergleichbare Komplexität heutiger Strukturen wie DNA/Protein hatten oder ob sie aus katalytischen RNA (RNA-Welt Hypothese) aufgebaut waren, ist bis heute noch nicht geklärt. Im Gegensatz zu Veröffentlichungen, die sich direkt mit der Frage nach dem Ursprung des Lebens befassen [11, 12, 13], soll die vorliegende Arbeit möglichst kontextfrei die Prinzipien räumlich aufgelöster Systeme behandeln. Welche Probleme es in diesem Zusammenhang gibt und wie mögliche Lösungsansätze aussehen, zeigen die folgenden Kapitel.

1.1.1 Koevolution und Parasiten

Wechselwirkende Systeme, die im Darwinschen Sinne als evolvierbar zu bezeichnen sind, haben fast immer das Problem, daß aufgrund von Mutationen Strukturen auftreten, die ungehemmt die Ressourcen des Systems ausnutzen und es unter Umständen destabilisieren können. Solche Strukturen, die nur zum eigenen Vorteil, ohne Gegenleistung und zum Schaden des Teil- oder gar des Gesamtsystems agieren, werden als parasitär bezeichnet. Sie treten in fast allen hinreichend komplexen, evolutiven Umgebung auf. Und obwohl es effiziente Methoden des Schutzes vor parasitären Angriffen gibt, zeigt die Erfahrung, daß es zumindest in komplexen, wechselwirkenden Systemen keine endgültige Abwehr vor Parasiten zu geben scheint, da sich diese ebenfalls evolutiv anpassen können und dabei versuchen, Schwachstellen des Systems zu enttarnen und auszunutzen, bzw. die neuen Schutzmechanismen für sich nutzbar zu machen. Die Veränderungen, die mit der Entwicklung eines besseren Schutzes vor Parasiten und den damit verbundenen Anstrengungen der Parasiten, sich wiederum effektivere Angriffsmethoden anzueignen, werden ebenso als Koevolution [14, 15] bezeichnet, wie auch die gemeinsame Entwicklung symbiotischer Spezies. Die besondere Art der Koevolution bei konkurrierenden Gruppen, die trotz enormer Weiterentwicklung keine der beiden in Vorteil bringt, wird als Red Queen Effekt bezeichnet [16].

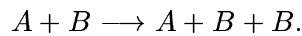
Aus physikalisch motivierten Aspekten, wie etwa eine reduzierte Erkennungsmöglichkeit, ist ersichtlich, daß es allgemeine Schutzmechanismen in interagierenden, molekularen Systemen gegen Parasitenbefall nicht geben kann. Deren Existenz ist folglich ein Kernpunkt, der bei der Behandlung evolvierbarer Systeme zu beachten ist.

1.1.2 Hyperzyklen und Informationskrise

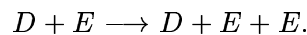
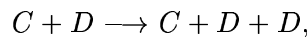
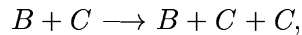
Neben der Frage nach dem Wechselspiel von Host und Parasit hat die Natur jedoch noch weitere Aufgaben zu lösen. Eine davon ist die Speicherung von (genetischer) Information. Das prinzipielle Problem der Informationsspeicherung bei den Selbstreplikatoren ist dabei zum einen, daß innerhalb des Modells immer kritische, maximale Sequenzlängen existieren, die direkt die Menge der Information pro Sequenz determinieren. Zum anderen erweist sich eine Aufteilung der Information auf unterschiedliche Sequenzen ebenso kritisch, da bereits marginale Unterschiede in den Replikationsraten dazu führen, daß in einer Population von Selbstreplikatoren nach einer gewissen Zeitdauer immer nur eine bestimmte Gruppe von Spezies (die Quasispezies) dominiert. Die Konzentrationen aller übrigen Spezies werden mit der Zeit exponentiell

unterdrückt, so daß konsequenterweise die Gesamtinformation fast vollständig verloren geht.

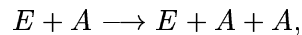
Gäbe es eine Möglichkeit, die Anzahl bestimmter Spezies unter allgemeinen Bedingungen zu stabilisieren, welche nämlich gerade die Träger der gewünschten Information sind, hätte man das Problem der Informationskrise überwunden. Gerade dieser Struktur entspricht das Hyperzyklus Modell, das Eigen 1977 veröffentlichte [4, 17, 18]. In diesem System wird die Idee eine direkten Kopplung von selbstreplizierenden Informationsträgern manifestiert. Die Kopplung erfolgt dabei über eine katalytische Replikation, deren Rate im Vergleich zu der Rate der Selbstreplikation dominiert und wird so definiert, daß sich eine Sorte von Molekülen nur mit Hilfe genau einer anderen Sorte vervielfältigen kann. Wird die Gesamtinformation beispielsweise auf fünf Sequenzen (zur Stabilität von Hyperzyklussystemen in Abhängigkeit der Spezies siehe [19]) A, B, C, D, E verteilt, so kann sich Sequenz B nur mit Hilfe Sequenz A katalytisch reproduzieren, was auch mit Reaktionsgleichungen verdeutlicht werden kann,



In der selben Weise reproduzieren sich auch alle anderen Sequenzen,



Nun wird noch Sequenz E als Replikator von A definiert,



womit sich der Kreis der Replikationen schließt. Abb. 1.1(a) zeigt solch einen 5 komponentigen Hyperzyklus schematisch. Das Prinzip des Hyperzyklus wurde von Eigen und Schuster im Zusammenhang mit Darwinschen Systemen als deren nächst höhere Organisationsstufe bezeichnet [20].

Das System weißt nun die interessante Eigenschaften auf, daß es repulsiv ist [14], was bedeutet, daß wenn die Konzentration einer Sequenz als Folge einer Störung nahe gegen Null geht (jedoch nicht Null wird) nach einer Relaxationszeit wieder beliebig hohe Werte erreichen wird. Einmalige Störungen treiben das System auf Grenzyklen, auf denen die Konzentration abwechselnd periodisch ihr Minimum und Maximum durchlaufen. Somit wird keine Sequenz des Hyperzyklus aus dem System verdrängt, selbst wenn sich die katalytischen Fähigkeiten unterscheiden. Nachdem nun eine Methode gefunden wurde, mehrere Sequenzen parallel nebeneinander zu stabilisieren und es schien, daß die Informationskrise auf diese Art gelöst werden konnte, zeigte sich erneut das Problem der Stabilität in der Gegenwart von Parasiten.

Entsteht unter dem Druck der Evolution eine parasitäre Sequenz, $C1$, die zwar wie Sequenz C von B kopiert wird, jedoch selbst keine weitere Sequenz des Hyperzyklus repliziert, vergl. Abb. 1.1(b), so kann man mit Hilfe von Differentialgleichungen zeigen, daß der Parasit $C1$ alle Mitglieder des Zyklus im Laufe der Zeit verdrängt.

Diese Erkenntnis stellte das ganze Konzept des Hyperzyklus in Frage.

1.1.3 Kompartimentierung, Spiralen und Parasiten

Eine Möglichkeit, einen - wenngleich auch unvollkommenen - Schutz gegen Parasiten zu entwickeln, ist der Natur zu entnehmen. Dort findet in allen höheren Lebensformen eine

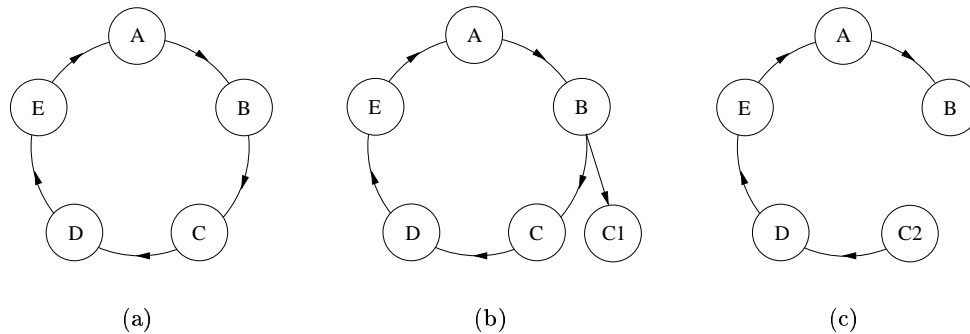


Abbildung 1.1: Schematische Darstellung eines perfekten (a) Hyperzyklussystems, eines Systems (b) mit Parasit $C1$ und eines aufgrund einer Mutation von $C \rightarrow C2$ unterbrochenen Zyklus (c).

(Zell)Kompartimentierung statt. Ihre separierende Wirkung verhindert ein direktes Ausbreiten eines Parasiten über das gesamte System. Er ist somit nicht mehr in der Lage, global Ressourcen zu beanspruchen. Ein lokaler Parasitenbefall hingegen führt in den seltensten Fällen zum vollständigen Kollaps des Systems. Bereits Eigen hatte schon in seiner frühen Arbeit zeigen können, daß Kompartimentierungen ein probates Mittel gegen den Befall von Parasiten sind [21]. Einen gewissen Durchbruch kam mit den Arbeiten von Hogeweg und Boerlijst, die zeigen konnten, daß allein der kontinuierliche Raum ausreichend ist, um die Hyperzyklen effektiv vor den Parasiten zu schützen [22, 22]. Bei ihren Untersuchungen auf einem zweidimensionalen Gitter zeigten sich in dieser räumlich strukturierten Umgebung völlig neue Eigenschaften. Es stellt sich heraus, daß sich aus anfänglich völlig ungeordneten Strukturen räumliche Muster etablieren - rotierende Spiralen, siehe Abb. 1.2. Diese Spiralen haben die Eigenschaft, eindringende Parasiten an deren Ausbreitung über das System zu verhindern, da sich die $C1$ Parasiten nur in der Nähe der B Sequenzen aufhalten können. Gelingt es jedoch den Parasiten in das Zentrum der Spirale zu gelangen, was gegen die Rotation erfolgen muß, so verdrängen sie allmählich von innen nach außen wachsend die C Sequenzen und zerstören vollständig die Spiralstruktur. Dies passiert jedoch nur dann, wenn sie eher zufällig in das Zentrum gelangen. Die Struktur der Spiralen erweist sich unter einem breiten Parameterbereich (Diffusionsrate, Replikationsraten) als sehr effektiver Schutz. Die Spiralmusterbildung und deren stabilisierende Wirkung konnte in nachfolgenden Arbeit verifiziert werden [23] und wurde im Hinblick auf die Bildungsgeschwindigkeit weiter untersucht [24]. In der Natur ist die Bildung von Spiralmuster bei den Belousov-Zhabotinsky Reaktionen zu beobachten [25, 26], die häufig mit den Methoden der partiellen Differentialgleichungen studiert werden [27].

Auch in weiteren biologisch motivierten Systemen zeigt sich der stabilisierende Einfluß der Musterbildung [28, 29]. Selbst begrenzte, lokale Strukturen wie die selbstreplizierenden Spots [30] oder die Clusterkompartimentierung [31] können bereits effektiv molekulare Systeme stabilisieren.

1.1.4 Kollaps durch Mutation

Trotz der scheinbaren Rettung des Hyperzykluskonzepts durch die Einbettung in den metrischen Raum, erweist sich der Ansatz der strengen Kopplung als recht artifizuell. Die Tatsache, daß immer genau eine Sequenz von genau einer anderen kopiert wird, impliziert eine 100%ige

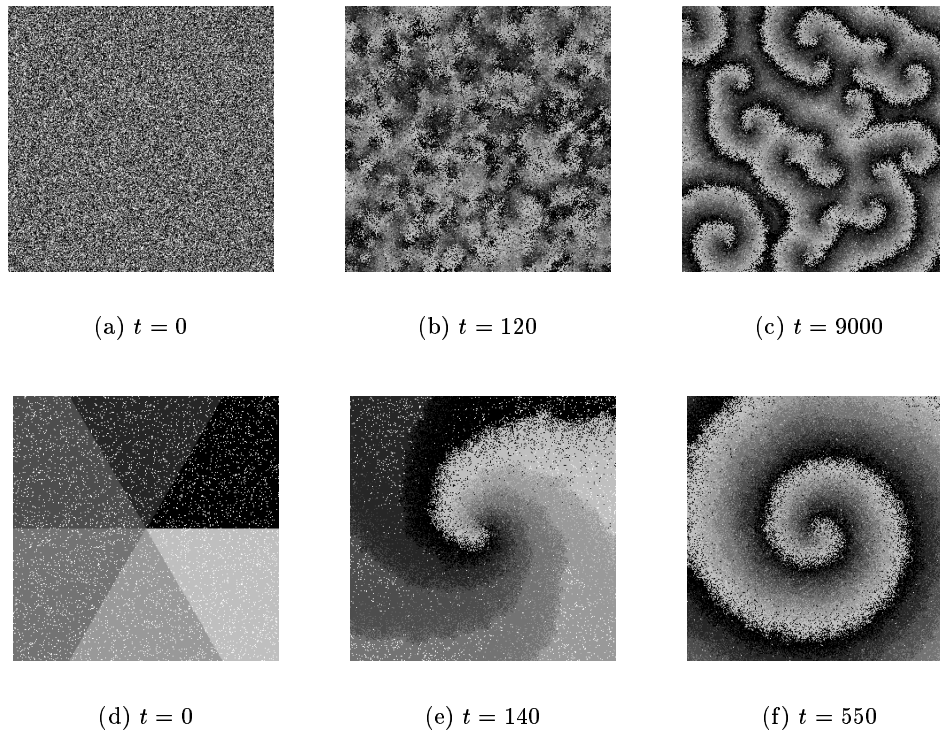


Abbildung 1.2: Die Entstehung von Spiralen in einem Hyperzyklussystem, mit ungeordneten (a-c) und geordneten (d-f) Anfangsbedingungen zu relativen Zeiten.

Erkennung. Eine Spezies kann von allen anderen fehlerfrei unterschieden werden, wenn ihre vollständige Sequenz gescannt und als die Richtige erkannt wird. Dies geschieht aber nur dann, wenn die zu kopierende Sequenz mit einer „internen Datenbank“ verglichen werden kann, was nichts anderes bedeutet, als daß der Kopierer neben seiner eigenen, auch noch die Information der zu kopierenden Sequenz enthalten muß. Letztlich führt das zurück zur Informationskrise.

Erkennt der Kopierer die zu kopierende Sequenz jedoch nicht zu 100%, so kann es vorkommen, daß bereits wenige Mutationen von zum Beispiel C zu $C2$ eine Erkennung unmöglich machen. Die Folge wäre, daß B fortan $C2$ nicht mehr kopiert, vergl. Abb. 1.1(c), und der komplette Hyperzyklus, angefangen bei $C2$ aussterben würde.

Die Notwendigkeit, (aus z.B. den beschriebenen Gründen der Informationsspeicherung) wechselwirkende Systeme zu etablieren, bringt immer das Problem mit sich, daß diese gegenüber Parasitenbefall geschützt werden müssen, wenn sie als evolvierbar bezeichnet werden sollen. Auch im Hinblick auf die prebiotische Evolution, die noch keinerlei übergeordnete Strukturen erlaubte, als auch aus Gründen einer universellen Verwendbarkeit muß die Stabilisierung unter sehr allgemeinen Bedingungen stattfinden.

1.1.5 Der Inhalt dieser Arbeit

Eine Antwort auf diese Problematik zu geben, hat sich diese Arbeit zum Ziel gesetzt. Dabei soll dies in genereller und analytischer Form und unter Berücksichtigung der diskreten Natur molekularer Prozesse geschehen. Daß diesem Anspruch nicht eine Person allein gerecht werden kann, liegt auf der Hand. So sind viele Konzepte und Analysen in der Kooperation mit

Prof. Dr. John McCaskill und Dr. Rudolf Füchsli entstanden.

Das aus dieser Kooperation entstandene Simplex Modell, daß in Kapitel 2 vorgestellt wird, stellt ein theoretisches Gerüst dar, mit dem das Verhalten von Reaktions-Diffusions Systemen vollständig analytisch zu berechnen ist und das dabei den diskreten Charakter molekularer Reaktionen, die insbesondere in dünn besiedelten Regionen unbedingt zu berücksichtigen sind [32, 33], modelliert.

Mit diesem System, daß aufbauend auf einer Inseltopologie die Berechnung der Zeitentwicklung von Zustandswahrscheinlichkeiten erlaubt, wird zunächst in Kapitel 3 der Einfluß des Raums auf die Klasse der Selbstreplikatoren untersucht, wobei sich zeigen wird, daß dieser Einfluß als durchweg negativ im Hinblick auf die Informationsspeicherung angesehen werden muß.

Anschließend wird in Kapitel 4 das einfachste, wechselwirkende System studiert, daß aus altruistischen Katalysten besteht, welche sich in der Gegenwart von Parasiten behaupten müssen. Zur Erklärung und Bedeutung des Altruismus in der Biologie siehe beispielweise [34, 35]. Es wird gezeigt, daß allein die Separation des Raums eine stabilisierende Kooperation zwischen den Altruisten induziert, die es erlaubt, der destabilisierenden Kraft der Parasiten entgegenzuwirken.

In einem weiteren Kapitel wird für das katalytische Modell die Wirkung von partieller Sequenzerkennung studiert. Die Möglichkeit der Faltung von Sequenzen wird sich im Hinblick auf eine Stabilitätershöhung als besonders wirkungsvoll herausstellen.

Für alle Modelle werden mit Hilfe von stochastischen Simulationen die Ergebnisse des PRESS Modells verifiziert und anschließend für physikalische Dimensionen bestätigt. In diesem Zusammenhang wird in Kapitel 5 der Parallelcomputer NGEN vorgestellt, für den eine dreidimensionale Topologie entworfen wurde. Dabei wurde ein Diffusionsalgorithmus konzipiert, mit dessen Hilfe in drei Dimensionen effizient Diffusion simuliert werden kann. Es wird gezeigt, daß der Algorithmus die Diffusionsgleichung erfüllt.

Kapitel 2

Methoden und Konzepte

2.1 Analytische Konzepte

Ein unendlich-dimensionales Inselmodell wird konzipiert, daß die analytische Untersuchung von Reaktions-Diffusionssystemen erlaubt und den diskreten Charakter molekularer Systeme berücksichtigt.

2.1.1 Das Press Modell

Die Grundidee des Modells ist es, auf einer ∞ -dimensionalen Simplextopologie (Inseltopologie) mit Hilfe der Diffusions- (Migrations-) und den Reaktionsraten die Besetzungswahrscheinlichkeit aller möglichen Zustände im stationären - nicht thermodynamischen - Gleichgewicht zu berechnen. Dabei wird die Beziehung zwischen den lokalen Besetzungszahlen und den Mittelwerten so ausgenutzt, daß selbstkonsistente Gleichungen entstehen, die implizit die Systemgrößen definieren. Die Idee einer Topologie, bei der unendlich viele einzelne „Inseln“ miteinander verbunden werden, wurde 1931 von Wright [36] erstmals zur Modellierung biologischer Prozesse vorgeschlagen.

Im Laufe der Zeit hat das Modell viele Veränderungen und Erweiterungen erfahren. Vor allem in der Populationsgenetik hat sich der Ansatz, eine Gesamtpopulation auf unendlich [37] oder endlich [38, 39] vielen Inseln zu verteilen, sehr bewährt. In einem ähnlichen Kontext, hat Kimura [40] die Simplex Topologie benutzt, um ein Modell herzuleiten, daß für den Grenzwert hoher Inselpopulationen eine Näherung für Diffusionsprozesse ergibt. Die hier verfolgte Idee zielt auf dünn besiedelte Populationen pro Insel ab und definiert sich zum einen über Zustandswahrscheinlichkeiten und trägt dem diskreten Charakter biomolekularer Prozesse Rechnung. Beide Eigenschaften zusammen definieren den Namen des PRESS Modells, probability reduced evolution of spatially-discrete species.

Um die Unterschiede zu den bisherigen Ansätzen deutlich zu machen, soll fortan neben den „Inseln“ auch der Begriff der „Parzellen“ benutzt werden, wenn explizit vom PRESS Modell die Rede ist. Ferner wird, um den molekularbiologischen Charakter der Untersuchungen zu betonen, neben der Sequenz auch der Begriff des Moleküls als Informationsträger eingeführt. Die abzuleitenden Gleichungen können bis zu einer gewissen Komplexität vollständig analytisch gehandhabt werden. Um einen Eindruck über die Grundlagen des Modells zu geben, werden im folgenden die Prozesse für ein nicht näher definiertes 1-Spezies System erörtert.

2.1.2 Eigenschaften des Press Modells

Gegeben sei eine Simplextopologie, welche aus unendlich vielen miteinander verbundenen Parzellen aufgebaut ist, siehe Abb. 2.1(a). Die Verbindungstopologie erlaubt es, daß einzelne Sequenzen beziehungsweise Moleküle von einer Parzelle zu einer beliebig anderen innerhalb eines Zeitschrittes migrieren können. Möglichen Reaktionen hingegen finden alle lokal auf den Parzellen statt.

Das Simplexmodell fordert für eine Exaktheit zwei fundamentale Eigenschaften:

1. Vollständige Symmetrie - alle Plätze sind bezüglich aller vorkommenden Prozesse gleichwertig.
2. Unendliche Population - obwohl die Zahl möglicher Moleküle pro Parzelle auf wenige beschränkt, ist die Gesamtzahl unendlich groß.

Diese Eigenschaften werden im folgenden fast durchgehend für die Rechtfertigung einzelner Schritte benötigt.

2.1.3 Zustände und Raten

Sei nun X eine beliebige Molekülsorte, die in beliebiger Weise reagieren kann. Ferner sei k die Anzahl dieser Moleküle und n die maximale Anzahl von Molekülen pro Parzelle, so daß $0 \leq k \leq n$ gilt. Nun sei P_k die Wahrscheinlichkeit, genau k Moleküle auf einer Parzelle zu finden. In Anbetracht der vollständigen Symmetrie des Simplexmodells - Eigenschaft 1 - sind auch die P_k für jede Parzelle identisch, das heißt, daß die Wahrscheinlichkeit (nicht die Zahl an sich) k Moleküle auf einer Parzelle zu finden unabhängig von der Wahl der Parzelle ist. Nun bestimmen die möglichen Reaktionen, die unabhängig voneinander in einem infinitesimalen Zeitschritt dt stattfinden, Übergangsraten, die die Zahl der Moleküle pro Parzelle verändern. So können Replikationen dazu führen, daß sich die Anzahl von Molekülen prinzipiell erhöhen kann. Zerfall hingegen kann bewirken, daß die Zahl von Molekülen auf einer Parzelle reduziert wird. Im allgemeinen wird es kinetische Prozesse geben, die mit einer gewissen Rate eine Veränderung der Population von k Molekülen auf k' bewirken.

Stellvertretend für mögliche Reaktionen soll $w_{k,k'}^{(\text{kin})}$ die Rate definieren, mit der diese Reaktionen ablaufen. Es wird sich zeigen, daß bei den betrachteten Reaktionen nur solche Prozesse vorkommen, die die Population um eins erhöhen, bzw. erniedrigen, also solche der Form $w_{k,k+1}^{(\text{kin})}$ und $w_{k,k-1}^{(\text{kin})}$. Die Reaktionsprozesse finden, wie bereits betont, nicht in Wechselwirkung mit anderen Parzellen sondern lokal statt.

Anders sieht es mit den möglichen Diffusionsprozessen, den Migrationen, aus. Dort kann eine einzelne Parzelle, die mit dem restlichen, unendlich großen Parzellenensemble in Verbindung steht, aufgrund der völligen Symmetrie mit jeder Parzelle des Ensembles zu einem beliebigen Zeitpunkt Teilchen austauschen. Die Wahrscheinlichkeit, daß aus dem Ensemble (innerhalb einer beliebig kleinen Zeitspanne dt) beispielsweise ein X -Molekül auf die Parzelle gelangt, ist proportional zu der gemittelten Anzahl aller X im Ensemble. Anschaulich kann das Parzellenensemble somit als eine unendlich große Parzelle betrachtet werden, siehe Abb. 2.1(b). Die Herausnahme der einzelnen Parzelle ändert wegen Eigenschaft 2 jedoch nicht den Zustand des unendlich großen Parzellenensembles im Vergleich zur Gesamtheit aller Parzellen.

Die Migration beschränkt sich im weiteren auf den Austausch eines Teilchen (oder Leerstelle) einer einzelnen Parzelle mit einem Teilchen (oder Leerstelle) aus dem Ensemble. Der

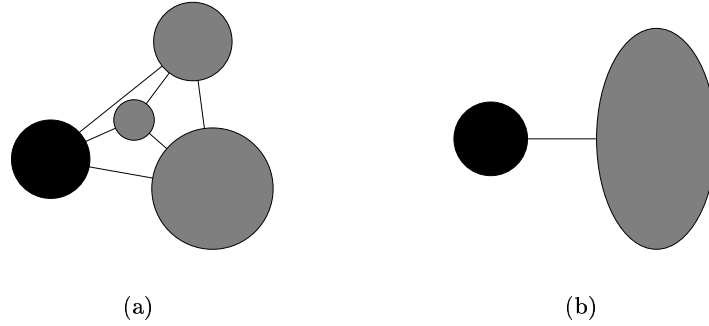


Abbildung 2.1: Die Abbildung (a) zeigt die Geometrie des (endlichen) Simplexmodells mit vier Parzellen. Greift man sich eine Parzelle heraus, so kann das übrige Ensemble aufgrund der völligen Symmetrie für den Migrationsprozeß als Pool betrachtet werden, Abb. (b)

Austauschprozeß garantiert, daß zu keinem Zeitpunkt mehr als n Moleküle pro Parzelle vorhanden sind, womit n für jede Parzelle konstant gehalten werden kann. Die Forderung nach einem variablen n würde die Analyse zum einen schwierig gestalten, da unter anderem auch die zentrale Eigenschaft 1 nicht mehr gewährleistet würde, und kann zum anderen durch größere n bei durchschnittlich kleiner Parzellenpopulation simuliert werden.

Um den Prozeß der Migration zu verdeutlichen, sollen die möglichen Veränderung in einem 1-Molekül System mit Leerstellen erörtert werden. In solchen Systemen gibt es zwei von vier möglichen Prozessen, die die Anzahl der X auf der einzelnen Parzelle verändern. Der erste Prozeß ist, daß ein X das Ensemble verläßt und zur einzelnen Parzelle migriert, wobei es mit einer Leerstelle ausgetauscht wird. Dies erhöht die Anzahl der X auf der Parzelle um eins und findet mit der Rate

$$w_{k,k+1}^{(\text{mig})} = m \frac{\bar{k}}{n} \cdot \frac{n-k}{n} \quad (2.1)$$

statt, wobei m den Migrationskoeffizient und \bar{k} die mittlere Anzahl von X im vollständigen Parzellenensemble darstellen. Der Term $(n-k)/n$ in Gl. (2.1) bezeichnet den Bruchteil von Leerstellen zur Gesamtzahl aller zur Verfügung stehenden Plätze auf einer Parzelle. Diesem Prozeß steht das emigrieren gegenüber bei dem ein X die einzelne Parzelle verläßt und mit einer Leerstelle aus dem Ensemble getauscht wird. Die Rate dazu ist

$$w_{k,k-1}^{(\text{mig})} = m \frac{n-\bar{k}}{n} \cdot \frac{k}{n}. \quad (2.2)$$

Mit den nicht näher spezifizierten Reaktions- und den Migrationsraten aus Gl. (2.1) und Gl. (2.2) ergeben sich nun die zusammengesetzten Übergangsraten, die letztlich die Anzahl der X auf den Parzellen bestimmen,

$$w_{k,k+1} = w_{k,k+1}^{(\text{kin})} + w_{k,k+1}^{(\text{mig})}, \quad (2.3)$$

$$w_{k,k-1} = w_{k,k-1}^{(\text{kin})} + w_{k,k-1}^{(\text{mig})}. \quad (2.4)$$

Die Addition ergibt sich aus der Unabhängigkeit der Prozesse, die aus den betrachteten infinitesimalen Zeitspannen resultieren.

2.1.4 Master Gleichung

Die Kinetik, also die tatsächliche Veränderung der Zustandswahrscheinlichkeiten, die durch die Übergangsraten induziert werden, wird stochastisch durch die Markov bzw. Pauli Master Gleichung definiert [41]. Sie resultiert direkt aus stochastischen Markov Prozessen, also solchen, deren Zeitentwicklung nur vom unmittelbar vorangehenden Zustand abhängig ist und - die vereinfacht gesagt - Prozesse ohne Gedächtnis sind. Die Master Gleichung hat sich wegen ihrer Universalität zu einem effektiven Werkzeug in der statistischen Physik und zur Beschreibung komplexer Systeme etabliert [42]. Sie ergibt sich aus der Überlegung, daß die Änderung der Wahrscheinlichkeit eines Zustands k von den Wahrscheinlichkeitszu- und abflüssen damit verbundener Zustände k' bestimmt wird. Die Flüsse ergeben sich aus den Summen der Produkte aus Wahrscheinlichkeit im Zustand k' zu sein, $P_{k'}(t)$, und der Rate, mit der sich dieser Zustand von k' zu k ändert, $w_{k,k'}$. Für den Zufluß ergibt sich somit $\sum_{k'} w_{k,k'} P_{k'}(t)$ und für den Abfluß $P_k(t) \sum_{k'} w_{k',k}$, wobei angenommen wird, daß die Raten explizit nicht zeitabhängig sind. Zusammen ergibt sich so die Master Gleichung

$$\dot{P}_k(t) = \sum_{k'} w_{k,k'} P_{k'}(t) - P_k(t) \sum_{k'} w_{k',k}. \quad (2.5)$$

Da nur solche Übergangsraten in Gl. (2.3) und Gl. (2.4) definiert wurden, die nur benachbarte Übergänge zulassen, reduziert sich der Wahrscheinlichkeitsfluß auf eine lineare Kette, siehe Abb. 2.2. Dies vereinfacht die Benutzung von Gl. (2.5), so daß man für die Wahrscheinlichkeitsänderung

$$\dot{P}_k(t) = w_{k+1,k} P_{k+1}(t) + w_{k-1,k} P_{k-1}(t) - (w_{k,k+1} + w_{k,k-1}) P_k(t) \quad (2.6)$$

schreiben kann. Da die einzelnen Zustände gerade mit der Anzahl von X pro Parzelle korrespondieren, $0 \leq k \leq n$, ergeben sich folglich die Randbedingungen $P_{-1}(t) = P_{n+1}(t) = 0$.

2.1.5 Stationäre Zustände

Im weiteren sollen stationären Zuständen untersucht werden, bei denen sich die $P_k(t)$ nicht mehr ändern, $\dot{P}_k(t) = 0$ und $P_k(t) = P_k$. Sie bestimmen die möglichen Gleichgewichtszustände des Systems, also solche, die in einer hinreichend langen Zeit erreicht werden. Mit Hilfe des durch Gl. (2.6) definierten gewöhnlichen Differentialgleichungssystem kann darüber hinaus auch (numerisch) der vollständige, zeitliche Verlauf untersucht werden.

Für den stationären Fall kann Gl. (2.6) nach den P_{k+1} aufgelöst werden,

$$P_{k+1} = \frac{(w_{k,k+1} + w_{k,k-1}) P_k - w_{k-1,k} P_{k-1}}{w_{k+1,k}}. \quad (2.7)$$

Für den Fall $k = 1$ erhält man wegen $P_{-1} = 0$ und $w_{0,-1} = 0$

$$P_1 = P_0 \frac{w_{0,1}}{w_{1,0}}. \quad (2.8)$$

Mit Hilfe der vollständigen Induktion kann gezeigt werden (siehe A.1), daß die allgemeine Lösung für das 1-Spezies System durch die Gleichung

$$P_k = P_0 \prod_{i=0}^{k-1} \frac{w_{i,i+1}}{w_{i+1,i}} \quad (2.9)$$

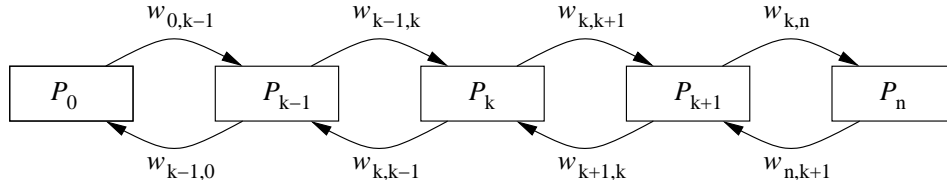


Abbildung 2.2: Veranschaulichung der kettenähnlichen Struktur der Wahrscheinlichkeitszu- und Abflüsse zu den $n + 1$ möglichen Zuständen.

bestimmt wird. Zu dieser Gleichung gelangt man auch direkt unter der Annahme, daß zu jedem Zustand k das detaillierte Gleichgewicht vorliegt, also das

$$P_{k-1}w_{k-1,k} = P_k w_{k,k-1} \quad (2.10)$$

gilt. Es soll angemerkt werden, daß das Prinzip des detailliertem Gleichgewichts in den meisten Fällen für Systeme vorausgesetzt werden kann, welche sich im thermischen Gleichgewicht befinden, bei denen also die Mikroreversibilität für die Erfüllung von Gl. (2.10) Sorge trägt. Daraus läßt sich vermutlich schließen, daß das 1-Spezies System sich in einem solchen thermischen Gleichgewicht befindet. Allerdings wurde bisher auf die Einführung einer Temperatur im gegebenen Kontext verzichtet, so daß letzte Aussagen nur als Anmerkung verstanden werden dürfen.

2.1.6 Selbstkonsistenz

Die Wahrscheinlichkeit kein X auf einer Parzelle zu haben, P_0 , wird durch die Normalisierungsbedingung $\sum_{k=0}^n P_k = 1$ beziehungsweise durch die Gleichung

$$P_0 = 1 - \sum_{k=1}^n P_k \quad (2.11)$$

bestimmt. Setzt man nun die P_k aus Gl. (2.9) in Gl. (2.11) ein und löst nach P_0 auf, so erhält man

$$P_0 = \frac{1}{1 + \sum_{k=1}^n \prod_{i=0}^{k-1} \frac{w_{i,i+1}}{w_{i+1,i}}}. \quad (2.12)$$

Die Einführung von $a_k = \frac{w_{k+1,k}}{w_{k,k+1}} \prod_{i=0}^k \frac{w_{i,i+1}}{w_{i+1,i}}$ erlaubt die Summation von $k = 0$ womit Gl. (2.9) zu $P_k = P_0 a_k$ und Gl. (2.12) zu $P_0 = 1 / \sum_{k=0}^n a_k$ werden, was zusammengesetzt, die P_k als

$$P_k = \frac{a_k}{\sum_{k=0}^n a_k} \quad (2.13)$$

definieren. Damit sind die einzelnen Zustandswahrscheinlichkeiten P_k eindeutig definiert, da die a_k Produkte der Übergangsraten $w_{k,k'}$ und diese bis auf die Mittelwerte \bar{k} allein von den Parametern (Migrations- und Reaktionsraten) abhängig sind.

Doch wie können die \bar{k} berechnet werden? Dazu werden erneut die beiden Haupteigenschaften des Systems, die vollständige Symmetrie und die unendlich große Population benutzt. Daraus läßt sich ableiten, daß der Mittelwert \bar{k} , über alle Parzellen zu einem gegebenen Zeitpunkt gemittelt, gleich dem Erwartungswert κ auf einer beliebigen Parzelle ist. Somit gilt

$$\bar{k} = \kappa = \sum_{k=0}^n k P_k. \quad (2.14)$$

Die Forderung dieser Eigenschaft soll als Selbstkonsistenz bezeichnet werden. Sie spiegelt die wechselseitige Abhängigkeit der (stationären) Wahrscheinlichkeiten von dem Mittelwert und *vice versa* wider. In Systemen mit endlicher Parzellen Zahl kann sie als Meanfield Näherung angesehen werden. Sie ist für Systeme mit unendlich vielen Parzellen exakt. Die Kombination aus Gl. (2.13) und Gl. (2.14) führt zu

$$\kappa \sum_{k=0}^n a_k = \sum_{k=0}^n k a_k, \quad (2.15)$$

was die Basisgleichung des Simplexmodells für alle 1-Spezies Systeme darstellt. Gl. (2.15) definiert implizit alle Parameter, die bei der Definition der $w_{k,k\pm 1}$ benutzt werden.

Es soll betont werden, daß durch die Benutzung der Selbstkonsistenzgleichung das PRESS Modell - unabhängig von den gewählten Reaktionen - eine quadratische Abhängig in den Zustandwahrscheinlichkeiten erfährt und sich damit von den gewöhnlichen Master Gleichungen unterscheidet.

2.2 Numerische Simulation

Die Grundlagen der stochastischen Simulationen werden behandelt, die sowohl eine Verifizierung der analytischen Ergebnisse als auch eine Analyse von molekularen Reaktionen in beliebigen Raumdimensionen erlauben.

2.2.1 Gittertopologie

Die mit dem PRESS Modell hergeleiteten analytischen Ergebnisse können mit Hilfe numerischer Simulationen verifiziert werden. Darüber hinaus erlauben sie die Untersuchung in niederen, physikalischen Raumdimensionen ($d = 1, 2, 3$), die mit dem PRESS Konzept nicht zugänglich sind.

In den folgenden Kapiteln wird kurz auf die Grundlagen der stochastischen Simulation für seriell arbeitende Rechner eingegangen. Es wurde ein Ansatz gewählt, der es ermöglicht, unterschiedliche Reaktionen modular in einer festen Simulationsumgebung einzubetten, um so eine Vergleichbarkeit der Simulationsergebnisse auch für unterschiedliche Reaktionssysteme zu erlauben. Eine Abhandlung für parallele Computer wird in Kapitel 5 gegeben.

Die Hauptidee ist es, eine Population von N individuellen Sequenzen auf einem rechtwinkligen, toroidalen Gitter unterzubringen, wobei jeder Gitterpunkt gerade n Individuen beherbergen kann. So entspricht ein Gitterpunkt gerade einer äquivalenten Parzelle im PRESS Modell. Die Verbindungstopologie, also die Anzahl der direkten Nachbarn, wird durch die Dimensionalität, d , bestimmt, vergl. Abb. 2.3. Aufgrund der gewählten Orthogonalität ist die Anzahl der nächsten Nachbarn gerade $2d$. Um Randeffekte zu vermeiden - die in unendlichen Populationen nicht auftreten würden - werden toroidale Gittertopologien benutzt, also Gitter mit periodischen Randbedingungen. Ferner sind die benutzten Gitter symmetrisch, das heißt, daß die Ausdehnungen in den einzelnen Dimensionen, L , gleich ist. Im d -dimensionalen Fall ergibt sich somit eine Gesamtpopulation von $N = nL^d$ Sequenzen.

2.2.2 Simulation von Diffusion

Die gewünschte Topologie wird durch eine Verknüpfung der Gitterpunkte erreicht, die als benachbart gelten sollen. Über diese Verknüpfungswege findet ein möglicher Teilchenaustausch statt. Die Simulation von Teilchenmigrationen wird wie folgt simuliert: Zuerst wird ein Gitterpunkt aus den L^d möglichen zufällig bestimmt. Um nun im weiteren der gewünschten Topologie Rechnung zu tragen, muß anschließend aus den $2d$ möglichen Nachbarn, genau ein weiterer Gitterpunkt gewählt werden. Befindet sich der erste Gitterpunkt am Rand, so wird sich aufgrund der toroidalen Natur des Gitters mindestens ein möglicher Nachbar auf der gegenüberliegenden Seite befinden. Ist unter der gegebenen Bedingung ein Gitterpunktpaar ermittelt worden, werden aus diesen jeweils ein aus den n möglichen Teilchen zufällig ermittelt. Anschließend wird der Inhalt dieser zwei Stellen vertauscht, unabhängig von deren Werten. Das bedeutet, daß unter Umständen auch zwei Leerstellen miteinander vertauscht werden können, was die Zustände genauso wenig ändert, wie das Vertauschen zweier identischer Sequenzen. Die Leerstelle nimmt keine besondere Position ein, was zum einen die Einhaltung der Maximalpopulation, n , pro Gitterpunkt als auch die Gesamtteilchenzahlerhaltung garantiert. Als alternative Diffusionsmethode kann für drei Dimensionen auch die im Kapitel 5.2.1 behandelte parallele Diffusionsmethoden benutzt werden (auch für ansonsten serielle Bearbeitung), die wegen ihrer Benutzung von Rotationsoperatoren die Nachbarschaftskondition

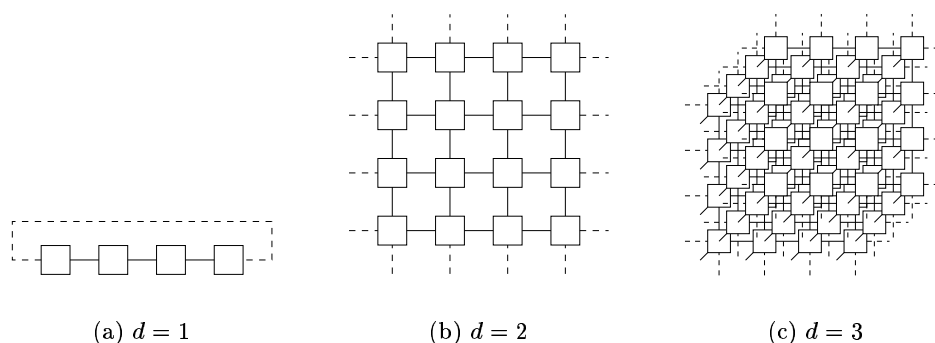


Abbildung 2.3: Der Zusammenhang zwischen Dimensionalität und Verbindungstopologie der Gitterpunkte - Anzahl nächster Nachbarn. Die gestrichelten Linien sollen, wie exemplarisch für $d = 1$ angedeutet, die Toroidalität darstellen.

berücksichtigt.

2.2.3 Simulation von Reaktionen

Die zweite Prozeßgruppe, die simuliert werden muß sind die möglichen Reaktion. Im Gegensatz zu den obigen Diffusionen finden diese isoliert auf einem Gitterpunkt statt, wie auch im PRESS Modell. Um die Simulation von Reaktionen durchzuführen, muß zuerst wieder ein Gitterpunkt aus den L^d möglichen zufällig ermittelt werden. Entsprechend der Natur der Reaktionen (uni-, bi- oder eventl. trimolekular), müssen für den gegebenen Gitterpunkt eine entsprechende Anzahl von Teilchen aus den n Positionen ermittelt werden. Diese werden dann in der ermittelten Reihenfolge gemäß der Reaktion abgearbeitet. Eine Permutation aller möglichen Kombinationen findet nicht statt. Dies reduziert zwar auf der einen Seite die Effizienz der Simulation, weil recht häufig keine Reaktion durchgeführt wird, trägt zum anderen jedoch die im PRESS Modell manifestierte besetzungsabhängige Wahrscheinlichkeit der Reaktionen Rechnung und erfüllt somit das chemische Massenwirkungsgesetz.

2.2.4 Simulationsprogramm

Damit die Simulation von Migration und unterschiedlichen Reaktionen mit den entsprechenden Raten durchgeführt wird, muß für jeden Simulationsschritt ermittelt werden, welcher Prozeß ausgeführt wird. Sei nun m die Migrationsrate und r_1 bzw. r_2 die Raten zweier unterschiedlicher Reaktionen. Nun können die Wahrscheinlichkeiten $p_m = \frac{m}{m+r_1+r_2}$, $p_{r_1} = \frac{m+r_1}{m+r_1+r_2}$ und $p_{r_2} = \frac{m+r_1+r_2}{m+r_1+r_2} = 1$ definiert werden. Mit Hilfe einer zu bestimmenden Zufallszahl z mit $0 \leq z < 1$ wird dann entschieden, welcher Prozeß simuliert werden soll, vergl. Abb. 2.4. Dieses Verfahren wird auch als Gillespie Algorithmus bezeichnet [43].

Die modulare Struktur erlaubt die unproblematische Einführung weiterer, möglicher Reaktionen mit entsprechenden Raten r_3, \dots, r_l und den damit verbundenen Wahrscheinlichkeiten p_{r_3}, \dots, p_{r_l} , die sich über die Beziehung $p_{r_j} = (m + \sum_{i=1}^j r_i) / (m + \sum_{i=1}^l r_i)$ herleiten.

Die Beschreibung des groben Simulationsprogramms verdeutlicht bereits die immense Verwendung von Zufallszahlen für alle angesprochenen Sektionen. Mit einer Reduzierung und der daraus resultierenden Ungenauigkeit von Zufallszahlen befaßt sich Abschnitt A.3

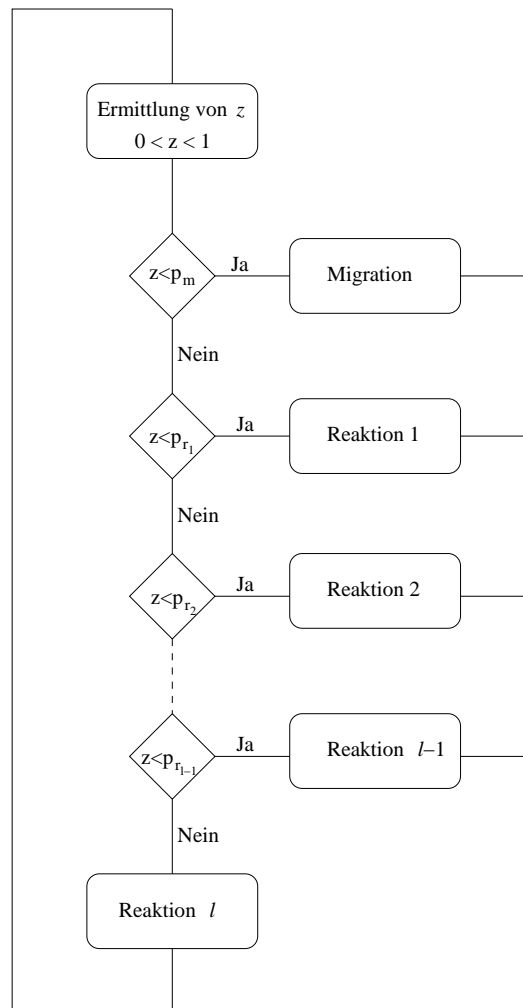


Abbildung 2.4: Das Schaubild verdeutlicht die Funktion des Simulationsalgorithmus (Gillespie Algorithmus) für den Fall zweier unterschiedlicher Reaktionen. Innerhalb der Schleife bestimmen Wahrscheinlichkeiten, proportional zu den entsprechenden Raten, den auszuführenden Simulationsteil. Die Module *Migration* und *Reaktion* sind in den entsprechenden Abschnitten beschrieben.

Kapitel 3

Selbstreplikatoren

3.1 Homogene Quasispezies

Das Quasispezies System als Modell für die Evolution wird in allgemeiner Form dargestellt und für weitere Untersuchungen vereinfacht.

3.1.1 Modell der Evolution

Das von Eigen 1971 veröffentlichte Modell der Quasispezies erlaubt Aussagen über das Verhalten von nicht wechselwirkenden, selbstreplizierenden und informationsspeichernden Sequenzen in einer homogenen Umgebung unter dem Druck von Mutation und Selektion zu treffen [3]. Es gibt erste Antworten auf die Fragen, inwiefern über einfache physiko-chemikalische Reaktionen Information verarbeitet werden kann.

Im folgenden werden die mathematischen Beziehung des Eigen Modells hergeleitet. Das Ergebnis ist die formale Darstellung der Lösung der Evolutionsgleichungen, die die Quasispezies darstellt. Um das Modell mit dem PRESS System untersuchen zu können, wird es an einigen Stellen vereinfacht, siehe Kap. 3.1.2.

Das Eigen Modell hat sich seit seiner Begründung zu einer wichtigen Theorie der molekularen Informationsverarbeitung entwickelt. Den strengen Bezug zur Physik konnte von Leuthäusser gezeigt werden, der das Quasispezies System exakt auf ein 2-dimensionales Ising Spinmodell abbildete [5, 44].

Bei der Entwicklung des Modells orientierte sich Eigen explizit an RNA/DNA-Sequenzen als Beispiel replizierender (Makro-) Moleküle. Die Struktur der Moleküle ist durch eine Aneinanderreihung von sogenannten Monomeren, den molekularen Einzelbausteinen, definiert. Im speziellen Fall von RNA-Molekülen kämen diese aus der Menge der vier unterschiedlichen Basen Guanin, Adenin, Cytosin und Uracil. Für eine gute Einführung in die Struktur und Manipulation von DNA bzw. RNA im Rahmen des DNA-Computings siehe [45] und allgemein [46]. Um keinen direkten Beschränkungen zu unterliegen, wird eine allgemeine Menge betrachtet, die aus λ Elementen besteht. Im obigen Fall der RNA-Moleküle wäre $\lambda = 4$. Der Prozeß der Selbstreplikation, der nicht ohne Energieaufwand und nur aus der Verwertung von Substraten einhergehen kann, ist in der Natur sehr komplex und wurde deswegen von Eigen auf wenige fundamentale Eigenschaften reduziert. So werden im Quasispezies Modell weder die angesprochenen Substrate, noch andere externe Ressourcen explizit modelliert.

Sei k_i die Anzahl von Molekülen mit der Sequenz i . Diese Moleküle bestehen aus ν Monomeren, wobei jedes Monomer aus der λ großen Sortenmenge stammt. Dies ergibt eine Variation von λ^ν unterschiedlichen Sequenzen, die in dem System vertreten sein können. Moleküle haben die Fähigkeit, sich mit einer gewissen Rate, A_i , zu replizieren. Diese sequenzspezifische Rate bestimmt, wie gut die Moleküle sich der Ressourcen bedienen können, und damit die Anpassung an ihre Umgebung. Je höher der Wert von A_i ist, desto schneller finden die Replikationen statt. Diese Abbildung von Sequenzen auf Replikationsraten wird als Fitneßlandschaft ([47, 16] (*NK-Modelle, Epistase*) und [48] für einen allgemeinen Überblick) bezeichnet. Im Gegensatz dazu gibt es aktuelle Arbeiten, die sich mit der Dynamik von explizit zeitabhängigen Fitneßlandschaften, $A_i = A_i(t)$, im Zusammenhang mit den Quasispezies beschäftigen [49, 50]. Da hier im weiteren jedoch der Einfluß räumlicher Korrelation im Fokus des Interesses steht, sollen die Replikationsraten A_i als konstant betrachtet werden. Bei einer Replikation kann es zu mutativen Fehlern kommen. Ausgehend von einer für alle Monomere gleichen Wahrscheinlichkeit q , bei einem Kopiervorgang korrekt repliziert zu werden, ergibt sich der Gesamtqualitätsfaktor $Q = q^\nu$, der die Wahrscheinlichkeit bestimmt, daß eine Replikation einer gesamten Sequenz fehlerfrei ist. Daraus läßt sich entsprechend auch die Fehlerwahrscheinlichkeit $R = 1 - Q$ definieren, die angibt, mit welcher Wahrscheinlichkeit eine Replikation fehlerhaft sein wird. Die Rate, mit der eine fehlerhafte Replikation von Sequenz i - wobei Sequenz j erzeugt wird - stattfindet, wird durch die Matrix $\mathbf{Q} = (Q_{ij})$ angegeben. Ihre explizite Darstellung wird unten hergeleitet. Auch sie soll als explizit nicht-zeitabhängig angenommen werden. Neben dieser prinzipiellen Erhöhung gibt es zwei weitere Prozesse, die die Anzahl von Molekülen erniedrigen können. Zum einen besteht die Möglichkeit, daß Moleküle zerfallen, was mit $-D_i k_i(t)$ beschrieben wird, wobei D_i als zeitlich konstant angesehen werden soll. Die ist auch der Unterschied zu der zweiten Möglichkeit, dem zeitlich abhängigen Ausfluß. Dieser mit einer Rate $-k_i(t)\Phi(t)$ stattfindende Prozeß kann als Verdünnung angesehen werden, so daß - wenn gewünscht - eine zeitlich konstante Molekülpopulation gewährleistet werden kann.

Die Zusammenfügung aller unterschiedlichen Terme ergibt für die Nettoproduktionsrate $\dot{k}_i(t)$

$$\dot{k}_i(t) = \left[A_i Q_{ii} - D_i \right] k_i(t) + \sum_{j \neq i} A_j Q_{ij} k_j(t) - k_i(t) \Phi(t). \quad (3.1)$$

Die Differenz von Replikations- und Zerfallsrate ergibt zusammen mit der Anzahl von Molekülen, $k_i(t)$, die Überschußproduktionsrate der Sequenz i . Die mittlere Überschußproduktionsrate $\bar{E}(t)$, ist somit

$$\bar{E}(t) = \sum_i k_i(t) \left[A_i - D_i \right]. \quad (3.2)$$

Da das System insofern als abgeschlossen betrachtet werden soll, daß keine neuen Moleküle von außen hinzugefügt werden können und sich die Produktion somit auf die „bekannten“ λ^ν Sequenzen reduziert, gilt die Erhaltungsbedingung

$$\sum_j Q_{ij} = 1, \quad (3.3)$$

was es erlaubt Gl. (3.2) als

$$\bar{E}(t) = \sum_i \dot{k}_i(t) + \Phi(t) \quad (3.4)$$

zu schreiben. Nun erkennt man, daß bei der Wahl $\Phi(t) = \bar{E}(t)$ die Gesamtpopulation $\sum_i k_i$ konstant gehalten werden kann.

Um weitere Rechnungen zu vereinfachen, soll o.B.d.A. $\Phi(t)$ gerade so gewählt werden. Ferner können die Prozesse der Replikation, der Mutation und des Zerfalls zu einer Matrix $\mathbf{W} = (W_{ij})$ zusammengefaßt werden. Unter Benutzung des Hamming Abstands, $d(i, j)$, der die Anzahl von Monomersubstitutionen angibt die notwendig sind, um von Sequenz i zur Sequenz j zu gelangen, ergibt sich

$$W_{ij} = A_j Q_{ij} - D_j \delta_{ij}, \quad (3.5)$$

mit

$$Q_{ij} = Q \left(\frac{q^{-1} - 1}{\lambda - 1} \right)^{d(i,j)}. \quad (3.6)$$

Durch die Transformation $x_i(t) = k_i(t) / \sum_j k_j(t)$, ist es möglich, relative Populationsvariablen, die es erlauben, Gl. (3.1) als

$$\dot{x}_i = \left[W_{ii} - \bar{E}(t) \right] x_i(t) + \sum_{j \neq i} W_{ij} x_j(t) \quad (3.7)$$

zu schreiben. Die $x_i(t)$ Abhängigkeit von $\bar{E}(t)$ macht Gl. (3.7) nicht-linear. Es ist jedoch möglich, diese exakt zu lösen [51]. Dabei wird der Term $\bar{E}(t)$ durch die Substitution

$$x'_i(t) = \exp \left(\int_0^t \bar{E}(\tau) d\tau \right) x_i(t) \quad (3.8)$$

eliminiert, was zu linearen Gleichungen der Form

$$\dot{x}'_i(t) = \sum_j W_{ij} x'_j(t) \quad (3.9)$$

führt. Da sich die $x_i(t)$ und die $x'_i(t)$ nur durch einen konstanten Faktor unterscheiden, ergibt sich die Rücktransformation einfach über die Beziehung

$$x_i(t) = \frac{x'_i(t)}{\sum_j x'_j(t)}. \quad (3.10)$$

Damit ist es möglich, daß Quasispezies Modell analytisch zu berechnen. Nach Gl. (3.6) ist \mathbf{Q} symmetrisch, also $Q_{ij} = Q_{ji}$. Unter der Annahme $A_i > 0$ und $D_i = 0$, $\forall i$ konnte Rumschitzki zeigen [52], daß unter Verwendung des Satzes von Frobenius-Perron [53] alle Eigenwerte der Matrix \mathbf{W} reell und positiv sind. Damit lassen sich mit den Eigenwerten $\lambda_0 \geq \lambda_1 \geq \lambda_2 \geq \dots$ und den dazugehörigen Eigenvektoren, $\phi^{(k)} = (\phi_i)^{(k)}$, die allgemeinen Lösungen von Gl. (3.9) als

$$x'_i(t) = \sum_k c_k \phi_i^{(k)} e^{\lambda_k t}, \quad (3.11)$$

darstellen, wobei die Koeffizienten c_k so gewählt werden, daß sie die Anfangsbedingung $x'_i(t = 0)$ erfüllen. Die Rücktransformation auf die Variablen $x_i(t)$ erfolgt gemäß Gl. (3.10)

$$x_i(t) = \frac{\sum_k c_k \phi_i^{(k)} e^{\lambda_k t}}{\sum_{i,k} \sum_k c_k \phi_i^{(k)} e^{\lambda_k t}}, \quad (3.12)$$

was nach Faktorisierung des größten Eigenwertes $e^{\lambda_0 t}$ die Beziehung

$$x_i(t) = \frac{c_0 \phi_i^{(0)} + \sum_{k>0} c_k \phi_i^{(k)} e^{(\lambda_k - \lambda_0)t}}{\sum_{i,k} c_k \phi_i^{(k)} e^{(\lambda_k - \lambda_0)t}} \quad (3.13)$$

ergibt.

Die Analyse von Gl. (3.13) läßt nun Rückschlüsse auf die Dynamik des Quasispezies Modells zu. Für $t \rightarrow \infty$ gehen die Werte $e^{(\lambda_k - \lambda_0)t}$ gegen Null, so daß nur die Komponenten des Eigenvektors $\phi^{(0)}$ zum dominanten Eigenwert λ_0 übrigbleiben. Das heißt zum einen, daß die Evolution des Quasispezies Modells asymptotisch in einem attraktiven Fixpunkt endet. Zum anderen können die Lösungen des stationären Zustands durch

$$x_i = x_i(t \rightarrow \infty) = \frac{\phi_i^{(0)}}{\sum_i \phi_i^{(0)}} \quad (3.14)$$

angegeben werden, was aber nichts anderes bedeutet, als daß sich als Ergebnis der Evolution in dem speziellen Modell der Quasispezies nicht die Sequenz mit der größten Replikationsrate, der Wildtyp - im Darwinschen Sinne der fitteste - sondern eine Wolke von Sequenzen durchsetzt. Diese Mischpopulation, die sich wie eine einzige Spezies verhält und sich nur für den artifiziellen Fall fehlerfreier Replikation auf den Wildtyp reduzieren läßt, wurde von Eigen als Quasispezies bezeichnet.

3.1.2 Vereinfachende Modellannahmen

Um nun im Weiteren prinzipielle Aussagen über die Evolution von Information zu treffen, reduziert man das Monomer Alphabet auf die binären Werte 0 und 1, was $\lambda = 2$ entspricht. Für die Untersuchung der homogenen Quasispezies wird der explizite Zerfall unterdrückt, $D_i = 0$. Des weiteren wird, um das spätere Interesse an dem Verhalten der Selbstreplikatoren in räumlichen Strukturen nicht unnötig zu erschweren, die einfachste, nichttriviale Fitneßlandschaft benutzt - die single-peaked Swetina-Schuster Landschaft [54]. Sie ordnet einer einzigen Sequenz $i = 0$ eine Replikationsrate $A_0 = \sigma$ (der „superiority Parameter“) zu und erscheint so im Vergleich zu den übrigen Sequenzen, die alle auf die Replikationsrate $A_{i \neq 0} = 1$ abgebildet werden, als Peak in einer Ebene. Um nun die immense Größe der Mutationsmatrix \mathbf{Q} zu reduzieren, werden die über den Hamming Abstand $d(i, j)$ definierbaren Fehlerklassen eingeführt. Dabei werden alle Moleküle zu der Fehlerklasse l zugeordnet, deren Sequenzen j den Hamming Abstand $d(0, j) = l$ zur Mastersequenz haben. Dies reduziert \mathbf{Q} von einer $(2^\nu \times 2^\nu)$ auf eine $(\nu \times \nu)$ Matrix und macht so eine Evaluierung zugänglich. Allerdings zieht dieses Verfahren auch die Limitierung mit sich, daß nicht jeder Sequenz ein separater Fitneßwert zugewiesen werden kann. Mit diesen Modellvoraussetzungen ist es möglich, die Konzentrationsverteilungen im stationären Zustand in Abhängigkeit von der Fehlerrate $R = 1 - Q$ zu berechnen.

Abb. 3.1 zeigt eine entsprechende Realisierung für den Fall $\sigma = 10$ und $\nu = 10$ (siehe Bildunterschrift für weitere Informationen). Beginnend bei $q = 1.0$, bei der nur die Mastersequenz im System vertreten ist, fällt deren Konzentration für fallende q stetig ab. Im gleichen Sinne beginnen die Konzentration der Sequenzen mit Hamming Abstand $d = 1, 2, \dots$ anzusteigen. Das Ensemble aller Sequenzen für ein gegebenes q definiert die Quasispezies. Auffallend ist eine kritische Fehlerrate $R_c = 1 - q_{\min}^\nu$, bei der die Korrelation der Quasispezies vollständig

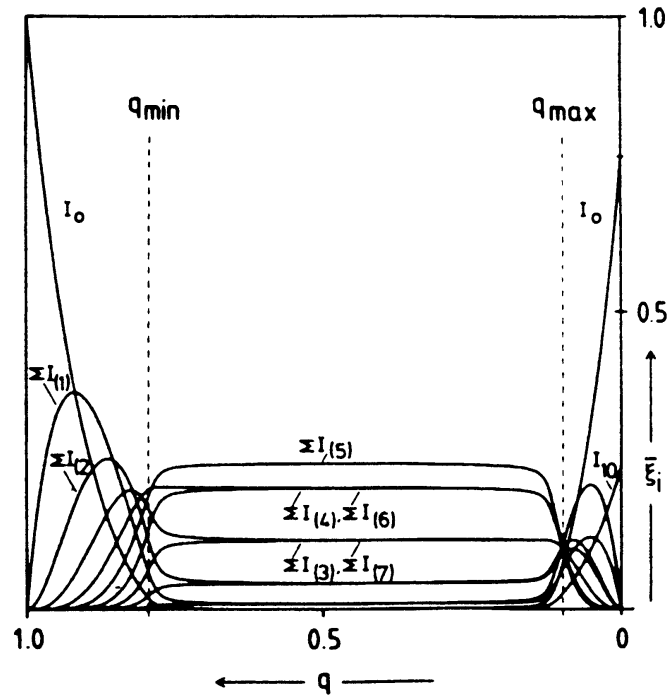


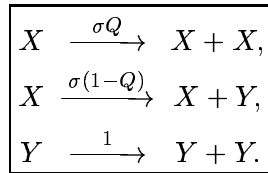
Abbildung 3.1: Die Abbildung zeigt (mit der Bezeichnung $x_i \rightarrow I_i$) die Konzentrationsverteilung, $\sum I_{(d)}$, der Fehlerklassen, d , in Abhängigkeit von der Monomer-mutationsrate q für Sequenzen der Länge $\nu = 10$. Man erkennt einen scharfen Übergang bei $q = q_{\min}$, zufällige Replikation zwischen $q_{\min} < q < q_{\max}$ und komplementäre Replikation für $q_{\max} < q$. (Die Abbildung stammt aus [10] mit freundlicher Genehmigung des John Wiley & Sons Verlag.)

zusammenbricht und zufällige Replikationen zu einer Gleichverteilung der Sequenzen führen. Da dies mit dem Verlust der Information einhergeht, wird der Übergang zur Gleichverteilung auch als Fehlerkatastrophe bezeichnet.

Für das weitere Vorgehen wird das Modell der Quasispezies noch um einen Punkt vereinfacht. Der Prozeß der Rückmutation zur Mastersequenz wird vernachlässigt, der durch den Vorgang, daß eine Sequenz $j \neq 0$ zu 0 mutiert, beschrieben wird. Da alle bis auf eine Sequenz gleichwertig behandelt werden, ist die Wahrscheinlichkeit dieses Prozesses sehr klein. Man kann in einer Näherung zeigen, vergl. Anhang A.2, daß die Wahrscheinlichkeit der Mutation einer Sequenz $j \neq 0$ aus einer gleichverteilten Population zu der Sequenz 0 bei kleinen Mutationswahrscheinlichkeiten proportional zu $\nu^{-1} \lambda^{-\nu}$ ist, was bei hinreichend langen Sequenzen beliebig klein wird und die Annahme rechtfertigt. Diese wird zudem noch durch die Tatsache begünstigt, daß in räumlich strukturierten Systemen, die im folgenden im Mittelpunkt des Interesses rücken werden, eine eventuelle Repopulation des Wildtyps in ein Gebiet vielmehr durch den Prozeß der Migration (bei physikalisch relevanten Diffusionsraten), als durch eine mögliche Neugeburt bestimmt würde.

Die speziellen Wahl der Fitneßlandschaft erlaubt eine Zerlegung des Sequenzraums in zwei Teile. Dabei wird die Konzentration $x_0(t)$ der Mastersequenz 0 fortan als $x(t)$ bezeichnet und die Summe aller übrigen Konzentrationen $\sum_{j \neq 0} x_j(t)$ als $y(t)$, siehe auch [55].

Die Prozesse der vereinfachten Quasispezies können wie folgt dargestellt werden:



Die deterministische Kinetik wird mit $x(t) = x$ und $y(t) = y$ durch

$$\dot{x} = \sigma Q x - x\Phi(t), \quad (3.15)$$

$$\dot{y} = \sigma(1 - Q)x + y - y\Phi(t), \quad (3.16)$$

beschrieben. Die kritische Fehlerschwelle R_c kann nun recht einfach berechnet werden. Dazu wird der stationäre Zustand $\dot{x} = \dot{y} = 0$ betrachtet und $\Phi(t)$ wieder so gewählt, daß $x + y = c = 1$ wird, woraus sich ein $\Phi(t) = \sigma x + y$ ergibt, was erneut der Rate der Gesamtproduktion entspricht. Nun kann Gl. (3.15) nach $y = \sigma(Q - x)$ aufgelöst und in Gl. (3.16) substituiert werden, was

$$x = \frac{\sigma Q - 1}{\sigma - 1}, \quad (3.17)$$

beziehungsweise

$$Q = \frac{1 + x(\sigma - 1)}{\sigma} \quad (3.18)$$

ergibt. Am Rande der Fehlerschwelle, bei der die Konzentration der Mastersequenz beliebig klein wird, $x \rightarrow 0$, ergibt sich für den Qualitätsfaktor Q der kritische Wert $Q_c = 1/\sigma$ bzw. der kritische Fehlerwert

$$R_c = 1 - 1/\sigma. \quad (3.19)$$

Die Benutzung des Limes $x \rightarrow 0$ als Definition der kritischen Mutationsrate kann nur in solchen Systemen vorgenommen werden, bei denen R eine stetige Funktion von x ist und x wiederum stetig gegen Null laufen kann. Ein System, bei dem dies nicht der Fall ist, wird in Abschnitt 4 von Interesse sein, wo die nicht-lineare Replikations- und die lineare Zerfallsdynamik gewisse Konzentrationsbereiche als nicht stabilisierbar deklarieren. In der vorliegenden reduzierten Form ist es möglich, das Verhalten der Selbstreplikatoren auf dem Simplex zu untersuchen.

3.2 Quasispezies ohne Zerfall im Press Modell

Der Einfluß des Raums auf Selbstreplikatoren wird mit Hilfe des Simplex Modells analytisch untersucht, wobei sich als Ergebnis eine Destabilisierung zeigt, die umso stärker wird, je größer des Einfluß des Raums wird.

3.2.1 Lokal konstante Populationen

Das Verhalten der Quasispezies, welches in nicht-räumlich aufgelösten Umgebungen bereits beschrieben wurde, wird im folgenden für den Fall einer Simplextopologie untersucht [56]. Es wird sich zeigen, daß für die Option einer lokal konstanten Population - ohne explizitem Zerfall - das Modell auf ein 1-Spezies System reduziert werden kann. Ausgehend von den Gl. (3.15) und (3.16), die das homogene Quasispezies Modell ohne Rückmutation und für die spezielle Wahl der Swetina-Schuster Fitneßlandschaft beschreiben, können die notwendigen Übergangsraten hergeleitet werden. Es wird angenommen, daß die explizit zeitabhängige Ausflußrate $\Phi(t)$ für jede einzelne Parzelle so eingestellt wird, daß die Anzahl von Molekülen zu jeder Zeit gerade der maximal möglichen Population n entspricht. Dabei wird eine beliebige, aber bereits vorhandene Sequenz von der Parzelle genommen, wenn eine neue als Folge einer Selbstreplikation entsteht. Diese Annahme einer lokalen Teilchenzahlregulierung entspricht dem Moran Modell [57]. Die resultierende Kopplung von Erzeugungs- und Zerfallsraten (durch das Entstehen eines Teilchens wird ein anderes vernichtet) kann ein als durch Ressourcen limitierter Prozeß verstanden werden. Obwohl diese Interpretation bei molekularen Systemen als etwas artifiziell erscheinen mag, zeigen sich im Vergleich zu der in Kapitel 3.3 behandelten Variante mit explizitem Zerfall kaum Unterschiede.

3.2.2 Das 1-Spezies Modell

Sei nun k die Anzahl der schnellreplizierenden Mastersequenzen, X , auf einer Parzelle, so ergibt sich die Zahl der langsamen Sequenzen, Y , eindeutig aus $n - k$. Somit reicht es aus, nur die X zu berechnen, was als ein 1-Spezies Modell angesehen werden kann und die Herleitungen und Gleichungen aus Kapitel 2.1 ihre Gültigkeit finden. Für den Fall unendlich großer Parzellen, $n \rightarrow \infty$, konvergiert $\frac{k}{n} \rightarrow x$ und $\frac{n-k}{n} \rightarrow y$, was die Beziehung zwischen den im homogenen Modell verwendeten Konzentrationen x, y und den im PRESS Modell benutzten (ganzzahligen) Werten $k, n - k$ aufzeigt. Entsprechend Gl. (3.15) und (3.16), ergibt sich somit für die Erhöhung der Anzahl von X (Erniedrigung der Anzahl von Y) auf einer beliebigen Parzelle gemäß Gl. (2.3) zu

$$w_{k,k+1} = \sigma Q \frac{k}{n} \cdot \frac{n-k}{n-1} + m \frac{\bar{k}}{n} \cdot \frac{n-k}{n}, \quad (3.20)$$

wobei der erste Term bestimmt, wie sich ein X mit der Wahrscheinlichkeit k/n und der Rate σ ohne Fehler, also mit der Wahrscheinlichkeit Q , repliziert - was als Ergebnis ein weiteres X liefert - und dabei ein Y mit der Wahrscheinlichkeit $(n-k)/(n-1)$ überschreibt. Entsprechend gilt bei der Erniedrigung der Anzahl von X (Erhöhung der Anzahl von Y),

$$w_{k,k-1} = \sigma(1-Q) \frac{k}{n} \cdot \frac{k-1}{n-1} + \frac{n-k}{n} \cdot \frac{k}{n-1} + m \frac{n-\bar{k}}{n} \cdot \frac{k}{n}, \quad (3.21)$$

mit ähnlichen Überlegungen zur Herkunft der Terme wie bei Gl. (3.20). Die zusätzlichen Migrationsterme mit (Diffusions-) Migrationskoeffizient m wurde im Kapitel 2.1.3 bereits erörtert und konnten unverändert übernommen werden.

Für den minimalen Fall $n = 2$, ergeben sich im 1-Spezies Modell genau drei verschiedene Zustände, wie in Abb. 3.2 zu sehen. Nach einsetzen der Raten Gl. (3.20) und Gl. (3.21) in Gl. (2.15), $\kappa \sum_{k=0}^n a_k = \sum_{k=0}^n k a_k$, kann man diese auf die Gleichung

$$\frac{2\kappa(2(Q-1)\sigma + m(2 + \kappa(\sigma - 1) - 2Q\sigma))}{(2 + m(2 - \kappa))(m(2 - \kappa) + 2(1 - Q)\sigma)} = 0 \quad (3.22)$$

reduzieren. Gl. (3.22) definiert implizit alle Abhängigkeiten der einzelnen Größen im stationären Zustand.

3.2.3 Kritische Fehlerrate $R_c(m)$

Um das Verhalten der Quasispezies auf dem Simplex mit dem homogenen Fall zu vergleichen, ist der Qualitätsfaktor Q erneut von Interesse. Löst man Gl. (3.22) nach $Q = Q(\kappa, m)$ ergibt sich

$$Q(\kappa, m) = \frac{\sigma + m(1 + \kappa \frac{\sigma-1}{2})}{\sigma(1 + m)}. \quad (3.23)$$

Die Stetigkeit von Q in κ erlaubt - wie im homogenen Fall - den kritischen Wert $Q_c(m)$ an der Stelle der Fehlerkatastrophe zu berechnen, $\kappa \rightarrow 0$, an der alle X aus dem System verdrängt werden. So ergibt sich für $Q_c(m) = Q(\kappa \rightarrow 0, m)$

$$Q_c(m) = \frac{\sigma + m}{\sigma(1 + m)}, \quad (3.24)$$

und für den entsprechenden kritischen Fehlerwert $R_c(m) = 1 - Q_c(m)$

$$R_c(m) = \frac{m(\sigma - 1)}{\sigma(1 + m)}. \quad (3.25)$$

Die Beziehung zwischen Fehlerschwelle $R_c(m)$ und Migrationskoeffizient m , die sich aus Gl. (3.25) ergibt, ist in Abb. 3.3 für die beiden Replikationsraten $\sigma = 2$ und 10 zu sehen. Sie werden mit den Werten der stochastischen Simulation verglichen, und zeigen trotz einer endlichen Simulationspopulation sehr gute Übereinstimmungen. Die abgebildeten horizontalen Linien verdeutlichen die entsprechenden Werte für den homogenen Fall. Diese ergeben sich aus Gl. (3.25) für den Grenzfall unendlicher Diffusion, bei der jegliche räumliche Korrelation verloren geht. So ergibt sich $R_c(m \rightarrow \infty) = 1 - 1/\sigma = R_c$, in Übereinstimmung mit den Herleitungen für den homogenen Fall, vergl. Gl. (3.19).

In gleicher Weise kann mit höheren Werten von n vorgegangen werden. Die Auflösung nach $Q(\kappa, m)$ bzw. $R_c(\kappa, m)$ wird jedoch aufgrund der ansteigenden Komplexität der Gleichungen mit zunehmenden n schwieriger. So ergeben sich für den Fall $n = 3$ aufgrund der quadratischen Abhängigkeit bereits zwei mögliche Lösungen. Die Physikalische kann erneut nach $R_c(m)$ umgeformt werden, was

$$R_c(m) = \frac{\sqrt{(1+m)(1+m+4m^3+8m^2\sigma+4m\sigma^2)} - (1+3m+2m^2)}{2(1+m)\sigma} \quad (3.26)$$

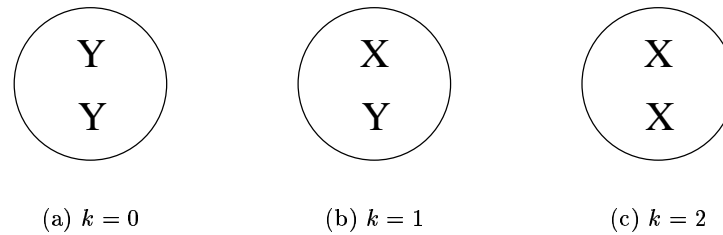


Abbildung 3.2: Die Unterschiedlichen Zustände pro Parzelle definiert über die Besetzungszahl k für das 1-Spezies Modell der Selbstreplikatoren mit Wildtyp, X , und übrigen Sequenzen, Y .

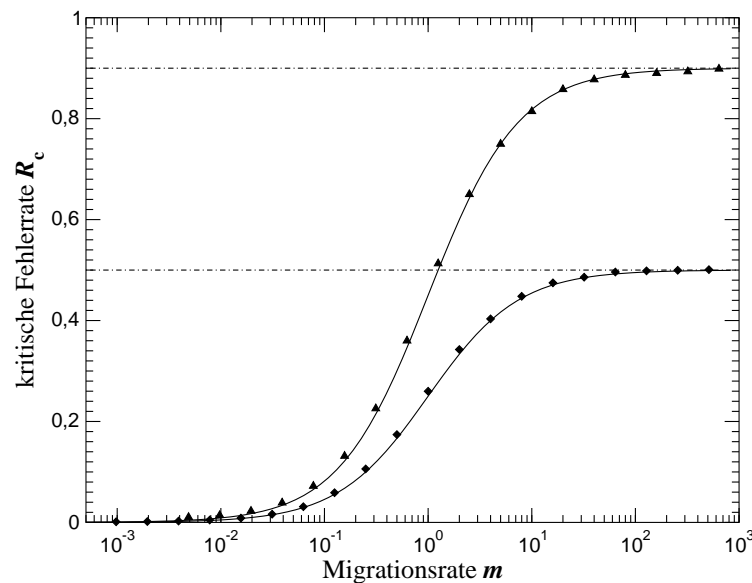


Abbildung 3.3: Die kritische Fehlerrate $R_c(m)$ als Funktion der Diffusionsrate m für zwei unterschiedliche „superiority“ Parameter $\sigma = 2$ und 10 . Die horizontalen Linien definieren die maximale Fehlerrate, $R_c = 1 - 1/\sigma$, identisch zum homogenen Quasispezies Fall. Die Symbole zeigen die entsprechenden Werte der stochastischen Simulation.

ergibt und die mit ansteigenden n zunehmende Komplexität verdeutlicht. Auch für $n = 3$ konvergiert Gl. (3.26) für $m \rightarrow \infty$ erneut gegen den Wert der homogenen Quasispezies, $R_c = 1 - 1/\sigma$. Die Darstellung der expliziten Lösung für $n = 4$ würde bereits einige Seiten benötigen, weshalb darauf verzichtet wird und statt dessen nur der entsprechende Plot, Abb. 3.4, gezeigt werden soll, der die kritische Fehlerrate $R_c(m)$ für $n = 2, 3$ und 4 vergleicht. Dieser zeigt, daß die fundamentalen Eigenschaften des Modells fast unabhängig von der Wahl der Parzellengröße ist und die destabilisierende Wirkung des Raums auf die Selbstreplikatoren mit zunehmenden n lediglich leicht abnimmt.

3.2.4 Skalierungsverhalten

Die schnell ansteigende Komplexität der Gleichungen für zunehmende n gestalten zwar eine analytische Lösung für hohe n schwierig, jedoch können diese problemlos numerisch gelöst werden. Trotzdem wäre eine approximative Beschreibung für das Verhalten von $R_c(m)$ für

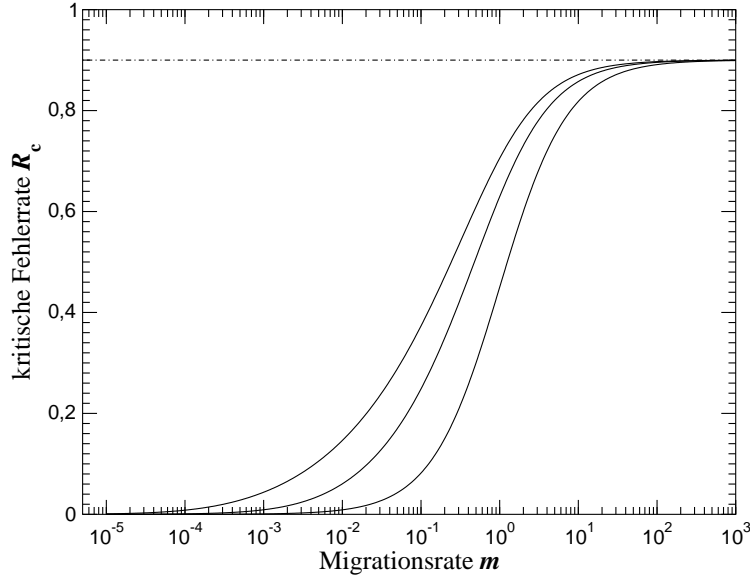


Abbildung 3.4: Die Abbildung zeigt die analytische Lösung von $R_c(m)$ für $n = 2, 3$ und 4 (von links nach rechts). Es zeigt sich eine Reduzierung der destabilisierenden Wirkung des Raumes für ansteigende Parzellengrößen, n .

unterschiedliche n wünschenswert. Dazu kann der Umstand genutzt werden, daß für die spezielle Topologie des Simplex ein Molekül, welches einmal eine Parzelle verlassen hat, auf diese - aufgrund der unendlichen Zahl von Parzellen - in endlicher Zeit nicht wiederkehren wird. Ein Molekül hat auf einer beliebigen Parzelle genau $n - 1$ Möglichkeiten sich per Überschreiben zu replizieren. Demnach hat es bei m Migrationsereignissen insgesamt $m(n - 1)$ Replikationsmöglichkeiten. Zwei unterschiedliche Werte für die Parzellengröße, n und n' können nun über eine Reskalierung der Diffusionsrate, m bzw. m' , so in Beziehung gesetzt werden, daß in einer Zeitspanne τ die Anzahl von möglichen Replikationen, $n_\tau = (1 + m\tau)(n - 1)$ gleich ist. Aus dieser Forderung, die einer Gleichsetzung einer effektiven Populationsgröße entspricht, ergibt sich

$$m' = \frac{n - 1}{n' - 1}m + \frac{n - n'}{n' - 1}1/\tau. \quad (3.27)$$

Die Zeit τ wird von der Größenordnung einer vollständigen Replikation sein und damit selbst wieder von m, n und n' abhängen. Ohne diese Abhängigkeiten im einzelnen zu bestimmen, kann jedoch für hinreichend große $m \gg 1/\tau$ Gl.(3.27) als

$$m' \approx \frac{n - 1}{n' - 1}m \quad (3.28)$$

approximiert werden. Ein Vergleich für Parzellengrößen n mit dem Referenzfall $n' = 2$ erhält man aus der Kombination von Gl. (3.25) und Gl. (3.28), was für die kritische Fehlerrate

$$R_c(n, m) \approx \frac{m(n - 1)(\sigma - 1)}{\sigma(1 + m(n - 1))} \quad (3.29)$$

ergibt.

Für den Fall einer unendlich hohen Population $n \rightarrow \infty$ pro Parzelle, was auch dem homogenen Fall entspricht, erreicht $R_c(n \rightarrow \infty, m)$ das korrekte Limit $1 - 1/\sigma$, wie bereits im

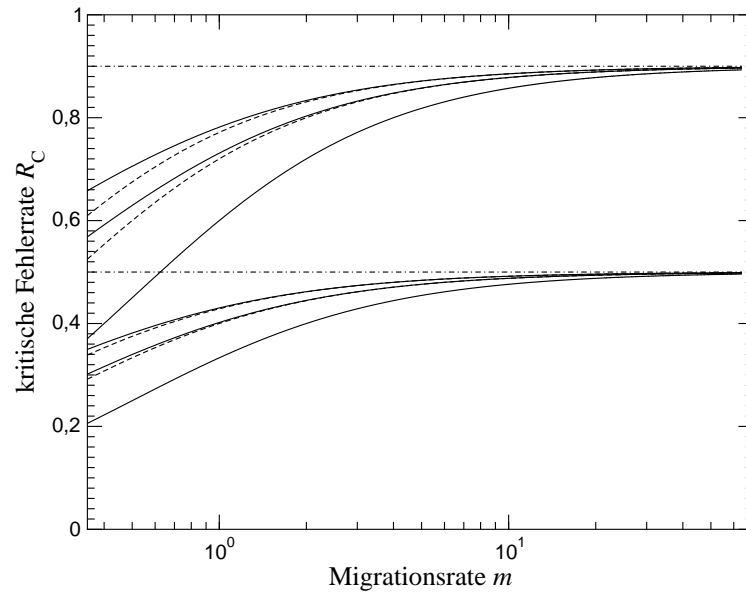


Abbildung 3.5: Vergleich der approximativen Skalierungsgleichung mit den exakten analytischen Ergebnissen der kritischen Fehlerrate $R_c(m)$ für $n = 2, 3, 4$ als Funktion moderater bis hoher Diffusion m .

Fall $m \rightarrow \infty$. Ein Vergleich der skalierten Werte mit den expliziten analytischen Ergebnissen für $n = 2, 3, 4$ wird für moderate und hohe Diffusionsraten in Abb. 3.5 gezeigt. Trotz der Näherung kann die approximative Skalierung zumindest für mittlere bis hohe Diffusionsraten das Verhalten für größere Parzellen beschreiben.

Um generelle Schlussfolgerungen zuzulassen, muß das Verhalten der Selbstreplikatoren in physikalischen Dimensionen untersucht werden. Um diesem Punkt Rechnung zu tragen, wurden entsprechende Simulationen auf ein-, zwei- und dreidimensionale Gitter ausgewertet und mit dem bekannten Ergebnissen verglichen, siehe Abb. 3.6. Man erkennt, daß die prinzipielle Abhängigkeit der kritischen Fehlerschwelle R_c von der Migrationsrate in allen Dimensionen identisch ist. Der einzige Unterschied wird durch eine Verschiebung der Fehlerschwelle gekennzeichnet, die sich aus der effektiv höheren Population von Systemen hoher Dimensionalität bei gleicher Diffusionsrate erklären läßt. Diese kommt einer Reskalierung der Migrationsrate gleich, ähnlich wie die bereits oben beschriebene Skalierung für unterschiedliche Teilchenzahlen.

3.2.5 Selbstreplikatoren im Raum

Für das einfache Beispiel der sich überschreibenden Selbstreplikatoren erweist sich das PRESS Modell als hervorragendes Werkzeug zur Beschreibung des räumlich aufgelösten Verhaltens der Quasispezies. Hervorzuheben ist, daß die Ergebnisse des PRESS Modells durch die Simulationen nicht nur bestätigt werden, sondern daß diese weiterhin zeigen, daß die Ergebnisse auch das Verhalten in physikalischen Dimensionen charakterisieren. Dabei bestätigt sich, daß der räumliche Einfluß eine Destabilisierung im Bezug auf die Erhaltung einer evolutiv ausgezeichneten Sequenz mit sich bringt und zwar umso mehr, je stärker dieser Einfluß (bei niedrigeren Migrationsraten) wird. Im Grenzfall unendlicher Diffusion, bei der die räumliche Struktur verloren geht, konvergiert das Modell gegen den Wert des homogenen Quasispezies Systems. Die entsprechenden Resultate zeigen eine streng monotone Abhängigkeit der kriti-

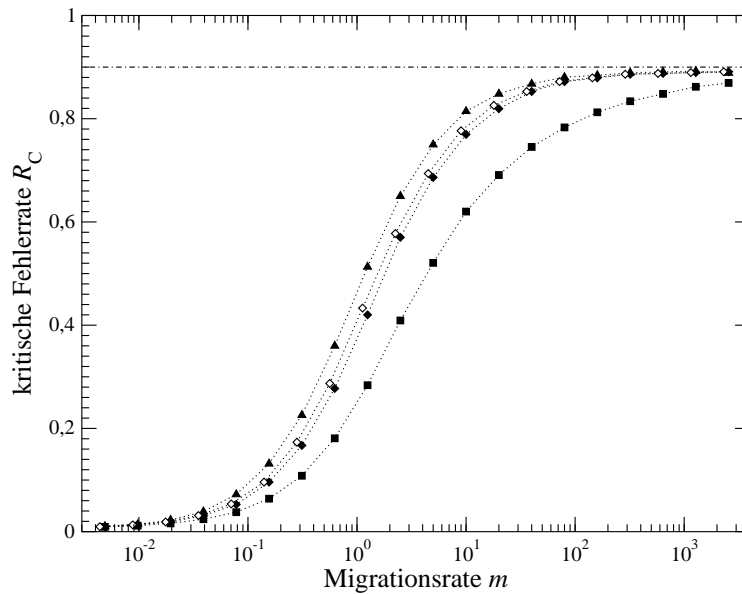


Abbildung 3.6: Vergleich der analytischen Ergebnisse des PRESS Modells für unendliche Dimension, $d = \infty$ (\blacktriangle), mit den Resultaten der numerischen Simulation für die physikalischen Raumdimensionen $d = 1$ (\blacksquare), $d = 2$ (\blacklozenge) und $d = 3$ (\blacklozenge) für den „superiority“ Parameter $\sigma = 10$.

schen Fehlerrate R_c von der Diffusionsrate m und damit der fundamental negative Einfluß des Raums auf die Stabilität von Selbstreplikatoren. Eine Destabilisierung, also die Reduktion von R_c wird bei den Quasispezies auch in homogenen Systemen mit endlicher (und geringer) Gesamtpopulation beobachtet [58, 59].

Allerdings ist nicht geklärt, in wieweit der räumliche Effekt das Resultat der lokalen Überschreibung ist, die Geburt und Zerfall streng miteinander koppelt. Um dies zu überprüfen, wird im nächsten Kapitel explizit der Zerfall modelliert, so daß sich Moleküle nur dann replizieren können, wenn eine entsprechende Leerstelle vorhanden ist. Dabei werden die Prozesse von Geburt und Zerfall wieder entkoppelt.

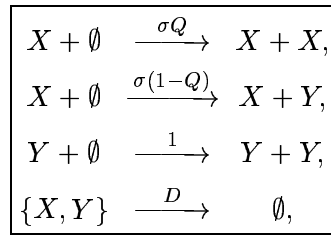
Um das erweiterte Quasispezies analytisch beschreiben zu können, wird das PRESS Modell auf ein multispezies System verallgemeinert, so daß es fortan zur Analyse beliebiger, räumlich aufgelöster, molekularbiologischer Reaktionen genutzt werden kann.

3.3 Quasispezies mit Zerfall im Press Modell

Das 1-Spezies Simplex Modell wird auf ein System erweitert, das prinzipiell die räumlich aufgelöste Dynamik beliebig vieler Spezies behandeln kann und das im Kontext der Quasispezies den durchweg negativen Einfluß des Raums auf die Stabilität der Selbstreplikatoren bestätigt.

3.3.1 Ratengleichungen mit Zerfall

In diesem Abschnitt wird die Annahme einer Überschreibung nicht weiter verfolgt und das Quasispezies System mit explizitem Zerfall modelliert. Allerdings muß zu diesem Zweck eine zweite Spezies aufgenommen werden. Im Kontext des Quasispezies Systems wird das bisher reduzierte PRESS Modell nun Schrittweise zu einem multispezies-Modell umgeformt. Die Grundmodellvoraussetzungen bleiben dabei erhalten und können prinzipiell wieder durch Gl. (3.15) und Gl. (3.16) beschrieben werden. Allerdings wird die explizit zeitabhängige Ausflußrate $\Phi(t)$ durch die konstanten und für alle Sequenzen gleichgewählte Zerfallsrate $D_i = D$ ersetzt. Eine Zusammenfassung der möglichen Reaktionen sieht wie folgt aus:



und führt schließlich zu den Gleichungen

$$\dot{x} = \sigma Q x(1 - x - y) - D x, \quad (3.30)$$

$$\dot{y} = \sigma(1 - Q)x(1 - x - y) + y(1 - x - y) - D y, \quad (3.31)$$

wobei die Geburts- und Zerfallprozesse lokal entkoppelt werden. Der Term $(1 - x - y)$ spiegelt die Tatsache wider, daß nur explizite Leerstellen per Replikationsprozeß okkupiert werden können. Insofern können die Leerstellen wie eine Ressource betrachtet werden, was die enge Verbindung zu den logistischen Gleichungen, $\dot{x} \propto x(1 - x)$, aufgezeigt. Die Benutzung einer konzentrationsunabhängigen Zerfallsrate kann unter bestimmten Voraussetzungen die vollständige Extinktion der gesamten Population mit sich führen.

Die Modellierung der kinetischen Gleichungen (3.30) und (3.31) auf die entsprechenden Raten geschieht prinzipiell wie schon in Kapitel 3.2.2. Mit den Bezeichnungen k_X und \bar{k}_X für die Anzahl von X pro Parzelle bzw. gemittelt für das Gesamtsystem und entsprechend für Y , k_Y bzw. \bar{k}_Y , ergeben sich die Änderung von k_X (bei konstanten k_Y) als

$$w_{k_X, k_X+1} = \sigma Q \frac{k_X}{n} \cdot \frac{n - k_X - k_Y}{n - 1} + m \frac{\bar{k}_X}{n} \cdot \frac{n - k_X - k_Y}{n}, \quad (3.32)$$

$$w_{k_X, k_X-1} = D \frac{k_X}{n} + m \frac{n - \bar{k}_X - \bar{k}_Y}{n} \cdot \frac{k_X}{n}, \quad (3.33)$$

und für die Änderung von k_Y (bei konstanten k_X) gemäß Gl. (3.31)

$$w_{k_Y, k_Y+1} = \sigma(1-Q) \frac{k_X}{n} \cdot \frac{n-k_X-k_Y}{n-1} + \frac{k_Y}{n} \cdot \frac{n-k_X-k_Y}{n-1} + m \frac{\bar{k}_Y}{n} \cdot \frac{n-k_X-k_Y}{n}, \quad (3.34)$$

$$w_{k_Y, k_Y-1} = D \frac{k_Y}{n} + m \frac{n-\bar{k}_X-\bar{k}_Y}{n} \cdot \frac{k_Y}{n}. \quad (3.35)$$

Die Raten sind erneut für infinitesimale Zeitabschnitte konstruiert, was die Möglichkeit von gleichzeitig stattfindenden Reaktionen ausschließt. Somit sind die einzigen Prozesse, die innerhalb des Zeitabschnitte die Zahl beider Spezies zu ändern vermag, Migrationen der Form

$$w_{k_X, k_X+1, k_Y, k_Y-1} = m \frac{\bar{k}_X}{n} \cdot \frac{k_Y}{n}, \quad (3.36)$$

$$w_{k_X, k_X-1, k_Y, k_Y+1} = m \frac{\bar{k}_Y}{n} \cdot \frac{k_Y}{n}. \quad (3.37)$$

3.3.2 Verallgemeinerung des Press Modells

Die Definition der Raten erlaubt erneut unter Benutzung der Master Gleichung die Berechnung der Belegungswahrscheinlichkeiten für X und Y . Allerdings muß beachtet werden, daß diese Wahrscheinlichkeiten nicht für eine, sondern für zwei (bzw. eine beliebige Zahl von) Spezies berechnet werden soll. Rechnet man das Quasispezies Modell nun mit explizitem Zerfall, bei dem die X - und Y -Moleküle separat betrachtet werden, korrespondieren zu einer maximalen Teilchenzahl von $n = 2$ pro Parzelle gerade 6 verschiedene Zustände, wie in Tabelle 3.1 zu sehen. Entsprechend Abb. 3.7 ergibt sich im zwei-Spezies Modell die Anzahl unterschiedlicher Zustände aus $N_Z = \frac{(n+1)(n+2)}{2}$. Eine allgemeine Beziehung zwischen der Anzahl benutzter Spezies s , der maximalen Parzellenpopulation n und den dazugehörigen Zuständen N_Z ergibt sich aus folgender Überlegung: Die Anzahl N_Z ist äquivalent zu der Fragestellung, wieviele Möglichkeiten es gibt, 0 bis n ununterscheidbare Teilchen auf s Kästen zu verteilen, bzw. genau n ununterscheidbare Teilchen auf $s + 1$ Kästen zu verteilen (wobei der $s + 1$ Kasten gerade der Spezies Leerstelle entspricht). Somit ergibt sich die Beziehung [41]

$$N_Z = \binom{s+n}{n}. \quad (3.38)$$

Die tatsächliche Aufstellung und Numerierung der N_Z Zustände ist dabei jedoch völlig beliebig. Um der Verknüpfung der unterschiedlichen Spezies und den damit verbundenen Wahrscheinlichkeitsflüssen Rechnung zu tragen, wird die Master Gleichung in Matrizenform umgeschrieben.

Da ein s Spezies System mit einer Population n pro Parzelle durch die N_Z Zustände, vergl. Tabelle 3.1, eindeutig charakterisiert wird, kann ein zu diesen Zuständen korrespondierender Wahrscheinlichkeitsvektor $\mathbf{P}(t)$ eingeführt werden. Sei $P(k_X, k_Y) = P(k_X, k_Y)(t)$ die Wahrscheinlichkeit zum Zeitpunkt t genau k_X Moleküle der Sorte X und k_Y der Sorte Y auf einer beliebigen Parzelle zu haben, so ergibt sich der Vektor $\mathbf{P}(t)$ für den speziellen Fall $s = 2$ und $n = 2$ gemäß Tabelle 3.1 zu

$$\mathbf{P}(t) = \left(P(0,0), P(1,0), P(2,0), P(0,1), P(1,1), P(0,2) \right)^T. \quad (3.39)$$

3.3. QUASISPEZIES MIT ZERFALL IM PRESS MODELL

Zustand	k_X	k_Y	$n - k_X - k_Y$
1	0	0	2
2	1	0	1
3	2	0	0
4	0	1	1
5	1	1	0
6	0	2	0

Tabelle 3.1: Darstellung der einzelnen Zustände im zwei Spezies Modell der Quasispezies mit explizitem Zerfall. Von links nach rechts sind die Zustandsnummer (Vektorkomponente), die Anzahl von X , die Anzahl von Y und die Gesamtzahl von Molekülen pro Parzelle angegeben.

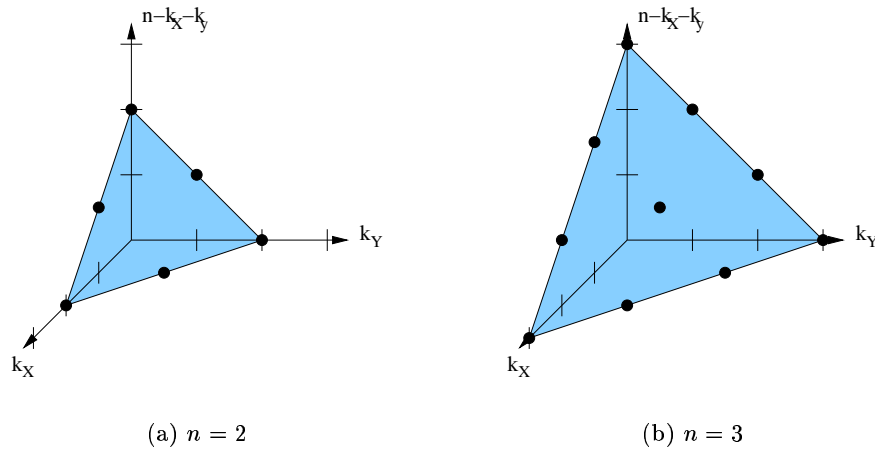


Abbildung 3.7: Die Anzahl unterschiedlichen Zustände pro Parzelle definiert symbolisiert durch \bullet in Abhängigkeit von der Parzellengröße n . Die Achsen bezeichnen die Besetzungsvariablen k_X , k_Y und die Anzahl resultierender Leerstellen $n - k_X - k_Y$. Für den Fall $n = 2$ ergeben sich 6 und für $n = 3$ gerade 10 unterschiedliche Zustände.

Die Zuordnung der Übergangsraten $w_{k,k'}$ zu den entsprechenden Wahrscheinlichkeiten wird mit Hilfe einer Matrix $\mathbf{L} = (L_{ij})$ erreicht, welche als

$$L_{ij} = L_{i,j} = w_{i,j} - \delta_{i,k} \sum_k w_{k,j}, \quad (3.40)$$

geschrieben werden kann. Die Übergangsraten $w_{k,k'}$, die sich auf den Zustandsraum beziehen ($k, k' \in \{0, \dots, N_Z\}$), ergeben sich aus Kompositionen der für den Fall der Quasispezies durch Gl. (3.32)-(3.37) gegebenen Raten, die als Index die tatsächlichen Populationsvariablen tragen (z.B. k_X, k_Y). In Anhang A.5 wird gezeigt, wie die Transformation für jede Komponente der Matrix \mathbf{L} mit Hilfe der Zustandsnummerierung manifestiert werden kann. Eine explizite Darstellung dieser Matrix ist schließlich für den gegebenen Fall der Quasispezies im Anhang A.6 gegeben. Ist \mathbf{L} definiert, ergibt sich entsprechend Gl. (2.5) für die Master Gleichung die Beziehung

$$\dot{\mathbf{P}}(t) = \mathbf{L}\mathbf{P}(t). \quad (3.41)$$

Im folgenden werden erneut die stationären Zustände behandelt, also solche bei denen $\dot{\mathbf{P}}(t) =$

0 und somit $\mathbf{P}(t) = \mathbf{P}$ gilt und das durch Gl. (3.41) definierte Gleichungssystem homogen wird.

Aus der Struktur der Matrix \mathbf{L} können ohne genauere Kenntnis der Komponenten bereits zwei wichtige Eigenschaften abgeleitet werden: Sie ist per Konstrukt gesamtwahrscheinlichkeitserhaltend und damit singular, siehe Abschnitt A.4. Die Singularität erlaubt prinzipiell eine Reduktion auf eine $((N_Z - 1) \times (N_Z - 1))$ Matrix. Allerdings wurde bisher die Forderung, das die Summe der Wahrscheinlichkeiten gleich 1 ist, also $\sum_i P_i(t) = 1$, noch nicht explizit formuliert. Sie kann elegant in das Gleichungssystem eingebaut werden, wie in Anhang A.7 gezeigt wird. Dabei ergibt sich eine Matrix $\mathbf{L}^{(\gamma)}$, die von gleicher Ordnung ist wie das ursprüngliche \mathbf{L} , und das Gleichungssystem

$$\delta_{i,\gamma} = \sum_j^{N_Z} L_{ij}^{(\gamma)} P_j. \quad (3.42)$$

bestimmt. Gl. (3.42) kann mit den Standardmethoden der linearen Algebra gelöst werden. Der Lösungsvektor \mathbf{P} enthält neben der Abhängigkeit von den Parametern σ, Q, m und D auch noch die gemittelten Größen \bar{x} und \bar{y} . Die letzten beiden können durch das Prinzip der Selbstkonsistenz eliminiert werden. Dabei wird, wie bereits im Kapitel 2.1.6 beschrieben, die Beziehung $k_X = \kappa$ und $k_Y = \zeta$ ausgenutzt, die sich aus der Geometrie des Simplex ergibt. Um die Erwartungswerte κ bzw. ζ mit \mathbf{P} zu verbinden, muß die durch Gl. (3.39) und Tabelle 3.1 definierte Transformation von k_X und k_Y auf k rückgängig gemacht werden. Dazu werden die beiden Vektoren \mathcal{X} und \mathcal{Y} definiert, deren k te Komponenten die entsprechenden Werte der k_X bzw. k_Y enthalten. Gemäß der in Tabelle 3.1 getroffenen Zuordnung erhält man

$$\mathcal{X} = \begin{pmatrix} 0, 1, 2, 0, 1, 0 \end{pmatrix}^T, \quad (3.43)$$

$$\mathcal{Y} = \begin{pmatrix} 0, 0, 0, 1, 1, 2 \end{pmatrix}^T. \quad (3.44)$$

Mit diesen ist es nun möglich, das Prinzip der Selbstkonsistenz für die beiden Molekülsorten X und Y elegant zu formulieren, wobei sich für κ und ζ

$$\bar{k}_X = \kappa = \sum_{i=1}^{N_Z} \mathcal{X}_i P_i, \quad (3.45)$$

$$\bar{k}_Y = \zeta = \sum_{i=1}^{N_Z} \mathcal{Y}_i P_i, \quad (3.46)$$

ergibt. Gl. (3.45) und Gl. (3.46) definieren implizit alle Abhängigkeiten der Parameter Q, m, D, σ und den Populationsgrößen κ und ζ und erlauben die Auflösung nach zwei unabhängigen Variablen. Allerdings können die beiden Gleichungen noch generell reduziert werden, was eine spätere Analyse und die explizite Auflösung nach einer der obigen Variablen vereinfacht. Dabei wird der Umstand ausgenutzt, daß die P_i als Lösung von Gl. (3.42) in der Form $P_i = p_i/q_i$ dargestellt werden können, wobei p_i und q_i Polynome der impliziten Parameter und der Populationsgrößen sind. Berechnet man den größten gemeinsamen Teiler - das Polynom q - so sind die P_i als $P_i = p'_i/q$ darstellbar. Aus der normierten Gesamtwahrscheinlichkeit ergibt sich außerdem die Beziehung $\sum_i p'_i = q$, so daß Gl. (3.45) und Gl. (3.46)

als

$$\sum_{i=1}^{N_Z} (\mathcal{X}_i - \kappa) p'_i = 0, \quad (3.47)$$

$$\sum_{i=1}^{N_Z} (\mathcal{Y}_i - \zeta) p'_i = 0, \quad (3.48)$$

geschrieben werden können, welche nur noch die reduzierten p'_i enthalten. Eine Koeffizientenanalyse zeigt zudem, daß es aus praktischer Sicht im allgemeinen sinnvoller ist, die beiden Gleichungen

$$\sum_{i=1}^{N_Z} (\zeta \mathcal{X}_i - \kappa \mathcal{Y}_i) p'_i = 0, \quad (3.49)$$

$$\sum_{i=1}^{N_Z} (\mathcal{Y}_i - \zeta) p'_i = 0, \quad (3.50)$$

zu verwenden, die von nun an als Basisgleichungen für alle $s = 2$ Systeme benutzt werden können und die es ermöglichen, zumindest prinzipiell nach einer impliziten Größe aufzulösen und dabei eine weitere zu eliminieren, z.B. nach dem Qualitätsfaktor Q , der nicht mehr von der Populationsvariablen ζ abhängig ist. Es zeigt sich, daß gerade diese Aufgabe je nach Beschaffenheit der zu untersuchenden Reaktionen durchaus problematisch sein kann, da auch Polynome mit höherem Grad als 5 entstehen können, die nicht mehr explizit auflösbar sind. Eine frühe Faktorisierung der beiden Gleichungen (3.49) und (3.50) und die Abtrennung nichtphysikalischer Lösungen kann jedoch die Komplexität stark reduzieren und eine analytische Behandlung im Allgemeinen ermöglichen. Im vorliegenden Fall der Quasispezies sind die Selbstkonsistenzgleichungen jedoch linear in Q , was eine straight-forward Evaluierung erlaubt.

3.3.3 Kritische Migrationsraten

Wie schon in den vorherigen Kapiteln soll die Stabilität des Quasispezies Systems untersucht werden, wobei sich nun der Einfluß des expliziten Zerfalls zeigen wird. Die interessante Kenngröße ist dabei wieder der Qualitätsfaktor $Q(\kappa, m, D)$ und insbesondere die kritische Qualität, die wiederum durch die Forderung $\kappa \rightarrow 0$ definiert werden kann, $Q_c(m, D) = Q(\kappa \rightarrow 0, m, D)$.

Ausgehend von den Selbstkonsistenzgleichungen (3.49) und (3.50) ergibt sich nach Umformungen die Beziehung

$$Q_c(m, D) = \frac{(2(D + m) + \sigma)((1 + D)(D + m) + |D^2 + Dm - m|)}{(D + 2D^2 + 6Dm + 2m(1 + 2m))\sigma}, \quad (3.51)$$

was die kritische Gesamtmutationsrate definiert, an der die schnellen Selbstreplikatoren X aufgrund der hohen Replikationsfehlern nicht mehr stabilisierbar sind und wegen des Zerfalls und der fehlenden Rückmutation vollständig aus dem System entfernt werden. Auffallend ist die Präsenz eines Betragsterms, $|D^2 + Dm - m|$. Er hebt einen von der Zerfallsrate D abhängigen Migrationswert m_2 hervor, bei dem $D^2 + Dm_2 - m_2 = 0$ wird,

$$m_2 = \frac{D^2}{1 - D}. \quad (3.52)$$

Für Migrationsraten kleiner als m_2 , $m < m_2$, ist der Term $D^2 + Dm - m$ positiv, so daß sich Gl. (3.51) mit $Q_c^{(1)}(m, D) = Q_c(m < m_2, D)$ zu

$$Q_c^{(1)}(m, D) = \frac{D(2(D+m) + \sigma)}{(D+2m)\sigma} \quad (3.53)$$

vereinfacht. Entsprechendes gilt für $m > m_2$. Hier ergibt sich mit $Q_c^{(2)}(m, D) = Q_c(m > m_2, D)$ die Vereinfachung

$$Q_c^{(2)}(m, D) = \frac{2(D+m) + \sigma}{\sigma(1+2(D+m))}. \quad (3.54)$$

Die kritischen Fehlerraten $R_c^{(1)}(m, D)$ und $R_c^{(2)}(m, D)$ sind über $R_c^{(i)}(m, D) = 1 - Q_c^{(i)}(m, D)$ definiert,

$$R_c^{(1)}(m, D) = \frac{2(\sigma m - Dm - D^2)}{(D+2m)\sigma}, \quad (3.55)$$

$$R_c^{(2)}(m, D) = \frac{2(D+m)(\sigma-1)}{(1+2(D+m))\sigma}. \quad (3.56)$$

In Gl. (3.55) fällt der Term $\sigma m - Dm - D^2$ auf. Dieser definiert eine weitere kritische Migrationsrate m_1 , für die $\sigma m_1 - Dm_1 - D^2 = 0$ gilt, womit sich

$$m_1 = \frac{D^2}{\sigma - D} \quad (3.57)$$

ergibt. Für alle $m < m_1$ wird $R_c^{(1)}(m, D)$ negativ, was physikalisch nicht sinnvoll ist und den vollständigen Zerfall der Population in diesem Bereich bedeutet. In dem Fall ist die Bestimmung einer kritischen Fehlerrate $R_c(m < m_1, D)$ nicht möglich. Sie kann stetig auf Null gesetzt werden,

$$R_c^{(0)}(m, D) = 0, \quad (3.58)$$

wobei die Abkürzung $R_c^{(0)}(m, D) = 1 - Q_c^{(0)}(m < m_1, D)$ verwendet wurde. Die vollständige Interpretation der kritischen Migrationsraten wird im Kapitel 3.3.4 gegeben. Da $\sigma > 1$ ist gilt die Beziehung $m_1 < m_2$, so daß die kritischen Migrationsraten drei aufeinanderfolgende Bereiche definieren, wie in Abb. 3.8 zu sehen. Die kritische Migrationsrate m_2 , definiert in Gl. (3.57), ist nur für $D < 1$ - also bei einer Zerfallsrate die kleiner als die Replikationsrate der Y-Sequenzen ist - positiv und somit physikalisch sinnvoll. Für $D \rightarrow 1$ konvergiert $m_1 \rightarrow \infty$, so daß für Zerfallsraten nahe eins der Verlauf der kritischen Fehlerrate ausschließlich durch $R_c^{(1)}(m, D)$ bestimmt wird. Abb. 3.9 zeigt die Abhängigkeit $R_c(m, D)$ von der Migrationsrate für unterschiedliche D und belegt die Behauptung.

Am Migrationspunkt m_2 weist die Kurve der kritischen Fehlerrate einen signifikanten Knick auf. Um dies zu verdeutlichen, wurden in den Abbildungen zusätzlichen die analytischen Fortsetzungen miteingezeichnet (unterbrochene Kurven). Im Limes $m \rightarrow \infty$, also $m > m_2$, konvergiert Gl. (3.56) gegen den Wert der homogenen Quasispezies $R = 1 - 1/\sigma$, verdeutlicht durch die horizontale Linie. Für den Fall von Zerfallsraten $1 < D \leq \sigma$ bestimmt m_1 die unterste Migrationsrate, bei der das System bei einem gegebenen D existieren kann. Für $D \sim \sigma$ wird m_1 - ähnlich zu m_2 im obigen Fall - beliebig groß. Dies bedeutet, daß nur für sehr hohe Migrationsraten die Population stabilisiert werden kann (siehe zur Interpretation von m_1 auch Kap. 3.3.4). Übersteigt D den Wert von σ , zerfallen also die Sequenzen häufiger als

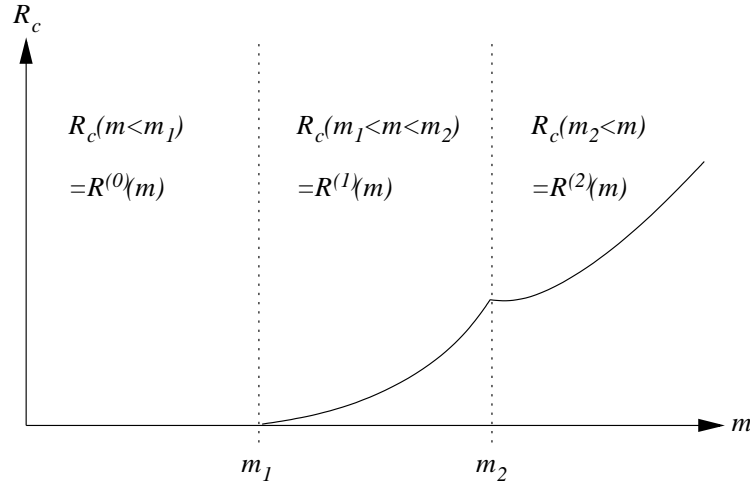


Abbildung 3.8: Einteilung der drei unterschiedlichen Bereiche, die durch die kritischen Migrationsraten m_1 und m_2 definiert werden. Zur Verdeutlichung ist zusätzlich schematisch die kritische Fehlerrate aufgetragen.

die Schnellsten sich replizieren können, kann keinerlei Information erhalten werden. Für $1 < D \leq \sigma$ kann die kritische Fehlerrate im Limes unendlicher Diffusion durch Gl. (3.51) bestimmt werden. Wie im homogenen Fall, siehe Kapitel A.8, ergibt sich für $Q_c(D) = Q_c(m \rightarrow \infty, D)$

$$Q_c(D) = \begin{cases} 1/\sigma & : D \leq 1 \\ D/\sigma & : D > 1 \end{cases} \quad (3.59)$$

Abb. 3.11 zeigt die analytischen Resultate für $D = 1$, $D = 2$ und $D = 5$. Um eine Bestätigung für die Korrektheit der gezeigten Ergebnisse zu erlangen, wurde mit Hilfe der in Kap. 2.2 besprochenen Simulationsmethode, die Analyse verifiziert. Abb. 3.12 zeigt den Vergleich der analytischen Ergebnisse mit den Daten der numerischen Simulation für den Fall $D = 0.1$ und $D = 0.5$. Es zeigt sich eine hervorragende Übereinstimmung, die sowohl die Vorhersage der beiden kritischen Migrationsraten m_1, m_2 als auch das Verhalten für sehr hohe Diffusionsraten bestätigt.

Schließlich stellt sich die Frage, ob die qualitative Übereinstimmung, die die beiden untersuchten Quasispezies Modelle im Zusammenhang mit der reduzierten Stabilität in räumlich aufgelösten Medien zeigen, auch quantitativ aufeinander abgebildet werden können. So sollte das Modell mit explizitem Zerfall im Limes $D \rightarrow 0$ auf das Überschreibungsmodell übergehen. Setzt man in Gl. (3.51) $D = 0$, ergibt sich

$$Q_c(m, D \rightarrow 0) = \frac{\sigma + 2m}{\sigma(1 + 2m)}. \quad (3.60)$$

Dies unterscheidet sich durch Gl. (3.24) nur durch den Faktor 2 vor den Migrationsraten m und ist erklärbar wie folgt: Für den Fall $D \rightarrow 0$ kann - wie beim Modell der Überschreibung - von vollständig belegten Parzellen ausgegangen werden. Im Fall $n = 2$ gibt es dann nur 2 Konfigurationen, bei denen eine Reaktion den Zustand einer Parzelle tatsächlich ändert, diese sind $\{X, X\}$ mit $X \rightarrow X+Y$ und $\{X, Y\}$ mit den Übergängen $X \rightarrow X+Y$ bzw. $Y \rightarrow Y+Y$. Ist bei der Überschreibungsvariante immer eine Reaktion möglich, so muß bei der Modellierung

des expliziten Zerfalls erst ein Molekül zersetzt werden um Platz für ein Reaktionsprodukt zu schaffen. Dabei wird im Zustand $\{X, X\}$ eines der X entfernt, womit die Konzentration und somit die Reaktionsrate auf die Hälfte reduziert wird. Im Fall $\{X, Y\}$ gilt ähnliches, wobei jeder der beiden Prozesse ebenfalls nur mit der halben Wahrscheinlichkeit stattfinden wird. Das bedeutet, daß die effektiven Replikationsraten im Vergleich zum Quasispezies Modell ohne Zerfall immer um die Hälfte reduziert sind, beziehungsweise, daß die Diffusion um den Faktor 2 erhöht ist.

3.3.4 Charakterisierung der m_1 und m_2 Übergänge

Die beiden kritischen Migrationsraten $m_1 = D^2/(\sigma - D)$ und $m_2 = D^2/(1 - D)$ beinhalten jeweils die Replikationsrate der X -Sequenzen, σ , und der Y -Sequenzen, 1. Dies läßt vermuten, daß das Verhalten der kritischen Fehlerrate an den gegebenen Migrationsstellen durch die Existenz bzw. nicht-Existenz der X - und Y -Sequenzen bestimmt wird. Aus diesem Grund sollen im folgenden die mittleren Populationsvariablen in Abhängigkeit der Diffusion untersucht werden. Ausgehend von den Selbstkonsistenzgleichungen (3.49) und Gl. (3.50) können $\kappa = \kappa(m)$ und $\zeta = \zeta(m)$ explizit berechnet werden.

An der Stelle m_2 weist die kritische Fehlerrate eine Knickstelle auf, so daß die erste Ableitung $\partial R_c(m, D)/\partial m$ an der Stelle $m = m_2$ unstetig ist. Die Frage nach der Ursache für dieses kritische Phänomen kann durch die Analyse von ζ beantwortet werden. Berechnet man die Y -Population, $\zeta(Q, m)$, an der Stelle m_2 , zeigt sich, daß diese für den Fall fehlerfreier Replikation, $Q = 1$, verschwindet. Dies wiederum bedeutet, daß ohne den Zufluß der schnellen jedoch fehlerhaft replizierenden X , die Replikationsrate der Y nicht ausreichend wäre, um gegen den Druck des Zerfalls bei geringen Migrationsraten zu bestehen. Da für $Q \leq Q_c$ die Konzentration der X per Definition gegen Null geht und somit der Zufluß zu der Y -Population versiegt, ist zu erwarten, daß für $m < m_2$ in dieser Stabilitätsregion die Y -Population aussterben sollte. Setzt man die kritische Fehlerrate $Q_c(m)$ - definiert durch Gl. (3.51) - in $\zeta(Q, m)$ ein, was sich zu

$$\zeta(Q_c, m) = \begin{cases} 0 & : m < m_2 \\ 2 - 2D - 2D^2/m & : m \geq m_2 \end{cases} \quad (3.61)$$

vereinfacht und entwickelt den Term $2 - 2D - 2D^2/m$ um die Stelle m_2 , so erhält man $\zeta(Q_c, m) = 2(1 - D)^2(m - m_2)/D$, also

$$\zeta(Q_c, m) \propto (m - m_2), \text{ für } m \gtrsim m_2, \quad (3.62)$$

was die Behauptung bestätigt. Es belegt ferner, daß die mittlere Anzahl der Y -Sequenzen in der Nähe der kritischen Migrationsrate $\gtrsim m_2$ und für $Q = Q_c$ proportional zu $(m - m_2)$ reduziert werden. Wird diese Schwelle unterschritten, werden die langsam replizierenden Moleküle aufgrund des Zerfalls und des fehlenden Mutationszufluß vollständig aus dem System entfernt. Für Migrationsraten, die deutlich größer als m_2 sind, konvergiert $\zeta(m \gg m_2) \rightarrow 2(1 - D)$. Diese Analyse erklärt den knickförmigen Verlauf der R_c Kurve an der Stelle $m = m_2$. In Anlehnung an die thermischen, kritischen Phänomene kann dieser Übergang auch als Phasenübergang zweiter Ordnung (allerdings im nicht-thermodynamischen Gleichgewicht) angesehen werden.

Für die folgende Interpretation der Migrationsrate m_1 ist es sinnvoll, daß Verhalten der Populationsgröße $\kappa(m)$ für Replikationsqualitäten $Q > Q_c$ zu studieren, da $\kappa(Q_c, m)$ definitionsgemäß Null für alle m ist und die Analyse erschwert. Dazu sollen die Nullstellen von

$\kappa(m', Q) = 0$ für beliebiges Q hergeleitet werden. Eine einfache Rechnung ergibt die einzige physikalische Nullstelle bei

$$m' = \frac{D(2D + \sigma(1 - Q))}{2(Q\sigma - D)}. \quad (3.63)$$

Die Diffusionsrate m' gibt in Abhängigkeit von der Fehlerrate Q die Stelle an, an der die Population des Wildtyps aussterben wird. Für $Q = 1$, also Replikationen ohne Fehler, wird Gl. (3.63) demnach die kleinste mögliche Migrationsrate definieren, bei der sich die Population gerade noch erhalten kann. In diesem Fall wird m' zu

$$m' = \frac{D^2}{\sigma - D} \quad (3.64)$$

also gerade zu der bekannten kritischen Migrationsrate m_1 . Es zeigt sich ferner, daß an der Stelle m_1 auch der Wert $\zeta(m)$ Null wird. Somit stellt m_1 eine auf dem Simplex universelle und von der Mutationsrate unabhängige Größe dar, die als unterste Migrationsschranke angesehen werden kann, bei der Selbstreplikatoren überleben können. Der Wert m_2 hingegen beschreibt die unterste Migrationsschranke, bei der beim Überschreiten der Fehlerschwelle die X -Population vollständig in die Y -Population übergeht, was der eigentlichen Definition der Fehlerschwelle entspricht. Unterhalb dieser Schranke, tritt beim Überschreiten dieser Grenze nur noch eine vollständige Extinktion ein. Abb. 3.10 zeigt schematisch die unterschiedlichen Bereiche.

3.3.5 Der Einfluß expliziten Zerfalls

Der Einfluß des expliziten Zerfall zeigt qualitativ keine Veränderung der kritischen Fehlerrate als Funktion der Migration für die nicht-wechselwirkenden Selbstreplikatoren. Der durchweg negative Einfluß auf die Stabilität der Information - gespeichert in den Wildtypsequenzen - bleibt auch in der biologisch relevanten Modellierung weiterhin bestehen. Allerdings ergeben sich mit der Einführung des Zerfalls einige Unterschiede zum vereinfachten Quasispezies System mit Überschreibung. Zum einen gibt es eine minimale Migrationsrate m_1 , bei der unabhängig jeglicher Mutation alle Sequenzen ausgelöscht werden und somit keinerlei Information gespeichert bzw. verarbeitet werden kann. Diese Rate definiert auch die Stelle, bei der die kritische Fehlerrate exakt Null wird, im Gegensatz zu dem Modell der Überschreibung, wo R_c auch für kleine Migrationsraten der x -Achse zwar beliebig nah kommt, aber dennoch ungleich Null bleibt. Zum anderen konnte eine weitere kritische Migrationsrate m_2 evaluiert werden, bei der die erste Ableitung von R_c in m unstetig wird. Es konnte gezeigt werden, daß dieses Phänomen mit dem Zerfall aller Y -Sequenzen in dem System für $m < m_2$ zu verstehen ist und die Knickstelle erklärt. Sie stellt im Hinblick auf die Definition der Fehlerschwelle die unterste Schranke dar, bei der beim Überschreiten dieser Schwelle die X -Population aufgrund der hohen Mutationswahrscheinlichkeit vollständig in die Y -Population übergeht. Für Migrationsraten kleiner als m_2 würde dieser Übergang zur vollständigen Extinktion des gesamten Systems führen.

Um die Verbindung zum vereinfachten Modell mit Überschreibung zu gewährleisten, konnte schließlich noch gezeigt werden, daß im Grenzwert $D \rightarrow 0$ beide Modelle auch quantitativ die gleichen Ergebnisse liefern (bis auf eine triviale Skalierung der Migrationsrate).

Die Resistenz von Selbstreplikatorpopulationen gegenüber Mutationen in homogenen Strukturen fällt deutlich in metrischen Räumen ab, offenbar fast unabhängig von der gegebene

nen Dimension. Auch eine moderate Vergrößerung der Parzellen - die eine leichte Erhöhung der Stabilität mit sich bringt - ändert nichts an der prinzipiellen Destabilisierung.

Schlußfolgernd kann aus den beiden Untersuchungen zum Quasispezies System die Erkenntnis gewonnen werden, daß sich der Effekt räumlicher Strukturen durchweg negativ auf die Informationserhaltung und -verarbeitung bezüglich einer ausgezeichneten Sequenz auswirkt. Die tolerierbaren Mutationen sind im Vergleich zu den bisher aus dem homogenen Modell bekannten Werten deutlich reduziert.

In den folgenden Kapiteln wird untersucht, wie sich wechselwirkende Systeme in räumlichen Umgebungen verhalten. Es wird sich herausstellen, daß die Einbettung in den Raum für diese Systeme in gewissen Bereichen eine stabilisierende Wirkung haben.

3.3. QUASISPEZIES MIT ZERFALL IM PRESS MODELL

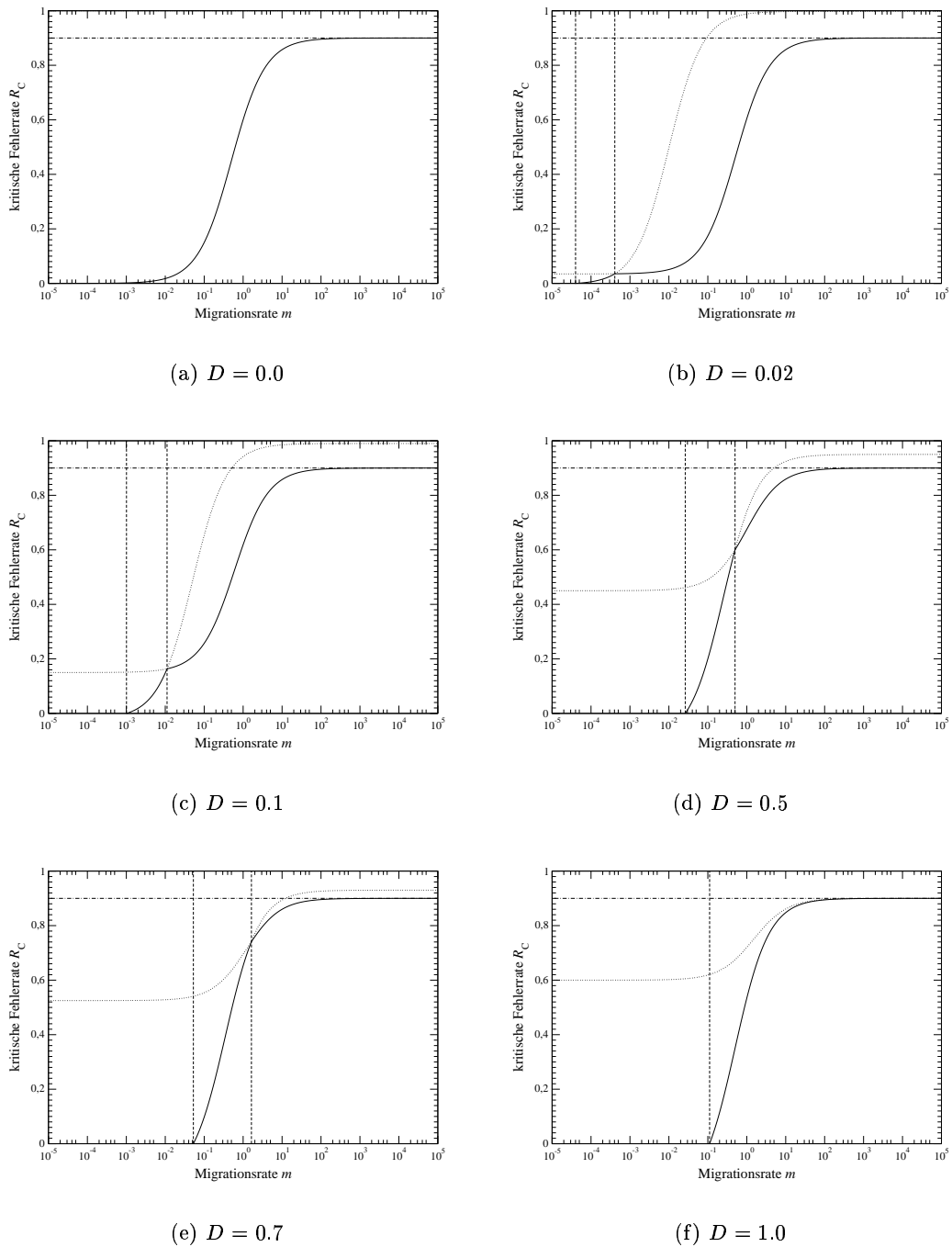


Abbildung 3.9: Die Graphen zeigen das Verhalten von $R_c(m)$ für unterschiedliche Zerfallsraten $0 \leq D \leq 1$. Zudem sind die kritischen Migrationsraten m_1 und m_2 (linke bzw. rechte vertikale Linie), die analytischen Fortsetzungen von $R_c(m)$ an der Stelle $m = m_2$ (unterbrochene Kurve) und der Limes der homogenen Quasispezies (horizontale Linie) eingezeichnet.

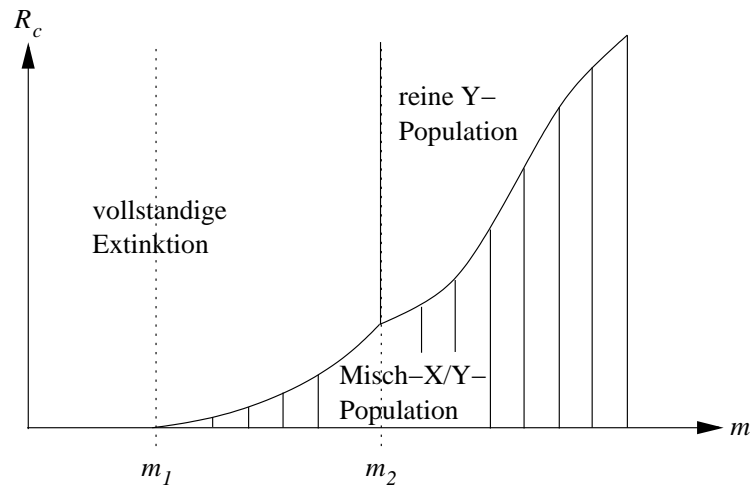


Abbildung 3.10: Die Graphik zeigt die Population in den drei durch m_1 und m_2 definierten Bereiche ober- und unterhalb der kritischen Fehlerschwelle. Unterhalb der Kurve (schraffierter Bereich) sind beide Populationen X und Y vorhanden. Oberhalb der Schwelle für $m > m_2$ nur die Y-Population stabilisiert werden. Alle übrigen Bereiche sind durch die vollständige Extinktion gekennzeichnet,

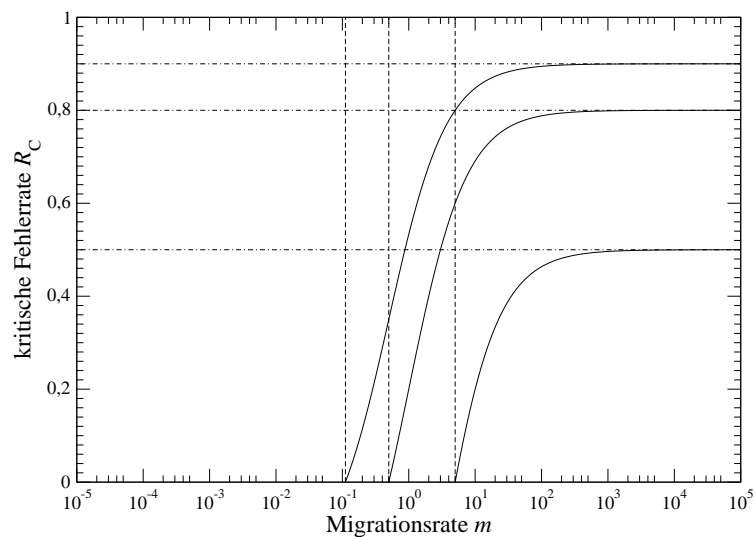
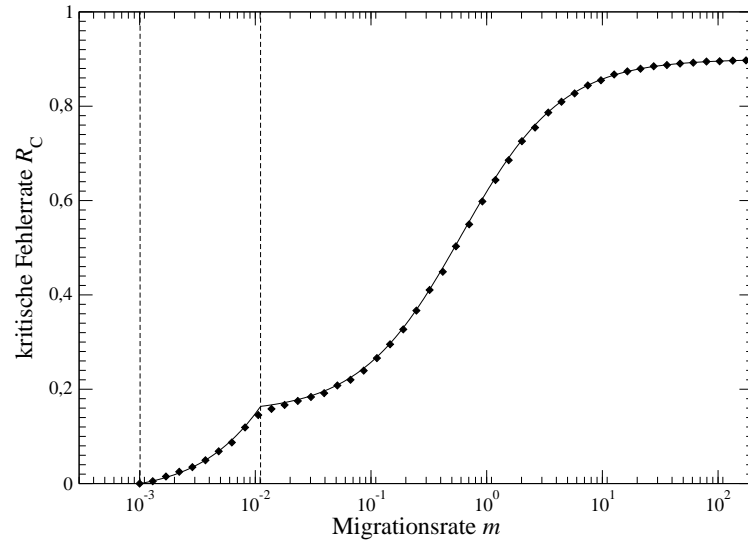
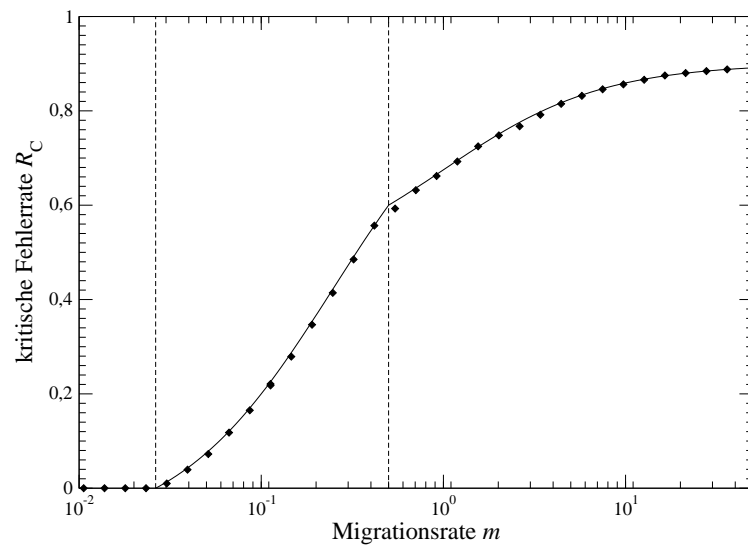


Abbildung 3.11: Die Graphen zeigen das Verhalten der kritischen Fehlerrate als Funktion der Migrationsrate für die hohen Zerfallsraten $D = 1, 2$ und 5 (von links nach rechts). Zusätzlich sind die jeweiligen Limite aus der homogenen Theorie (horizontale Linien) und die kritische Migrationsrate m_1 eingezeichnet.



(a) $D = 0.1$



(b) $D = 0.5$

Abbildung 3.12: Die Graphen vergleichen die numerischen Ergebnisse der Simulation und die der Analyse von $R_c(m)$ für $D = 0.1$ (a) und $D = 0.5$ (b), wobei sich sehr gute Übereinstimmung zeigt, die auch die Existenz der kritischen Migrationsraten (vertikale Linien) bestätigt.

Kapitel 4

Kooperierende Katalysatoren

4.1 Katalytisches Modell ohne Erkennung

Mit Hilfe des Simplex Ansatzes wird ein Modell von katalytischen Altruisten untersucht, das im homogenen Fall keinerlei Resistenz gegenüber Mutationen zeigt aber durch die Einbettung in den Raum stabilisiert werden kann.

4.1.1 Modell altruistischer Katalysatoren

Das bisher untersuchte Modell der Quasispezies reagierte auf die Einbettung in eine räumliche Umgebung mit einer deutlichen Reduktion der Fehlerschwelle, also mit einem Verlust an Stabilität gegenüber dem Informationsverlust durch Mutation. Der einzelne Selbstreplikator zeichnet sich vor allem dadurch aus, daß er kaum von seiner direkten Umgebung abhängig ist. Im Fall der Überschreibung kann er zwar von einem benachbarten Molekül eliminiert werden und im Modell mit Zerfall gilt es, die Ressource Leerstelle schneller als andere Moleküle zu okkupieren, jedoch bedingt die aktuelle Umgebung eines Moleküls nicht dessen Replikationseffektivität. Bei den Selbstreplikatoren gibt es keinerlei Möglichkeit der direkten Interaktion oder des Informationsaustauschs mit seiner Umgebung. Betrachtet man sich jedoch biologische Prozesse so stellt man fest, daß diese im allgemeinen wechselwirkende Systeme sind, vor allem, wenn man sich auf immer kleinere Größenskalen begibt, bei denen die Vereinfachung der Selbstreplikation nicht mehr möglich ist. Auch im Hinblick auf hoch komplexe und adaptive Systeme macht es keinen Sinn, die vollständige Funktionalität - die ein einzelner Selbstreplikator haben muß - in jedes Teilsystem zu integrieren, da der Kodierungsaufwand und der Overhead beliebig groß würden. Nur das Wechselspiel von Spezialisten und die Verteilung von Aufgaben auf unterschiedliche Subsysteme ermöglicht auf Dauer die notwendige, ressourcenschonende Effizienz und damit die Sicherung der Gesamtfunktionalität. Dabei muß, um eine universellen Anwendung zu sichern, die Distribution der Ressourcen selbstorganisiert gesteuert werden; also ohne übergeordnete Steuerungsmechanismen.

Wie zu Beginn dieser Arbeit gezeigt, unterliegen solche wechselwirkende Systeme immer dem Problem der Parasiten, deren globale Ausbreitung verhindert werden muß. Im Hinblick auf die ersten Momente der Evolution, als auch aus ganz allgemeiner Sicht, ist dabei vor allem das Verhalten des einfachsten katalytischen Systems interessant, das vollkommen ohne die Fähigkeit der Selbstreplikation auskommen und in einem möglichst allgemeinen Kontext stabilisierbar sein muß. Nur unter diesen Voraussetzungen kann das Minimalssystem als

Grundstufe für weitere evolutionäre Änderungen angesehen werden.

Im folgenden soll gezeigt werden, wie das einfachste aller denkbaren, wechselwirkenden Systeme, daß die Grundlage für alle komplexeren, interagierenden Modelle ist, allein durch die Einbettung in den Raum - also ohne die Notwendigkeit von globalen räumlichen Mustern - stabilisiert werden kann und sich damit vollkommen anders verhält, als es die homogene Prognose postuliert. Dabei bedarf es in der einfachen Form keinerlei - weder starken noch schwachen - Kopplungen wie im Hyperzyklus Modell oder den autokatalytischen Netzwerken [60]. Es ist ferner völlig frei von artifiziellen Annahmen, die eine Stabilisierung fördern würden und bleibt vollständig evolvierbar. Für weitere technische Details und Informationen siehe auch [61].

4.1.2 Reduzierung der Funktionen

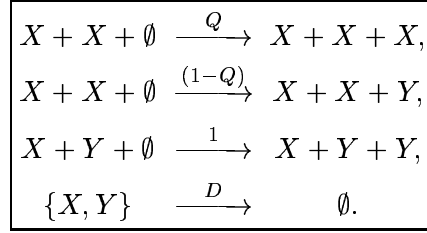
Die Vielzahl möglicher Funktionen, die in allgemeinen molekularbiologischen Systemen vorliegen können, wird auf die fundamental wichtige Operation reduziert, die für die Aufrechterhaltung eines biomolekularen Systems notwendig ist: Die Fähigkeit der Replikation. Dieses Funktional, das Replikase-Molekül, soll in Anlehnung an die vorherigen Kapitel wieder mit X bezeichnet werden. Es hat die Fähigkeit, andere Moleküle mit einer bestimmten Rate zu replizieren. Dabei soll es in der einfachsten Form keine Möglichkeit haben, die zu replizierenden Moleküle anhand ihrer Sequenzen zu erkennen. Somit wird jedes Molekül, welches mit dem Altruisten X wechselwirkt, vervielfältigt, völlig unabhängig von der Art des Moleküls. Dem gegenüber zerfallen die Moleküle mit einer gewissen Rate, wobei bei diesem Prozeß Platz für neue Sequenzen geschaffen wird. Es wird sich in den weiteren Analysen herausstellen, daß der explizite Zerfall für kooperativen Systeme deutlich wichtiger sein wird, als bei den Selbstreplikatoren.

Um die Vergleichbarkeit mit diesen zu gewährleisten und aus Gründen der Vereinfachung wird die Fähigkeit zur Replikation erneut durch die single-peaked Swetina-Schuster Landschaft beschrieben. In diesem Fall ist der Begriff einer Fitneßlandschaft irreführend, da das einzelne Molekül nicht direkt von seiner Fähigkeit andere zu replizieren profitieren kann. Im Gegensatz zum Quasispezies Modell diskriminiert die Landschaft die Sequenzen nicht in zwei Gruppen unterschiedlich ausgeprägter Funktionalität, sondern streng in funktionell und nicht-funktionell, also eine ausgewählte Sequenz - die Replikatorspezies X - im Pool völlig katalytisch inaktiver aber parasitärer Sequenzen - zusammenfassend wieder mit Y bezeichnet. Es gilt für die katalytische Replikationsrate $A_0 = 1$, sonst $A_{i \neq 0} = 0$. Der gesamte Sequenzraum soll erneut als so groß angesehen werden, daß Rückmutationen zur Mastersequenz problemlos vernachlässigt werden können.

Das System muß folglich die Aufgabe bewältigen, unter dem Druck von Mutationen und des Zerfalls, die katalytische Sequenz zu jedem Zeitpunkt zu stabilisieren. Dies kann nur dann gewährleistet werden, wenn die altruistischen, funktionellen Elemente kooperieren und sich somit einen - wenngleich minimalen - Vorteil gegenüber den parasitären und funktionslosen Sequenzen sichern.

Zusammenfassend können die Reaktionen des katalytischen Modells wie folgt dargestellt

werden:



Das führt zu den Differentialgleichungen

$$\dot{x} = Qx^2(1 - x - y) - Dx, \quad (4.1)$$

$$\dot{y} = (1 - Q)x^2(1 - x - y) + xy(1 - x - y) - Dy. \quad (4.2)$$

In Anhang A.9 wird gezeigt, daß das homogene Kooperationsmodell, also ohne räumliche Auflösung, für $Q < 1$ nicht stabilisierbar ist. Es wird sich zeigen, daß die alleinige Einführung räumlicher Struktur das Verhalten strikt ändert.

4.1.3 Zustände und Raten

Um den Aspekt der räumlichen Auflösung tatsächlich zu integrieren, wird das verallgemeinerte PRESS System benutzt, das in Kapitel 3.3.2 beschrieben wird. Wie schon in den vorherigen Kapiteln, müssen dazu die Gl. (4.1) und Gl. (4.2) in die entsprechenden Ratengleichungen transformiert werden, was für die Änderung von k_X (bei konstantem k_Y)

$$w_{k_X, k_X+1} = Q \frac{k_X}{n} \cdot \frac{k_X}{n-1} \cdot \frac{n - k_X - k_Y}{n-2} + m \frac{\bar{k}_X}{n} \cdot \frac{n - k_X - k_Y}{n}, \quad (4.3)$$

$$w_{k_X, k_X-1} = D \frac{k_X}{n} + m \frac{n - \bar{k}_X - \bar{k}_Y}{n} \cdot \frac{k_X}{n}, \quad (4.4)$$

und für die Änderung von k_Y (bei konstanten k_X) gemäß Gl. (3.31)

$$\begin{aligned}
 w_{k_Y, k_Y+1} &= (1 - Q) \frac{k_X}{n} \cdot \frac{k_X}{n-1} \cdot \frac{n - k_X - k_Y}{n-2} \\
 &\quad + \frac{k_X}{n} \cdot \frac{k_Y}{n-1} \cdot \frac{n - k_X - k_Y}{n-2} + m \frac{\bar{k}_Y}{n} \cdot \frac{n - k_X - k_Y}{n},
 \end{aligned} \quad (4.5)$$

$$w_{k_Y, k_Y-1} = D \frac{k_Y}{n} + m \frac{n - \bar{k}_X - \bar{k}_Y}{n} \cdot \frac{k_Y}{n} \quad (4.6)$$

ergibt. Und für die gleichzeitige Veränderung beider Populationsgrößen

$$w_{k_X, k_X+1, k_Y, k_Y-1} = m \frac{\bar{k}_X}{n} \cdot \frac{k_Y}{n}, \quad (4.7)$$

$$w_{k_X, k_X-1, k_Y, k_Y+1} = m \frac{\bar{k}_Y}{n} \cdot \frac{k_X}{n}. \quad (4.8)$$

wie schon in den beschriebenen Kapiteln bezeichnet k_X und k_Y wieder die Anzahl der X - und Y -Moleküle pro Parzelle und \bar{k}_X bzw. \bar{k}_Y die entsprechenden Zahlen gemittelt über das vollständige System. Die wechselwirkende, bi-molekulare Natur der Reaktionen bedingt eine kleinste Gesamtpopulation von $n = 3$ pro Parzelle. Mit den zwei Spezies X und Y ergibt sich

für $s = 2$ nach Gl. (3.38) genau $N_Z = 10$ unterschiedliche Zustände, die für einen speziellen Fall in Tabelle 4.1 zu sehen sind und den Vektor

$$\mathbf{P}(t) = \left(P(0, 0), P(1, 0), P(2, 0), P(3, 0), P(0, 1), \right. \\ \left. P(1, 1), P(2, 1), P(0, 2), P(1, 2), P(0, 3) \right)^T \quad (4.9)$$

definieren, wobei die einzelnen Komponenten des Vektors $\mathbf{P}(t)$ die Wahrscheinlichkeit $P(k_X, k_Y)$ angeben, genau k_X und k_Y Moleküle vom Typ X bzw. Y auf einer Parzelle zu finden. Mit diesen Zuordnungen ist es nun möglich, laut Anhang A.5 zuerst die Matrix \mathbf{L} (eine explizite Darstellung ist in Anhang A.6 gegeben) und darauf folgend gemäß Anhang A.10 die Matrix $\mathbf{L}^{(\gamma)}$, mit der für den stationären Fall die Gleichung

$$\delta_{i,\gamma} = \sum_j^{N_Z} L_{ij}^{(\gamma)} P_j \quad (3.42)$$

erneut zu lösen ist. Die P_i definieren über die Bedingung $P_i = \frac{p'_i}{q}$ erneut die reduzierten Komponenten p'_i mit denen die Selbstkonsistenzgleichungen als

$$\sum_{i=1}^{N_Z} (\zeta \mathcal{X}_i - \kappa \mathcal{Y}_i) p'_i = 0, \quad (3.49)$$

$$\sum_{i=1}^{N_Z} (\mathcal{Y}_i - \zeta) p'_i = 0, \quad (3.50)$$

geschrieben werden können und die fundamentalen Beziehungen darstellen, mit deren Hilfe nach den einzelnen Variablen aufgelöst werden kann. Die Vektoren \mathbf{x} und \mathbf{y} sind wieder für die Rücktransformation auf die einzelnen Zustände notwendig und lesen sich gemäß der in Tabelle 4.1 gemachten Konvention als

$$\mathbf{x} = \left(0, 1, 2, 3, 0, 1, 2, 0, 1, 0 \right)^T, \quad (4.10)$$

$$\mathbf{y} = \left(0, 0, 0, 0, 1, 1, 1, 2, 2, 3 \right)^T. \quad (4.11)$$

Es zeigt sich, daß es bei der Benutzung des PRESS Modells für das kooperative System formal keinen Unterschied zu den Quasispezies mit Zerfall gibt. Allerdings muß betont werden, daß die Komplexität der Gleichungen mit der Anzahl der Zustände N_Z deutlich zunimmt und sich die expliziten Lösungen der Selbstkonsistenzgleichungen entsprechend schwieriger gestalten.

4.1.4 Nichtlineare Dynamik

Bevor auf die Stabilisierung des Systems und damit auf die Behandlung der kritischen Fehler-raten eingegangen wird, soll auf einige Besonderheiten im Vergleich zum Quasispezies Modell aufmerksam gemacht werden. Diese Unterschiede sind die Folge der nicht-linearen Reaktions-kinetik. Sind die Erzeugungs- und Zerfallsraten im Quasispezies Modell beide linear in den

Zustand	k_X	k_Y	$n - k_X - k_Y$
1	0	0	3
2	1	0	2
3	2	0	1
4	3	0	0
5	0	1	2
6	1	1	1
7	2	1	0
8	0	2	1
9	1	2	0
10	0	3	0

Tabelle 4.1: Darstellung der einzelnen Zustände im zwei Spezies Modell des einfachen katalytischen Modells mit explizitem Zerfall. Von links nach rechts sind die Zustandsnummer (Vektorkomponente), die Anzahl von X , die Anzahl von Y und die Anzahl von Leerstellen pro Parzelle angegeben.

relativen Molekülpopulationen (abgesehen vom Term der Ressourcenbeschränkung), so wird die Dynamik im Fall der Wechselwirkung durch das Zusammenspiel von quadratischer (Erzeugung) und linearer (Zerfall) Abhängigkeit geprägt. Das hat zur Folge, daß kritische Konzentrationen vorhanden sein können, unterhalb denen keine Stabilisierung möglich ist. Anhand des homogenen Spezialfalles soll dies demonstriert werden: Sei die Erzeugung proportional zu x^2 und der Zerfall eine lineare Funktion, Dx , so könnte ein entsprechendes System nicht unterhalb der kritischen Konzentration $x_c = D$ stabilisiert werden, wie eine einfache Rechnung zeigt. Die Schlußfolgerung angewandt auf das vorliegende System bedeutet, daß die bisherige Definition $Q_c = Q(\kappa \rightarrow 0)$ nicht gültig ist. Jede Berechnung von Q_c muß demnach den Aspekt der Stabilität berücksichtigen, d.h. auch, wie sich die Lösungen des einfachen kooperativen Modells in der Gegenwart von Perturbationen auf die Besetzungswahrscheinlichkeiten und den damit verbundenen Populationsvariablen κ und ζ verhalten.

Um die Methoden der linearen Stabilitätsanalyse für die Klasse der PRESS Modelle anwenden zu können, muß die Dynamik bestimmende Matrix \mathbf{L} transformiert werden, da zur Bestimmung der Stabilität von Lösungen homogener Gleichungssystemen, also solche der Art

$$\dot{\mathbf{P}} = \mathbf{L}\mathbf{P}, \quad (4.12)$$

prinzipiell die Charakteristiken der Eigenwerten der $(N_Z \times N_Z)$ Matrix \mathbf{L} mit Hilfe der Determinanten zu berechnen sind, die Matrizen aber in den hier besprochenen Fällen singulär sind. Daher werden diese auf eine nicht-singuläre $((N_Z - 1) \times (N_Z - 1))$ Matrix reduziert, wobei die homogene Form des Gleichungssystems beibehalten wird.

Das Ergebnis dieser Matrixtransformation soll fortan mit \mathbf{R} bezeichnet werden. In Anhang A.10 wird explizit und allgemein gezeigt, wie dies für die wahrscheinlichkeitserhaltenden \mathbf{L} Matrizen funktioniert. Zu beachten ist, daß auch die entsprechenden Vektoren \mathbf{P} um eine Dimension reduziert werden. Die eliminierte Komponente P_j , die den Vakuumzustand repräsentiert, wird aus den restlichen Komponenten über die Beziehung $P_j = 1 - \sum_{i \neq j}^{N_Z} P_i$ berechnet. In der gemäß Tabelle 4.1 getroffenen Konvention ist $j = 1$.

Eine Einführung in die lineare Stabilitätsanalyse und die entsprechenden Umwandlungen auf die Variablen des PRESS Modells ist in Anhang A.11 gegeben. Wie bereits gesagt, erlaubt die Komplexität des Kooperationsmodells keine direkte analytische Nutzung dieser Ergebnis-

se. Allerdings kann die im Anhang A.10 hergeleitete Matrix \mathbf{R} , die die Zeitentwicklung der Belegungswahrscheinlichkeiten in homogener Form präsentiert,

$$\dot{\mathbf{P}} = \mathbf{R}\mathbf{P}, \quad (4.13)$$

benutzt werden, um die zeitabhängigen Lösungen $\mathbf{P}(t)$ bzw. die Eigenwerte numerisch zu berechnen. Dazu muß \mathbf{R} aber noch insofern umgeändert werden, daß alle Matrixterme, die explizit κ oder ζ enthalten mit Hilfe der Selbstkonsistenzgleichungen durch die entsprechende Linearkombination der Wahrscheinlichkeitskomponenten substituiert werden. Dies transformiert $\mathbf{R}(\kappa, \zeta)$ zu $\mathbf{R}(\mathbf{P})$ und erzeugt die quadratische Abhängigkeit der Gl. (4.13) von \mathbf{P} .

Wie die weiterführenden Berechnungen der Stabilitätsanalyse im einzelnen aussehen, wird ebenfalls in Anhang A.11 gezeigt.

4.1.5 Stabilisierung von Kooperation

Bevor auf eine Analyse der kritischen Fehlerrate eingegangen wird, sollen die Populationszahlen vorgestellt werden, die sich unter dem Druck von Mutation auf dem Simplex ergeben. Der linke Teil von Abbildung 4.1 zeigt die stationären Lösungen, für den Fall $Q = 0.999$ und $Q = 0.99$ und für die angegebenen Zerfallsraten.

Zu erkennen sind Populationszahlen ungleich Null der X - und somit auch Y -Moleküle bei $Q < 1$. Es fällt auf, daß dies jedoch immer nur für Migrationsbereiche möglich sind. Die Startwerte dieser Bereiche sollen mit m_a und die Endwerte mit m_b bezeichnet werden. Sie sind aufgrund der Komplexität der Gleichungen nicht mehr vollständig analytisch lösbar. Die speziellen Reaktionen des Kooperationsmodells erlauben keine Existenz der Y -Moleküle ohne die Gegenwart der katalytischen X -Sequenzen, so daß der Bereich, in dem $\zeta > 0$ ist, ebenfalls durch m_a und m_b definiert wird. Auffallend ist, daß die absolute Höhe der X -Population für die untersuchten Zerfallsraten nahezu konstant bleibt, bis schließlich bei $D \approx 10^{-2}$ ein deutlicher Einbruch stattfindet. Wie bereits erwähnt, stellen die beiden linken Graphen aus Abbildung 4.1 die stationären Lösungen dar, die ohne Beachtung der Stabilität berechnet wurden. Wird diese als weiteres Kriterium hinzugezogen - was unter Beachtung der Eigenwerte geschieht - so reduzieren sich die vorherigen Bereiche auf die der asymptotisch stabilen Zustände, die die rechten Graphen von Abb. 4.1 beschreiben. Die Zustände liegen nun zwischen m_a und m_c , wobei für alle Parameterwerte $m_a < m_c < m_b$ gilt. Sie sind dadurch gekennzeichnet, daß „kleine“ Störungen, welche das System aus ihrer Gleichgewichtslage bringen, im Laufe der Zeit kompensiert werden können, so daß für $t \rightarrow \infty$ der Gleichgewichtszustand wieder eingenommen wird. Wie die Zustände jenseits von m_c zu charakterisieren sind, wird in Kapitel 4.1.6 behandelt.

Die Stabilisierung der X -Population bedeutet, daß prinzipiell eine Population von nicht-perfekt kopierenden und völlig altruistisch handelnden Molekülen in einer parasitären Umgebung existieren können, ganz im Gegensatz zum homogenen Fall, bei dem selbst beliebig kleine Mutationsraten den totalen Informationsverlust bedeuten. Das heißt, daß die Anwesenheit des Raums es erlaubt, eine kooperatives Wechselspiel zwischen den (prinzipiell altruistischen) funktionalen Elementen zu etablieren, was diesen im zeitlichen Verlauf einen gewissen Vorteil zu den nicht-funktionalen, parasitären Molekülen zuspricht. Es stellt sich die Frage, wie dies im konkreten Fall zu erklären ist. Dazu ist die Kinetik auf den einzelnen Parzellen wichtig, die für $n = 3$ in drei unterschiedliche Fälle klassifiziert werden kann und die wie folgt aussehen:

$$(1) : \{X, X, \emptyset\}, \quad (2) : \{X, Y, \emptyset\}, \quad (3) : \{Y, Y, \emptyset\}. \quad (4.14)$$

4.1. KATALYTISCHES MODELL OHNE ERKENNUNG

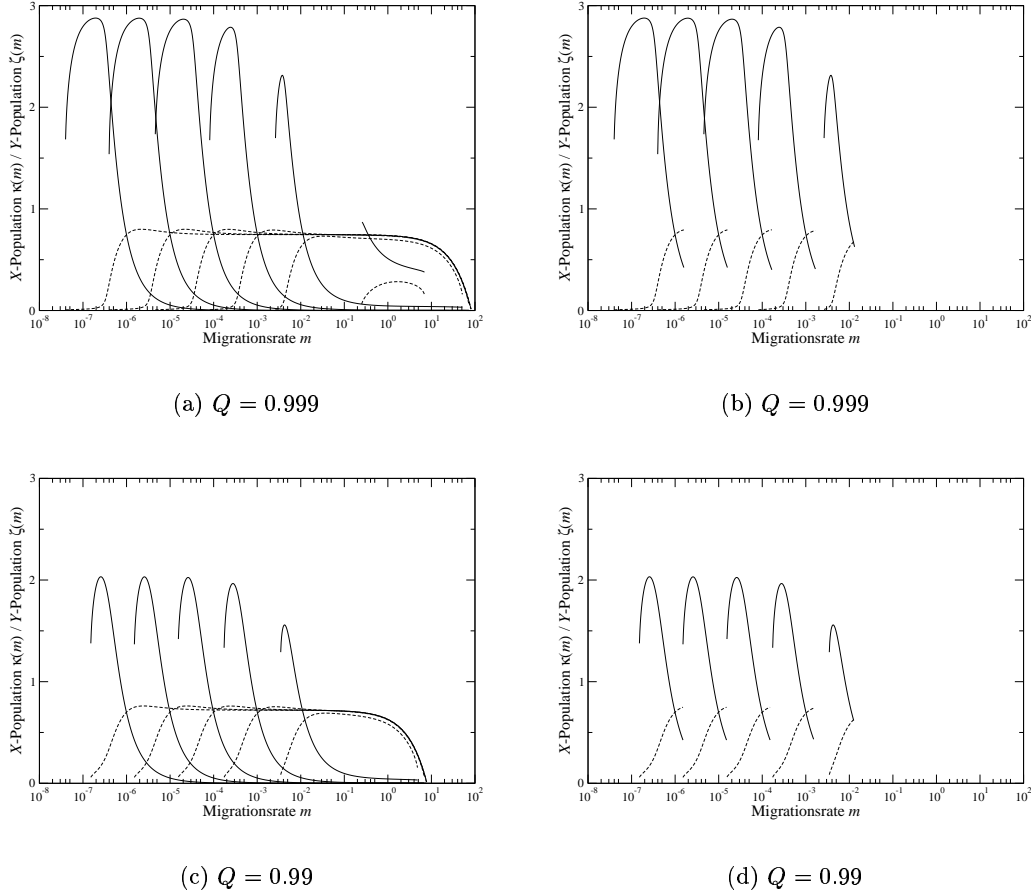


Abbildung 4.1: Die X - und Y -Populationen κ (durchgezogen) und ζ (gestrichelt) für angegebene Q und $D = 10^{-6}, 10^{-5}, 10^{-4}, 10^{-3}, 10^{-2}$ und 10^{-1} (jeweils von links nach rechts). Dabei stellen die Graphen auf der linken Seite die stationären und die der rechten Seite die asymptotisch stabilen Lösungen dar. Die Lösung bei $Q = 0.999$ und $D = 0.1$ existiert nur für die stationäre Lösung und zeigt sich als vollständig asymptotisch instabil. Für $Q = 0.99$ ist sie nicht mehr existent.

Alle weiteren Zustände sind nicht von Interesse, da bei diesen keine Veränderung aufgrund von Reaktionen möglich sind. Gemittelt über viele solcher Zustände und über die Zeit, wird Zustand 1 mit der Rate 1 ein neues Molekül erzeugen, Zustand 2 mit der Rate $1/2$ - da nur das X reaktiv ist - und Zustand 3 mit der Rate Null. Insbesondere Zustand 3 verdeutlicht, wie die Separation der Y - von den X -Molekülen den Einfluß der Parasiten reduzieren kann. Solch eine strikte Differenzierung zwischen den Zuständen und den daraus resultierenden Erzeugungsraten ist im homogenen Fall nicht möglich, da dort zu jedem Zeitpunkt ein Y von einem X repliziert werden kann. In räumlich strukturierten Systemen bedeutet dies, daß die Gesamterzeugungsraten der Y Sequenzen - sofern man den Mutationsfluß nicht berücksichtigt - reduziert ist. Somit wird eine Mutationsrate ungleich 0 überhaupt erst möglich. Allerdings gibt es zwei gegenspielende Prozesse, die eine mögliche Stabilisierung verhindern können. Zum einen ist eine zu hohe Migrationsrate zu nennen, bei der die Zustände zu schnell vermischt werden, so daß es keine effektive Separation geben kann. Ist hingegen die Migrationsrate zu klein, können sich die katalytischen Elemente nicht schnell genug über das System ausbreiten und

müssen sich auf wenige, einzelne Parzellen beschränken. Dies führt für große Zeiten aber dazu, daß irgendwann der absorbierende Zustand - kein X im System - erreicht wird. Aus dieser Überlegung muß es demnach einen Migrationsbereich geben, bei dem es eine Stabilisierung gibt. Interessanterweise kommt diese Stabilisierung dabei völlig ohne langreichweitige Muster aus; die einfache Trennung zwischen den Molekülen, induziert durch die Parzellen, ist bereits ausreichend. Somit läßt sich ableiten, daß Kooperation zwischen Molekülen in räumlich strukturierten Umgebungen für ein völlig frei evolvierbares Modell möglich ist. Daß die im Simplex getroffenen Aussagen auch in physikalischen Dimensionen relevant sind, wird in Kapitel 4.1.8 gezeigt.

Es ist evident, daß das vorliegende Modell aufgrund seiner Minimalität nicht 1:1 auf ein reelles biologisches System abbildbar ist. Ein erst kürzlich von Bartel veröffentlichtes Experiment [62], zeigt aber, daß - unter Laborbedingungen - RNA-Moleküle existieren, die gerade die Art von Polymerisation katalysiert, die für RNA-Replikation notwendig ist. Dieser Form von replizierendem System kommt das untersuchte Modell jedoch selbst in der vorliegenden Minimalität bereits recht nah.

4.1.6 Hopf Bifurkation und Grenzzyklen

Die vorhergehende Analyse ist von einer störungsfreien Umgebung ausgegangen, die in reellen Systemen nicht vorliegt. In diesen kann es immer vorkommen, daß ein im Gleichgewicht befindliches System durch äußere Einflüsse gestört wird. Kann diese Störung nicht kompensiert werden - liegt also kein asymptotischer Fixpunkt vor - so kann das System unter Umständen in den Einzugsbereich des Vakuums gelangen, was mit der völligen Extinktion einhergehen würde.

Um zu bestimmen, ob ein berechneter Zustand asymptotisch stabil, stabil oder nicht stabil ist, bedarf es in Systemen, die durch mehr als eine Zustandsvariable gekennzeichnet sind, die Berechnung der Eigenwerte einer Ableitungsmatrix. Für das PRESS Modell wurde dies bereits im Kapitel 4.1.4 erörtert, so daß an dieser Stelle nur noch einmal ergänzend auf Anhang A.11 hingewiesen werden soll, daß eine Einführung in die lineare Stabilitätsanalyse bietet. Berücksichtigt man bei der Berechnung von κ und ζ schließlich die Stabilität, so zeigt sich eine Reduktion des Migrationsbereich auf die asymptotisch stabilen Lösungen, die zwischen m_a und m_c liegen. Dabei bezeichnet m_c die Migrationsrate, bei der nicht mehr alle Eigenwerte aus der Stabilitätsanalyse negativ sind und vorherige Attraktoren (asymptotisch stabile Fixpunkte) verloren gehen, siehe Abb. 4.4 für eine Darstellung des Stabilitätsbereichs im Zusammenhang mit der kritischen Fehlerrate. Eine Untersuchung des Verhalten der Population jenseits von m_c zeigt, daß das Überschreiten dieser Grenze nicht direkt mit dem vollständigen Zerfall der Population verbunden ist. Statt dessen erscheinen periodische Lösungen, die sich stabil um den Fixpunkt der stationären Lösungen bewegen. Solche stabilen jedoch nicht asymptotisch stabilen Lösungen sind als Grenzzyklen bekannt. Der Übergang von asymptotisch stabilen Zuständen zu Grenzzyklen in Abhängigkeit eines Kontrollparameters, m , an einer kritischen Stelle, m_c , wird als Hopf Bifurkation bezeichnet. Zu deren Theorie gibt ebenfalls Anhang A.11 Auskunft. Diese Grenzzyklen werden für den exemplarischen Parametersatz $Q = 0.99$ und $D = 0.00001$ genauer untersucht. Dazu werden spezielle Migrationswerte ausgewählt, die einen repräsentativen Eindruck über das Verhalten der Grenzzyklen wiedergeben. Die Migrationswerte sind in Abb. 4.2 zu sehen. Für jeden Parametersatz wird mit Hilfe der Matrix \mathbf{R} die explizite Zeitabhängigkeit untersucht. Da die Analyse der stationären Zustände gezeigt hat,

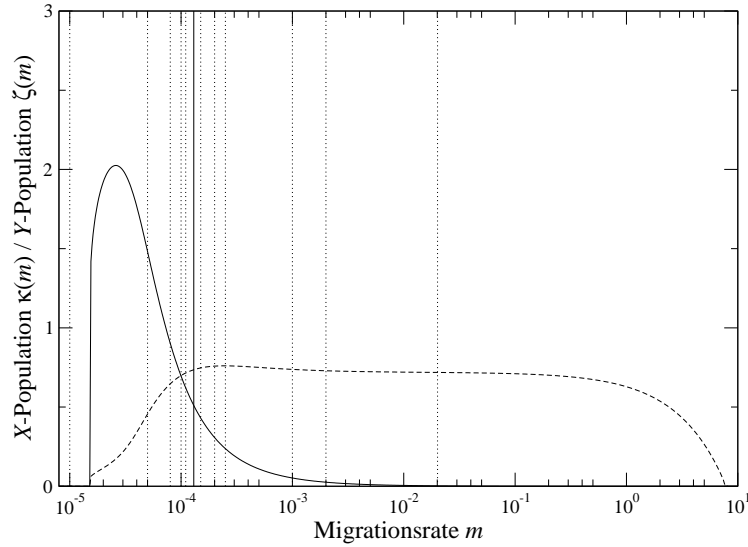


Abbildung 4.2: Die Abbildung 4.1(b) - $Q = 0.99$ und $D = 0.00001$ wird vergrößert dargestellt um die Migrationsschnitte darzustellen, bei denen die Stabilität numerisch untersucht werden soll. Das resultierende Verhalten von κ und ζ kann in Abb. 4.3 betrachtet werden. Die durchgezogene, vertikale Linie markiert die Grenze der Hopf Bifurkation.

daß es neben dem absorbierenden Vakuumfixpunkt nur eine weitere stationäre und gleichzeitig physikalische Lösung gibt, können bei der numerischen Berechnungen auch Anfangswerte benutzt werden, die hinreichend weit vom Vakuumspunkt entfernt sind, sonst aber beliebig sein dürfen. Dabei hat sich vor allem der Punkt $\kappa = n, \zeta = 0$, bzw. $P_4 = 1, P_{i \neq 4} = 0$ in den Wahrscheinlichkeitskoordinaten (Tabelle 4.1), als günstiger Startwert erwiesen. Die ermittelte Zeitabhängigkeit der Belegunswahrscheinlichkeiten und damit die Dynamik von $\kappa(t)$ und $\zeta(t)$ wird in Abb. 4.3 durch die Phasenraumdarstellung (wobei es sich im gegebenen Fall um einen von zehn auf zwei Dimensionen reduzierten Phasenraum handelt) und die explizite Darstellung der Zeitentwicklung illustriert.

Die Anwesenheit von Grenzzyklen zeigt sich auch in weiteren, biologisch motivierten Modellen. Vor allem treten sie in solchen Systemen auf, bei denen es eine starke Kopplung zwischen den einzelnen Systemkomponenten (Molekülspezies) gibt. Als Beispiel sei erneut das bereits häufig erwähnte Hyperzyklussystem zu nennen. Hier formieren sich zeitlich periodische Lösungen, wobei diese korrespondierend zu der Konzentration der sich katalytisch replizierenden Moleküle phasenverschoben auftreten. ähnliche zeitliche Korrelation erscheinen bei vielen nicht-linearen Systemen, so beispielsweise auch bei der Klasse der Räuber-Beute (Lotka-Volterra) Modelle [63] und im Zusammenhang mit Populationskonflikten bei der Suche nach evolutionär stabilen Strategien im Kontext der Koevolution [64, 65].

Die periodischen Schwankungen lassen sich fast immer dadurch erklären, daß (beispielsweise für den hier behandelten Fall) eine hohe Konzentration von X -Sequenzen zwangsläufig die Y -Moleküle häufiger repliziert und somit deren Population erhöht. Die nun häufig auftretenden Y beanspruchen jedoch entsprechend mehr Platz und verdrängen damit teilweise die katalytischen Elemente, was sowohl deren als schließlich auch die Y -Population reduziert. Die neu entstandenen Leerstellen können aber rasch von den X okkupiert werden, was die Konzentration anschnellen läßt und den periodischen Kreislauf schließt. Im Gegensatz zu stark gekoppelten Systemen, findet der Einfluß der Y -Moleküle auf die Replikatoren nur über die

Leerstellen statt, da die Parasiten katalytisch inaktiv sind. Im Fall von $n = 3$ ist die Rückkopplung recht stark, da die Besetzung einer Leerstelle durch einen Parasiten bereits die Replikation eines X auf einer Parzelle verhindert. Anders verhält es sich, wenn die Zahl der Leerstellen mit ansteigenden n vergrößert werden und die Besetzungsproblematik reduziert wird. Die direkte Wirkung der Y auf die X wird mit zunehmenden n beliebig klein und damit auch die treibende Kraft, die in gekoppelten Systemen Grenzzyklen entstehen läßt. Eine Untersuchung dieser Hypothese wird in Kapitel 4.1.9 gegeben.

4.1.7 Kritische Fehlerraten

Nachdem die Dynamik der Populationen behandelt wurde, sollen - auch um einen direkten Vergleich mit den Quasispezies zu erlauben - die kritischen Fehlerraten bestimmt werden. Wie bereits erläutert, kann dazu nicht der explizite Limes $Q_c = Q(\kappa \rightarrow 0)$ benutzt werden. Statt dessen muß für gegebene m und D ein κ ermittelt werden (im allgemeinen ungleich 0), bei dem Q maximal wird. Diese maximale Q kann dann als Q_c interpretiert werden, da ein Überschreiten dieses kritischen Wertes den sofortigen Zerfall der Population mit sich führen, ohne daß κ stetig gegen Null laufen würde. Die Bestimmung von Q_c ist jedoch nicht mehr vollständig analytisch möglich, kann aber für explizite Werte numerisch aus $Q(\kappa, m, D)$ berechnet werden. Abbildung 4.4 zeigt eine schematische Darstellung mit Bezug auf die kritischen Migrationsraten und Abb. 4.5 schließlich die Ergebnisse der Analysen. Dabei wird zwischen der stationären Lösungen (inklusive dem Bereich der Grenzzyklen) und den entsprechenden Lösungen unter Berücksichtigung der asymptotischen Stabilität unterschieden. Es zeigt sich, daß letztere erwartungsgemäß innerhalb eines Subbereichs der stationären Lösungen liegen. Die Gleichheit beider Bereiche in der Nähe von m_a folgt aus der Tatsache, daß die stationären Lösungen dort ebenfalls asymptotisch stabil sind. Da die Bestimmung von etwaigen Grenzzyklen sich bei der Komplexität der Ausdrücke als extrem schwierig erweist, sollen für die weitere Behandlung nur die asymptotisch stabilen Werte benutzt werden.

4.1.8 Simulationsergebnisse und euklidischer Raum

Die in den vorherigen Kapiteln hergeleiteten analytischen und numerischen Ergebnisse aus dem PRESS Modell werden mit Hilfe der stochastischen Simulationen verifiziert. Bei einer Übereinstimmung ist es möglich - wie bereits im Quasispezies Kapitel - das Verhalten des kooperativen Systems auch in niederen, physikalischen Raumdimensionen per Simulation zu studieren.

Abbildung 4.6 zeigt die mit Hilfe der in Kap. 2.2 beschriebenen Simulationemethode gewonnenen Ergebnisse bis zum Punkt der Hopf Bifurkation für den Parametersatz $Q = 0.999$ und $D = 0.02$ in ein, zwei, drei und unendlichen (Simplex-Geometrie) Dimensionen. Die durchgezogene Linie gibt die analytische Kurve für den angegebenen Parametersatz im PRESS Modell wider. Man erkennt gute Übereinstimmungen zwischen den stochastischen Simulationen und den analytischen Ergebnissen. Es läßt sich erkennen, daß in allen Dimensionen die prinzipielle Kurvenform beibehalten wird. Je höher die Dimension wird und sich somit dem Simplex nähert, desto ähnlicher wird auch der Kurvenverlauf zu dem des analytischen Modells. Allerdings ist eine systematische Verschiebung des Kurvenmaximas zu beobachten, die mit der Dimensionsreduzierung einhergeht und eine Erhöhung der Migrationsrate zur Folge hat. Geht man davon aus, daß es keine neuen Effekte (langreichweitige Musterbildung) in den niederen Dimensionen gibt, die die Stabilität verändern könnten, ist es möglich, die unterschiedlichen

Diffusionsbereiche durch eine einfache Approximation aufeinander abzubilden. Dazu wird die Annahme benutzt, daß die Dimensionalität nur die Anzahl direkter Nachbarn festlegt, die mit einem Migrationsereignis erreicht werden können, wobei die Anzahl nächster Nachbarn auf den benutzen rechtwinkligen Gittern gerade $2d$ ist. Nach einem Migrationsereignis, wird ein ausgewähltes Molekül auf einer anderen Parzelle landen. Springt es von dieser zufällig zu einer weiteren, so gibt es - abhängig von der Dimension - eine Wahrscheinlichkeit, $1 - p$, zurück auf die erste Parzelle zu gelangen. Die Größe $p = (2d - 1)/(2d)$ gibt an, mit welcher Wahrscheinlichkeit ein Molekül nicht auf die Parzelle zurückspringt, von der es gekommen ist. Im Simplex ist diese Größe aufgrund der unendlichen Anzahl von Nachbarn Null.

Vergleicht man eine Migrationsrate für eine beliebige endliche Dimension, m_d , mit der Migrationsrate, die im PRESS Modell benutzt wird, m_∞ , ergibt sich mit Hilfe der Wahrscheinlichkeit p die einfache Beziehung

$$m_d = \frac{2d - 1}{2d} m_\infty. \quad (4.15)$$

Abbildung 4.7 zeigt die Anwendung dieser Skalierung auf die Werte der unterschiedlichen Parameter. Die gute Übereinstimmung der Lage der Maxima (als Referenzpunkte) belegt die Korrektheit der Skalierungsgleichung.

Es muß betont werden, daß das unendlich dimensionale Selbstkonsistenzmodell nur eine Näherung für das Verhalten in euklidischen Dimensionen ist, was auch an der unterschiedlichen Breite der Kurven aus Abb. 4.7 zu sehen ist, die nicht auf einfache Art wegskaliert werden können. Das liegt vor allem daran, daß es räumliche Symmetriebrechung in niederen Dimensionen und damit globale Musterbildung geben kann. Die Existenz dieser Muster können aber, wie bereits zu Beginn dieser Arbeit anhand der Hyperzyklen demonstriert, die Stabilität verändern und somit Unterschiede zum Verhalten auf dem Simplex manifestieren. Für das katalytische Modell scheint dies - für den untersuchten Parameterbereich - jedoch nicht der Fall zu sein. Die leichte Erhöhung des Kurvenmaximas und die geringe Ausdehnung des Stabilitätsbereichs lassen - wenn überhaupt - nur auf leichte Raumkorrelationen schließen. Positiv in jedem Fall ist die Tatsache, daß die Stabilisierungseffekte in physikalischen Dimensionen immer größer zu sein scheinen, als die aus dem PRESS Modell berechneten. Der Simplex gibt somit eine Art worst-case im Bezug auf die rauminduzierte Kooperation an.

4.1.9 Vergrößerung der Parzellenpopulation

Um die Untersuchungen abzuschließen muß letztlich noch gezeigt werden, daß auch die Parzellengröße keinen entscheidenden Einfluß auf die Stabilität des Systems zeigt. Dazu wird das Verhalten der Population für $n = 4$ bis $n = 10$ untersucht. Da nach Gl. (3.38) bereits $N_Z = 15$ bzw. $N_Z = 66$ Zustände notwendig sind, kann dies (im Rahmen der gegebenen Hilfsmittel) nur noch numerisch behandelt werden. Dazu wird wie in Anhang. A.5 beschrieben die Matrix \mathbf{L} berechnet und gemäß Anhang. A.10 zur Matrix \mathbf{R} transformiert, mit der numerische Ergebnisse gewonnen werden können. Abbildung 4.8 zeigt das normierte Verhalten der X -Population und Abbildung 4.9 der Y -Population für $Q = 0.999$ und $D = 0.02$ normiert auf die unterschiedlichen n .

Bei der Untersuchung zeigt sich eindeutig, daß die analytisch für $n = 3$ im Hinblick auf die Stabilität gefundenen Ergebnisse kein Artefakt der kleinstmöglichen Parzellenpopulation sind. Es ist sogar so, daß der Stabilitätsbereich für größer werdende n wächst. Normiert auf die maximale Parzellengröße steigen die Peaks der X - und Y -Population aus Abb. 4.8

bzw. Abb. 4.9 mit n erst rasch an, um dann beim Wert $\kappa/n \approx 0.92$ zu verweilen. Es ist evident, daß die Stabilität für beliebig große n auch wieder reduziert werden muß, da der Fall $n \rightarrow \infty$ den homogenen Fall pro Parzelle darstellen würde. Ob es jedoch ein kritisches n gibt, bei dem die Stabilität schlagartig abfallen würde, und wo dieses liegen könnte, bleibt bisher unbeantwortet.

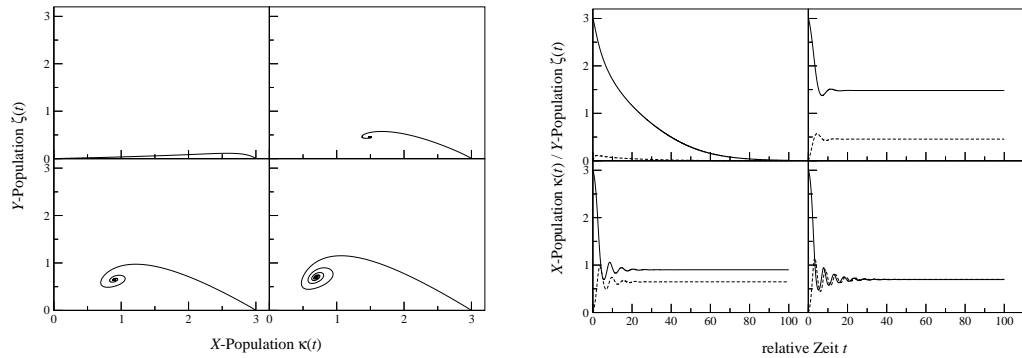
Dafür zeigt sich ein anderes interessantes Ergebnis, daß sich mit der Frage nach der Stabilität in Systemen mit $n > 3$ Parzellen befaßt. Die Hypothese, daß die Vergrößerung der Parzellen den Rückkopplungseffekt der Parasiten auf den Katalysten reduziert, zeigt sich recht deutlich mit ansteigenden n , so daß bereits bei $n = 5$ für den gegebenen Parametersatz keine Grenzzyklen mehr gefunden wurden. Der gesamte Stabilitätsbereich kann somit als asymptotisch stabil bezeichnet werden.

Für den Fall $n = 4$ zeigt sich ein Bereich, bei dem die asymptotische Stabilität erst verloren geht, um anschließend bei höheren Migrationsraten wieder aufzutauchen. Die Ursache für dieses Verhalten sei auf weiterführende Untersuchungen verlegt. Es kann jedoch festgestellt werden, daß die numerisch untersuchten Parameterbereiche belegen, daß eine Stabilisierung des katalytischen Systems für $n > 3$ nicht nur möglich ist, sondern sogar deutlich ausgeprägter zu sein scheint. Der Fall $n = 3$ erscheint bei den Analysen als minimaler Spezialfall.

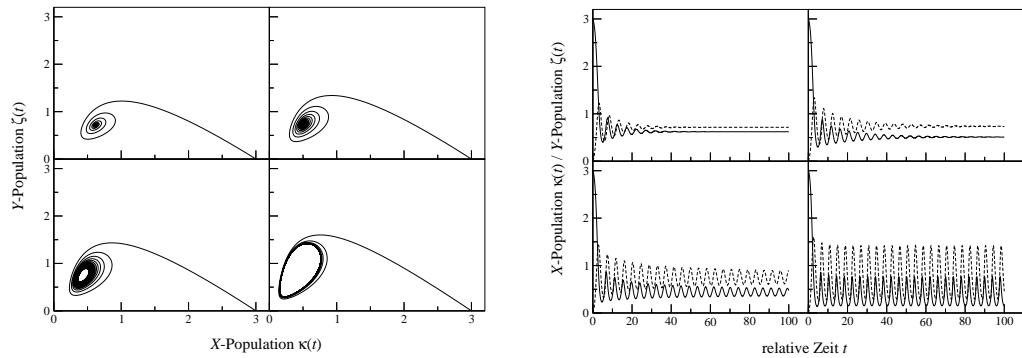
Die bisherigen Kapitel haben deutlich gezeigt, daß es möglich ist, das einfachste wechselwirkende System im Raum zu stabilisieren. Unabhängig von der Raumdimension und der Größe der Parzellen (moderate n). Die vereinfachenden Annahmen erhöhen dabei die Problematik der Stabilisation, als daß sie sie vereinfachen würden. So erzeugt die Vernachlässigung der Rückmutation einen absorbierenden Vakuumzustand, der - einmal erreicht - jegliche Rückkehr zur Funktionalität verbietet. Auch die gewählte single-peaked Landschaft, erlaubt kein evolutives Hillclimbing, wie es z.B. bei einer glatten „Fujiyama“ Landschaft möglich ist [16]. Daher kann davon ausgegangen werden, daß die gefundenen Ergebnisse auch für Umgebungen angewandt werden können, die deutlich bessere Voraussetzungen liefern. Dort sollten die zu erwartenden Ergebnisse entsprechend besser sein.

Im folgenden wird eine Methode vorgestellt, die zeigt, wie eine mögliche Stabilität ausgebaut werden kann, die partielle Erkennung. Dabei stellt sich heraus, daß eine besondere Eigenschaft der RNA-Moleküle - die Faltung - im Zusammenhang mit partieller Sequenzerkennung eine Stabilitätssteigerung generiert, die eine deutliche Sequenzverlängerung zuläßt.

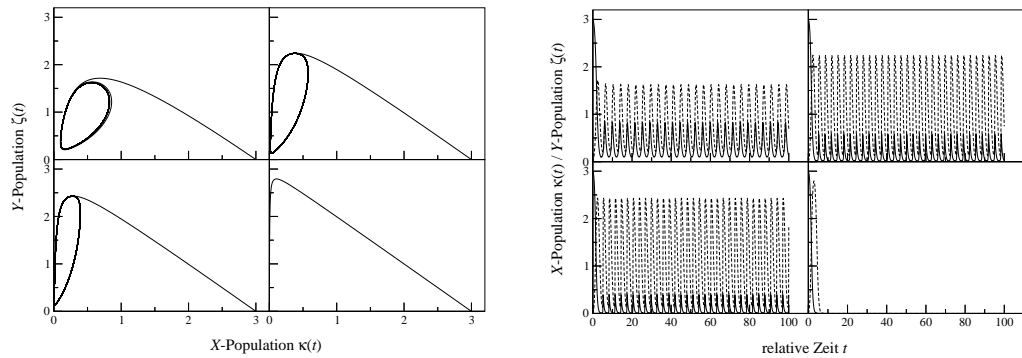
4.1. KATALYTISCHES MODELL OHNE ERKENNUNG



(a) $m = 0.00001, 0.00005, 0.00008, 0.001$



(b) $m = 0.00011, 0.00013, 0.00015, 0.00020$



(c) $m = 0.00025, 0.001, 0.002, 0.0201$

Abbildung 4.3: Für den exemplarischen Parametersatz $Q = 0.99$ und $D = 0.0001$ wird die Stabilität numerisch untersucht. Die linken Graphen stellen die Phasenraumplots - aufgespannt durch κ und ζ - für die unterschiedliche m (jeweils von links oben nach rechts unten) und die rechten den entsprechenden Zeitverlauf für κ (durchgezogen) und ζ dar.

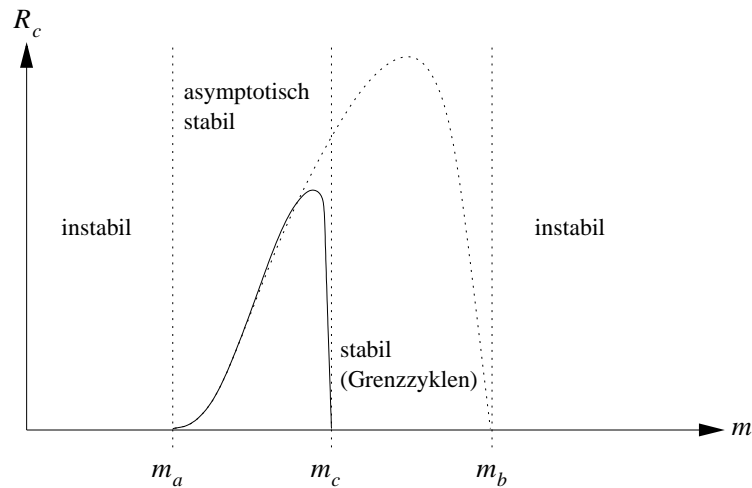


Abbildung 4.4: Die Abbildung zeigt die kritischen Migrationsraten (vertikale Linien) und die kritische Fehlerrate als Funktion der Migrations, wobei zwischen asymptotisch stabilen (durchgezogen), $m_a \leq m < m_c$, und nicht asymptotisch stabilen (gepunktet), $m_c < m \leq m_b$ Zuständen unterschieden wird.

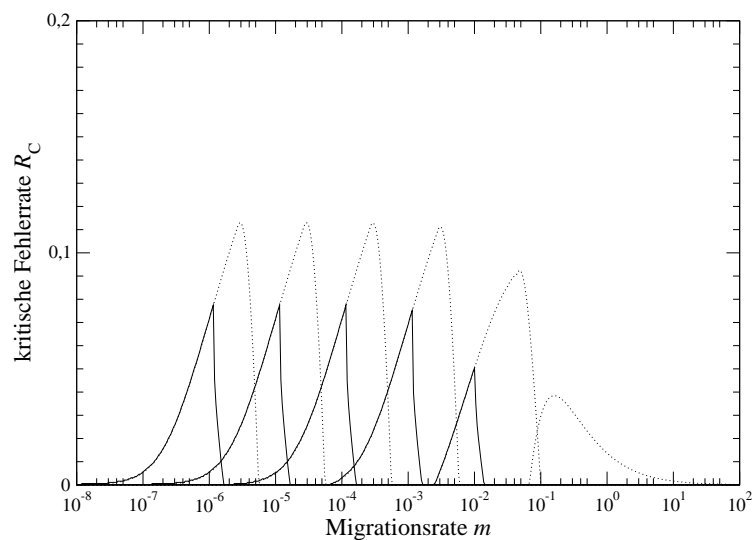


Abbildung 4.5: Die Abbildung vergleicht die kritischen Fehlerraten für den stationären Fall (gepunktet) und für den Unterbereich der asymptotischen Stabilität. Dabei wurden die Standardzerfallsraten $D = 10^{-6}, 10^{-5}, 10^{-4}, 10^{-3}, 10^{-2}$ und 10^{-1} (jeweils von links nach rechts) benutzt.

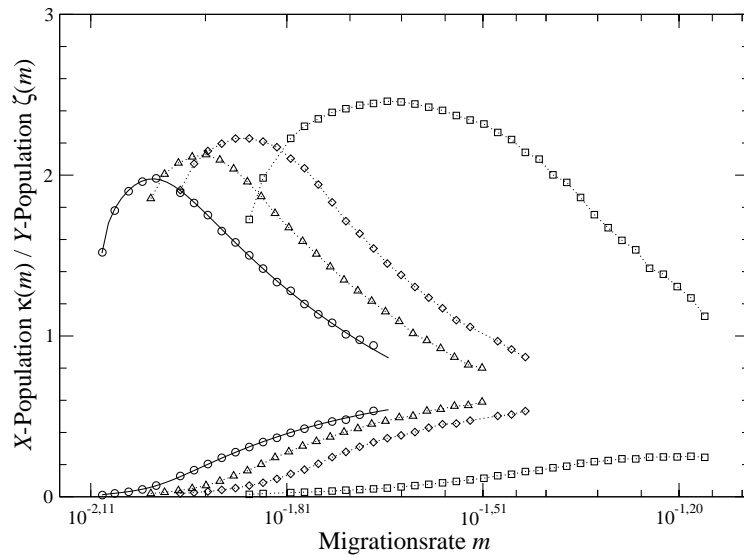


Abbildung 4.6: Die Abbildung vergleicht für den Parametersatz $Q = 0.999$ und $d = 0.02$ zum einen für den Simplex Fall die analytischen Ergebnisse (durchgezogene Kurve) mit den stochastischen Simulationen (\circ), wobei sich eine hervorragende Übereinstimmung zwischen den beiden Methoden zeigt. Weiterhin zeigt die Abbildung für obige Parameter das Verhalten für $d = 1$ (\square), $d = 2$ (\diamond) und $d = 3$ (\triangle).

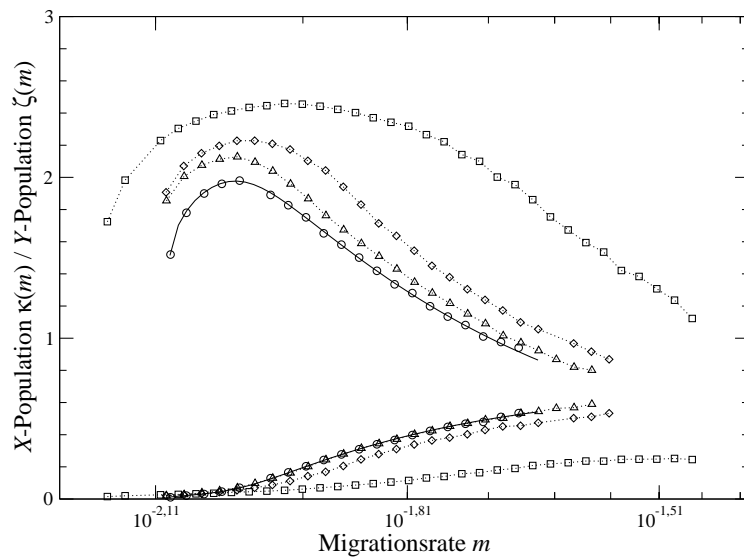


Abbildung 4.7: Die Abbildung zeigt die Reskalierung der Migrationsraten entsprechend Gl. (4.15). Für die Bedeutung der Symbole siehe Abb. 4.6

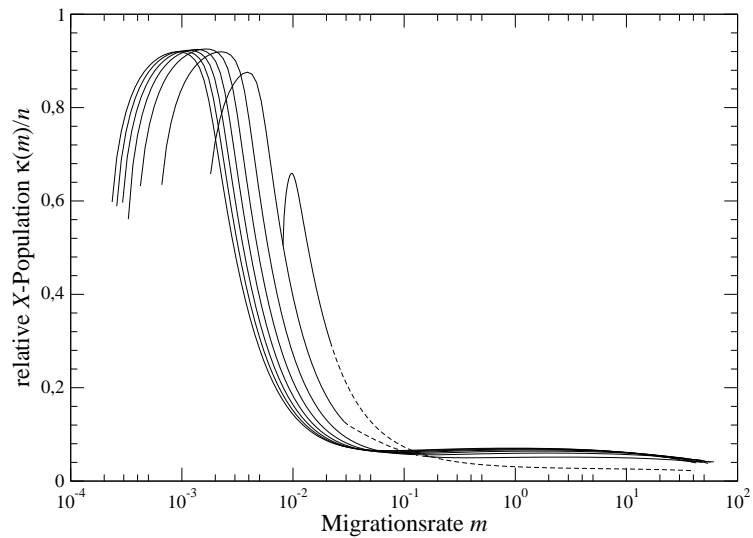


Abbildung 4.8: Die Abbildung zeigt das Verhalten der auf die Parzellengröße normierten X -Population für $Q = 0.999$ und $d = 0.02$ für $n = 3, \dots, 10$ (von rechts nach links). Die gestrichelten Abschnitte verdeutlichen die Existenz von Grenzzyklen, die ab $n = 5$ vollständig verschwunden sind und bei $n = 4$ nur noch in einem kleinen, von asymptotisch stabilen Zuständen eingerahmten Bereich zu finden sind.

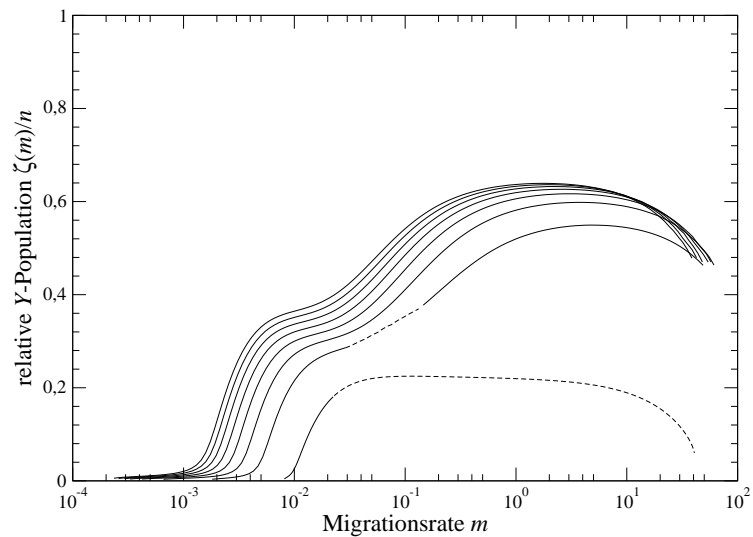


Abbildung 4.9: Die Abbildung zeigt das Verhalten der auf die Parzellengröße normierten Y -Population für $Q = 0.999$ und $d = 0.02$ für $n = 3, \dots, 10$ (von unten nach oben). Die Existenz von Grenzzyklen wird durch die gestrichelten Bereiche gekennzeichnet, siehe auch Abb. 4.8

4.2 Partielle Sequenzerkennung

Das Prinzip der partiellen Sequenzerkennung wird im Hinblick auf eine Erhöhung der evolutiven Stabilität untersucht, wobei sich die Faltung als effektive Methode herausstellt, mit der Sequenzverlängerungen möglich sind.

4.2.1 Reduzierung des Sequenzraums

Bisher wurde das einfache Kooperationsmodell ohne die Möglichkeit der Sequenzerkennung behandelt. Wie zu Beginn dieser Arbeit bereits erörtert, ist eine vollständige, d.h. 100%ige Erkennung allein schon aus physikalischer Sicht problematisch, da die vollständige Oberfläche zweier Makromoleküle verglichen werden müßte. Statt dessen beschränkt die Natur sich im allgemeinen darauf, nur Teile dieser Oberflächen für eine Teilerkennung zu nutzen.

Eine Erkennung impliziert für evolvierbare Systeme immer zwei Vorteile:

1. Symbiotische Spezies werden erkannt und eine Wechselwirkung mit ihnen wird ermöglicht.
2. Parasiten werden erkannt und eine Wechselwirkung mit ihnen wird verhindert.

Die auf den ersten Blick als Trivialität erscheinende Aufzählung zeigt jedoch zwei verschiedene evolutive Aspekte auf. Eine Wechselwirkung mit Symbionten erlaubt beispielsweise eine Aufteilung von genetischer Information auf individuelle (aber genetisch verwandte) Träger, die zusammen die Fitneß der Gruppe bzw. Sippe (im Gegensatz zur individuellen Darwinschen Fitneß) bestimmen. Diese Art der Gruppenzugehörigkeit wird auch in der Populationsbiologie benutzt und dort mit dem Begriff „kin Selection“ in Verbindung gebracht [66]. Auch im Zusammenhang mit Altruismus spielt dieser Begriff eine wichtige Rolle, siehe [35] als Review.

Im Rahmen des einfachen katalytischen Modells soll nur der zweite Aspekt der Erkennung behandelt werden, die Erkennung von möglichen Parasiten, wobei der Effekt der Teilsequenzerkennung im Hinblick auf die Erhöhung der kritischen Fehlerrate, R_c , untersucht wird. Die partielle Erkennung wird dabei wie folgt modelliert: Sei X_i ein wie in Kapitel 3.1.1 beschriebenes Molekül mit der Binärsequenz i der Länge ν , wobei die spezielle Sequenz X_0 erneut dem katalytischen Molekül X entspricht und alle anderen Sequenzen als nicht-funktionale Elemente Y bezeichnet werden und sei $d(i, j)$ der Hamming Abstand zwischen den Sequenzen i und j . Aus den ν Monomeren werden nun k_r an beliebigen aber festen Positionen ausgewählt, die als Erkennungsregion für den Kopierer und dem zu Kopierenden dienen. Nur unter der Voraussetzung, daß diese Bits bei beiden Sequenzen übereinstimmen, kann ein bestimmtes Molekül an den Kopierer andocken und ein Kopiervorgang wird initiiert. Da die exakte Position der k_r aufgrund der vollständigen Symmetrie des Binärstrings keinerlei Einfluß hat, sollen des besseren Verständnis wegen die ersten k_r Bits einer jeglichen Sequenz gewählt werden. Ist eins der k_r Erkennungsbits nicht identisch, findet der Replikationsprozeß nicht statt, vergl. auch Abb. 4.10.

Aus der Modellierung wird klar, daß die Erkennung bzw. das Andocken an die funktionelle Replikase für Moleküle mit exakt der gleichen Sequenz am besten funktioniert. Diese Einschränkung erweist sich als weniger drastisch, wenn man berücksichtigt, daß nur ein kleiner Teil der Gesamtsequenz zur Erkennung herangezogen wird und davon ausgegangen werden kann, daß sich im Laufe der Evolution durchaus eine Replikatorsequenz entwickelt hat, die

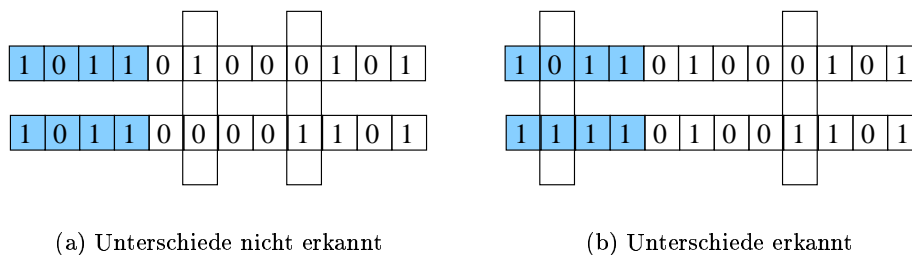


Abbildung 4.10: Darstellung zweier $\nu = 12$ Bit langen Sequenz mit $k_r = 4$ Erkennungsbits, wobei die untere ein 2 Fehlermutant ist. In Bild (a) liegen die Mutationen alle außerhalb der Erkennungsregion (dunkel), weshalb die zweite Sequenz nicht als falsch erkannt und somit von der ersten kopiert würde. Bild (b) zeigt die Situation, bei der mindestens eine der Mutationen in der Erkennungsregion liegt und die zweite Sequenz als falsch erkannt würde.

gerade diese Forderung erfüllt. Zur Festigung dieser These soll auf die Theorie der neutralen Netze hingewiesen werden [67], die sich mit der Fragestellung der Veränderung von Sequenzen unter Erhaltung der Funktionalität beschäftigt.

Da bei der Sequenzerkennung prinzipiell der vollständige Sequenzraum modelliert werden muß, eine Behandlung der 2^ν unterschiedlichen Sequenzen im Kontext des PRESS Modells analytisch jedoch wenig aussichtsreich erscheint, wird im folgenden eine Art Meanfield Verteilung der Moleküle im Sequenzraum benutzt um anschließend die Wahrscheinlichkeit einer Erkennung zu berechnen. Sei q wieder die Wahrscheinlichkeit mit der einzelne Monomere richtig kopiert werden, was entsprechend $Q = q^\nu$ als Wahrscheinlichkeit einer vollständig korrekten Replikation definiert. Teilt man den Sequenzraum in Fehlerklassen - definiert über den Hamming Abstand zur katalytischen Sequenz X_0 - auf, so ergibt sich für die Wahrscheinlichkeit, nach einer Replikation einen m Fehlermutanten als Ergebnis zu haben über die Binomialverteilung

$$P(m) = \binom{\nu}{m} q^{\nu-m} (1-q)^m. \quad (4.16)$$

Dabei entspricht $P(m=0)$ - die Wahrscheinlichkeit, das bei der Reaktion kein Fehler gemacht wurde - logischerweise gerade dem Faktor Q . Nur die Veränderungen nach einer einzigen Replikation sind relevant, da der Prozeß der Erkennung vor jeder Replikation stattfindet und somit mehrfach ausgeführte Replikationen ohne Erkennung nicht möglich sind.

Für die weiteren Untersuchungen wird es wichtig sein, die bedingte Wahrscheinlichkeit einen m Fehlermutanten nach einer Replikation zu erhalten, unter der Voraussetzung, daß diese tatsächlich fehlerbehaftet war, also mit der Wahrscheinlichkeit $1-Q$ keine exakte Kopie erzeugt hat. Somit ergibt sich für die mittlere Fehlerklasse \bar{m}

$$\bar{m} = \frac{1}{1-Q} \sum_{m=1}^{\nu} m P(m) = \nu \frac{1-q}{1-Q}. \quad (4.17)$$

Der Zusammenhang mit der Erkennung wird über die verschiedenen Permutationen der mutierten Bits manifestiert. Dabei entspricht die Wahrscheinlichkeit r , daß m Mutationen so angeordnet werden, daß nicht eine in den Bereich der ersten k_r Bits (Monomere) fällt, gerade der Wahrscheinlichkeit mit der ein Y nicht als solches erkannt und somit von einem X repliziert wird. Diese Wahrscheinlichkeit r ergibt sich aus der kombinatorischen Anzahl von

Mutationen, die nicht im Erkennungsbereich liegen zur Anzahl aller Kombinationen, also

$$r = \frac{\binom{\nu - k_r}{\bar{m}}}{\binom{\nu}{\bar{m}}} = \frac{(\nu - k_r)! (\nu - \bar{m})!}{\nu! (\nu - k_r - \bar{m})!}. \quad (4.18)$$

Da \bar{m} im allgemeinen eine reelle Zahl sein wird, soll für die weiteren Berechnungen die Eulersche Gammafunktion benutzt werden, definiert über $\Gamma(z) = \int_0^\infty t^{z-1} e^{-t} dt$. Für den Fall eines ganzzahligen z ergibt sich der Zusammenhang mit der Fakultätsfunktion, $\Gamma(z+1) = z!$. Somit erhält man für reellwertige \bar{m}

$$r = \frac{\Gamma(\nu - k_r + 1) \Gamma(\nu - \bar{m} + 1)}{\Gamma(\nu + 1) \Gamma(\nu - k_r - \bar{m} + 1)}, \quad (4.19)$$

eine Verallgemeinerung, die die Benutzung der Erkennungsabhängigkeit für beliebige Abbildungen erlaubt, bei denen eine effektive mittlere Fehlerklasse berechnet werden kann. (Siehe dazu auch die Anmerkung im Kapitel 4.2.2.) Dementsprechend liegen einzelne Mutationen mit der Wahrscheinlichkeit $1 - r$ im Erkennungsbereich der k_r Monomere. Dieser Teil der Y -Sequenzen - mit Y_1 bezeichnet - kann nicht an das X andocken und wird folglich auch nicht kopiert. Der Anteil r hingegen - Y_0 - hat nur an den Stellen Mutationen, die nicht zur Andocksequenz gehören und werden von den X katalysiert. Hierdurch wird der Sequenzraum, der zur Spezies Y korrespondiert, in zwei Teile zerlegt wird. Die Y_1 -Sequenzen, also der nicht kopierbare Anteil, $1 - r$, erfährt nur noch einen Zufluß aufgrund der fehlerhaften Kopien bei der Replikation der X - bzw. Y_0 -Moleküle.

Im homogenen Fall würde ein noch so geringer Anteil von Y_0 -Sequenzen dafür sorgen, daß jeglicher Kopierfehler der X -Sequenzen deren Aussterben zur Folge hat. Das bedeutet aber, daß das Prinzip der partiellen Sequenzerkennung nicht in homogenen Räumen funktioniert. In räumlich aufgelösten Systemen gibt es jedoch zwischen den Migrationsraten m_a und m_b einen Bereich, der eine Stabilisierung der X -Spezies trotz Mutationen, $Q \neq 1$, erlaubt. Das heißt, daß dort Replikationen stattfinden, bei denen fehlerhafte Kopien entstehen, aber das Gesamtsystem trotz dieser Mutationen stabil bleibt. Ein Teil dieser Kopien, nämlich die Y_1 -Sequenzen werden nun nicht weiter von den X repliziert, so daß ihre Konzentration reduziert wird. Daß nur der Bruchteil r der Parasiten von X repliziert wird, der zur Klasse der Y_0 -Sequenzen gehört hat zur Folge, daß bei einer gegebenen Mutationsrate weniger echte Parasiten entstehen, was eine entsprechende zusätzliche Stabilisierung gewährleistet und höhere Fehlerraten erlaubt.

Als Konsequenz kann abgeleitet werden, daß partielle Sequenzerkennung nur in Systemen funktioniert, die *per se* eine minimale Grundstabilität gegenüber mutativen Veränderungen aufweisen, da die Erkennung diese zwar ausbauen (vervielfachen) kann, aber von sich aus nicht in der Lage ist, diese aus dem Nichts zu erzeugen. Für den Fall der einfachen (linearen) partiellen Sequenzerkennung, kann der Faktor r für realistische Qualitätsfaktoren stark vereinfacht werden. Die mittlere Anzahl von Mutationen \bar{m} wurde für den Fall $\nu = 20$ und unterschiedlichen Q berechnet, siehe Tabelle 4.2. Man sieht, daß für die mittlere Fehlerklasse \bar{m} erst bei Qualitätsfaktoren, die unter den Wert 0.5 liegen, deutliche Abweichungen von 1 zu erwarten sind. Da die bisherigen Ergebnisse beim Kooperationsmodell jedoch keine Werte $Q < 0.5$ zeigten, kann die Näherung $\bar{m} = 1$ gemacht werden. Dabei muß jedoch beachtet werden, daß eine partielle Erkennung auch eine Erniedrigung von Q mit sich bringt, so daß die Approximation nur für hinreichend kleine k_r gemacht werden kann. Substituiert man \bar{m}

Q	\bar{m}	Q	\bar{m}	Q	\bar{m}
0.99	1.005	0.90	1.051	0.50	1.360
0.98	1.010	0.80	1.101	0.40	1.500
0.97	1.015	0.70	1.178	0.30	1.670
0.95	1.025	0.60	1.261	0.20	1.933

Tabelle 4.2: Abhängigkeit der mittleren Anzahl von Mutationen von der Qualität Q für fehlerbehaftete Sequenzen. Es zeigt sich, daß für \bar{m} erst bei $Q < 0.5$ eine deutliche Abweichung von 1 zu sehen ist.

in Gl. (4.19) mit 1, vereinfacht sich r zu

$$r_1 = \frac{\nu - k_r}{\nu}. \quad (4.20)$$

Der Index 1 soll dabei den linearen Modus der Erkennung verdeutlichen, im Gegensatz zur jetzt vorgestellten Faltungsmethode.

4.2.2 Information und Struktur

Interpretiert man partielle Erkennung als das Ergebnis eines Andockprozeß eines Moleküls an ein anderes, so muß in diesem Zusammenhang die Faltungseigenschaft reeller RNA-Sequenzen bzw. Proteinen berücksichtigt werden. Diese Faltung hat zur Folge, daß die Moleküle neben ihrer primären Struktur - also der eigentlichen Sequenz - auch noch sekundäre und tertiäre Strukturen aufweisen.

Ob zwei RNA-Stränge oder Proteine andocken können um so in irgendeiner Form zu wechselwirken, wird im allgemeinen nicht durch ihre primäre, sondern vielmehr über ihre räumliche, also den sekundären und tertiären Strukturen bestimmt. Dabei ist der Zusammenhang zwischen den unterschiedlichen Strukturen nicht perfekt vorhersagbar. Stabile Faltungen sind prinzipiell durch das Minimum ihrer freien Energie charakterisiert, doch ist die exakte Berechnung der sekundären Struktur aus einer gegebenen Sequenz aufgrund der extrem hohen permutativen Kombinatorik nur für relativ kurzkettinge Moleküle möglich. Für langkettige Moleküle gibt es zwar effektive Algorithmen [68, 69], doch können sie letztlich auch nur approximative Lösungen geben.

Hat man für eine Sequenz i eine Abbildung gefunden, so zeigt sich, daß eine Sequenz j , die sehr ähnlich zu i ist ($d(i, j)$ ist klein) eine ähnliche oder vollständig andere Faltung erfahren kann. Dies hängt davon ab, an welcher Stelle sich die Monomere befinden, die Sequenz i von j unterscheiden. Über die Häufigkeit und die Verteilung von kritischen Monomersubstitutionen haben in letzter Zeit vor allem Reidys *et. al.* mit Hilfe der Random Graph Modelle [70, 71] Aussagen treffen können.

Für die weiteren Untersuchungen soll der Extremfall modelliert werden, daß sich jede Monomersubstitution in eine neue völlig unkorrelierte Sekundärstruktur resultiert. Zusammen mit der bereits erörterten linearen Erkennung werden so die zwei möglichen Extremfälle studiert, zwischen denen alle realen Faltungsmodelle liegen müssen. Kann bei realistischen Modelle die gemittelte Fehlerklasse \bar{m} berechnet (oder abgeschätzt) werden, so erlaubt die Benutzung von Gl. 4.19 eine Implementierung in das PRESS Modell. Für den hier behandelte Extremfall einer perfekten Faltung sollen folgende Bedingung gelten:

1. Die Abbildung ist frei von Korrelationen, also rein zufälliger Natur.

2. Ist eine Abbildung gewählt, so gilt sie für alle Instanzen der Sequenzen.
3. Die Abbildung ist bijektiv.

Die Forderung nach Korrelationsfreiheit bestimmt die Qualität der zufällig gewählten Zuordnung und ist wichtig für die weiteren Ableitungen. Bei jeder Abbildung, die so konstruiert wird, daß nach und nach jeder einzelnen Sequenz eine Struktur zugeordnet wird, ist Forderung 2 automatisch erfüllt. Zeigen alle diese Abbildung auf unterschiedliche Strukturen, so ist die Abbildung zusätzlich bijektiv, wie z. B. in Abb. 4.11 zu sehen. Das entscheidende Kriterium solcher Abbildungen ist es, daß aufgrund der zufälligen Faltungsabbildung zwischen der Ähnlichkeit zweier benachbarter Strukturen keine (oder in realistischeren Modellen wenig) Korrelation mit den entsprechenden Sequenzen und *vice versa* gibt.

Erkennung im Strukturraum bedeutet, daß zwei Moleküle andocken können, wenn Teile ihrer sekundären Struktur zusammen passen. Das Maß bzw. die Strenge der Erkennung wird wie schon im vorherigen Kapitel über k_r Bits definiert, diesmal jedoch im Strukturraum, also bezüglich der gefalteten Sequenzen. In diesem Raum können die k_r Bits als die Größe der Erkennungsregion interpretiert werden. Wie aufgrund der Faltung die Erkennungsbits über die Sequenz verteilt werden, zeigt Abb. 4.12 schematisch. Greift man zwei Sequenzen heraus, die im Sequenzraum nicht identisch sind, also mindestens einen Hamming Abstand von 1 besitzen, so sind ihre Strukturen aufgrund der gewählten zufälligen Abbildung unkorreliert. Dies vereinfacht die Herleitung der Erkennungswahrscheinlichkeit, $1 - r$, enorm, da keinerlei explizite Rechnungen über die Verteilung von eventuellen Mutationen gemacht werden müssen. Hat man zwei Strings (die Numerierungsstrings der Strukturen) mit zufälliger Binärsequenzen der Länge ν , wobei k_r übereinstimmen sollen, so gibt es genau $(2^{\nu-k_r} - 1)$ verschiedene. Nun wird eine Sequenz genau dann kopiert, wenn gerade die ersten k_r Bits übereinstimmen und die Sequenz somit als eine Y_0 -Sequenz erkannt wird. Da es insgesamt $(2^\nu - 1)$ Sequenzen gibt, ergibt sich

$$r_f = \frac{2^{\nu-k_r} - 1}{2^\nu - 1}. \quad (4.21)$$

Für hinreichend große Gesamtsequenzlängen ist der Term 1 im Nenner zu vernachlässigen. Fordert man weiterhin kleine Erkennungsregionen, kann Gl. (4.21) zu $r_f \approx 2^{-k_r}$ vereinfacht werden, womit es quasi Gesamtlängenunabhängig wäre. Dies würde jedoch bedeuten, daß die Sequenz beliebig verlängert werden kann, ohne daß die Wahrscheinlichkeit der Erkennung reduziert würde, im Gegensatz zum linearen Modus. Um die Unterschiede der beiden Erkennungsmechanismen zu demonstrieren, werden r_1 und r_f als Funktion von k_r/ν in Abbildung 4.13 verglichen.

4.2.3 Reaktionen bei Teilerkennung

Trotz der partiellen Erkennung bleiben die Grundeigenschaften des kooperativen Modells erhalten. Lediglich die Anteile der fehlerhaften Kopien auf den Y_0 - und Y_1 -Sequenzraum werden angepaßt. Dazu wird das System kurzzeitig auf ein drei Spezies Modell, nämlich

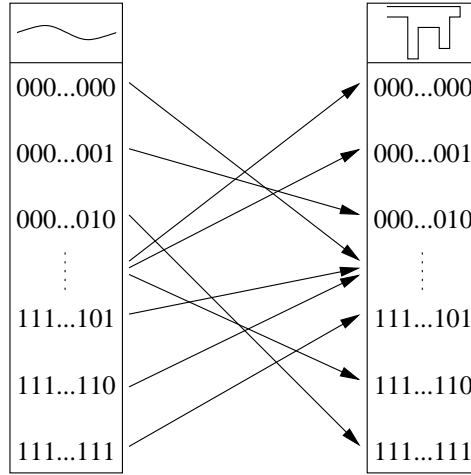
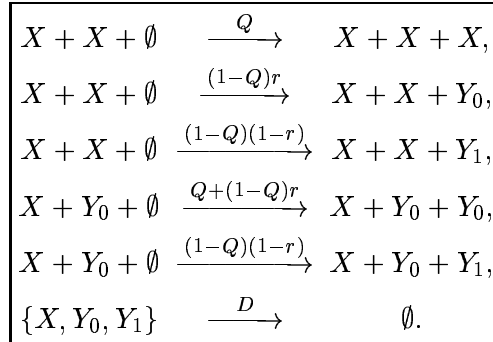


Abbildung 4.11: Die Abbildung zeigt einen Ausschnitt aus dem Sequenzraum (links) und aus dem Strukturraum (rechts). Da es bei einer bijektiven Abbildung gleichviele Strukturen wie Sequenzen gibt, macht es Sinn, diese ebenfalls per Binärstring durchzunummerieren. Die Pfeile verdeutlichen, daß jeder Sequenz zufällig genau eine Struktur zugeordnet wird. Würde eine entsprechende Abbildung einmal getroffen, gilt sie für alle im System befindlichen Sequenzen.

X , Y_0 und Y_1 erweitert, daß wie folgt zusammengefaßt werden kann:



Dabei wird dem Umstand Rechnung getragen, daß die Monomermutationswahrscheinlichkeit q für alle Sequenzen gleich ist und somit auch die Gesamtqualität $Q = q^n$ unabhängig von Typ der Sequenzen benutzt werden kann. Die Differentialgleichungen, die das Verhalten im homogenen Fall beschreiben, sind

$$\dot{x} = Qx^2(1 - x - y_0 - y_1) - Dx, \quad (4.22)$$

$$y_0 = ((1 - Q)rx^2 + (Q + (1 - Q)r)xy_0)(1 - x - y_0 - y_1) - Dy_0, \quad (4.23)$$

$$y_1 = (1 - Q)(1 - r)(x^2 + xy_0)(1 - x - y_0 - y_1) - Dy_1. \quad (4.24)$$

Aus der obigen Diskussion ist bereits klar, daß das Prinzip der Teilerkennung ($r > 0$) in homogenen Räumen nicht funktioniert, da Erkennung nur eine bereits vorhandene Stabilität gegen Mutation erhöht. Die Differentialgleichungen können jedoch genutzt werden, um eine einfache Konzentrationsabschätzung zwischen den beiden katalytisch inaktiven Spezies Y_0 und Y_1 zu treffen. Löst man nach den beiden Variablen y_0 und y_1 im (virtuellen) stationären Zustand auf, so erhält man für die Approximation weniger Mutationen ($Q \approx 1$) und einer

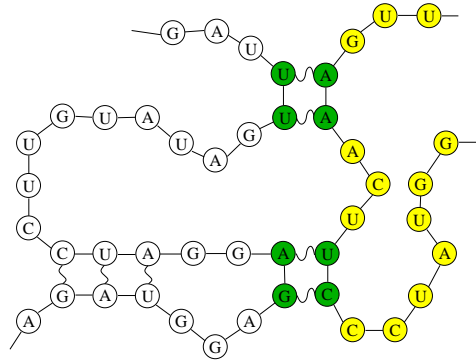


Abbildung 4.12: Die Abbildung zeigt Ausschnitte aus zwei gefalteten RNA-Sequenzen (weiß und hellgrau), die über vier Basen (dunkelgrau) gebunden werden (Wasserstoffbrückenbindungen). Man erkennt, wie aufgrund der Faltung Sequenzteile räumlich zusammen- bzw. auseinanderrücken können. Diese (modellerte) Korrelationsfreiheit von primärer und sekundärer Struktur kann genutzt werden, um die Effekte der partiellen Sequenzerkennung deutlich zu erhöhen.

hinreichend kleinen X -Population die Abschätzung

$$\frac{y_1}{y_0} \approx (1 - Q)(1 - r). \quad (4.25)$$

Sie besagt, daß aufgrund der ausbleibenden katalytischen Unterstützung durch X die Konzentration der Y_1 - im Vergleich zu den Y_0 -Sequenzen reduziert ist. Dies erlaubt, daß System wieder auf ein effektives zwei Spezies Modell zu reduzieren, bei dem die Population der Y_1 -Sequenzen vernachlässigt wird und somit Y_0 in Y übergeht.

4.2.4 Teilerkennung im Simplexmodell

Nachdem die Modellierung der zwei Extremfälle einer Teilerkennung diskutiert wurde, wird in diesem Abschnitt deren explizite Anwendung für das katalytische Modell im Rahmen des PRESS Systems behandelt.

Um erneut eine direkte Vergleichbarkeit mit vorherigen Ergebnissen zu sichern, werden die kritischen Q_c für jeweils zwei verschiedene Zerfallswerte $D = 10^{-5}$ und 10^{-2} benutzt, die entsprechend der zweiten und vierten Kurve aus Abb. 4.5 zugeordnet werden können. Für diese Parametersätze wird die Stabilität für $k_r = 1$ bis 8 analysiert. Dies entspricht bei $\nu = 20$ einer prozentualen Erkennungsregion von 5% bis 40%, relativ zur Gesamtlänge. Die entsprechenden Werte für die Erkennungswahrscheinlichkeiten der unterschiedlichen Modi können der Tabelle 4.3 entnommen werden. Da die Gesamtlänge ν und die Anzahl der Erkennungsbits k_r letztlich nur mit Hilfe von Gl. (4.20) und Gl. (4.21) die Erkennungswahrscheinlichkeiten definieren, können die Ergebnisse auch für andere ν und k_r extrapoliert werden. Abb. 4.14 zeigt die Ergebnisse für obige Werte, wobei der lineare und der Faltungsmodus gegenüber gestellt werden. Auffallend für alle gezeigten Kurven ist die Bestätigung der Hypothese, daß Teilerkennung eine gegebene Stabilität nur erhöhen, nicht aber direkt erzeugen kann, was durch die Unabhängigkeit der Migrationsraten m_a und m_b von den k_r belegt wird.

Die Gegenüberstellung von linearem und gefaltetem Modus läßt den Vorteil der letzten Methode deutlich werden. Ändert sich die Fehlerrate mit der Erweiterung der Erkennungsregion im linearen nur sehr langsam, so entspricht eine Teilerkennung von wenigen Bits im Faltungsmodus fast schon einer vollständigen Erkennung.

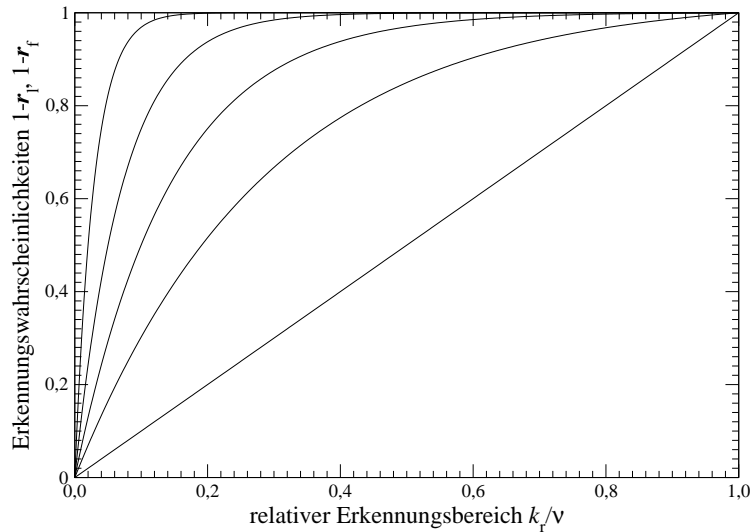


Abbildung 4.13: Es werden die Erkennungswahrscheinlichkeiten $1 - r_l$ und $1 - r_f$ für $\nu = 5, 10, 20, 50$ verglichen. In relativen Einheiten bleibt r_l konstant (Gerade), wohingegen die $1 - r_f$ umso schneller gegen 1 konvergieren, je länger die Gesamtsequenz ist. Die Zuordnung der Gesamtlänge geht von $\nu = 50$ (linke Kurve) absteigend bis $\nu = 5$ (rechte Kurve).

k_r	$1 - r_l$	$1 - r_f$	k_r	$1 - r_l$	$1 - r_f$
1	0.0500	0.5000	5	0.2500	0.9688
2	0.1000	0.7500	6	0.3000	0.9843
3	0.1500	0.8750	7	0.3500	0.9921
4	0.2000	0.9375	8	0.4000	0.9960

Tabelle 4.3: Gegenüberstellung der Erkennungswahrscheinlichkeit für den linearen (r_l) und den Faltungserkennungsmodus (r_f) für den in Abb. 4.14 dargestellten Fall von $k_r = 1, \dots, 8$. Zu beachten ist die unterschiedliche Skalierung der y-Achse.

4.2.5 Sequenzverlängerung

In der bisherigen Modellierung wurde davon ausgegangen, daß Teile der kodierenden Bereiche gleichzeitig auch als Erkennungsregion dienen, was im Hinblick auf die Relevanz der Faltung durchaus gerechtfertigt ist. Die bereits diskutierte Existenz von neutralen Netzen, die sich auch im Kontext der RNA-Sekundärstrukturen ergibt [71], gibt evolutiven Systemen die Möglichkeit, Sequenzen leicht zu modifizieren, ohne daß sich deren Funktionalität ändert. Diese Änderung kann entsprechend auch eine redundante Verlängerung der Sequenz sein. Dabei werden Monomere an oder in die Sequenz addiert, die keinen Einfluß auf den Charakter, also die Funktionalität des Gesamtmoleküls haben. Diese zusätzlichen Monomere könnten aber für eine Erkennung durchaus dienlich sein, die die Stabilität erhöhen würde. Durch eine Erweiterung der Sequenz wird jedoch auch immer der Gesamtqualitätsfaktor $Q = q^\nu$ aufgrund der ansteigenden Länge reduziert.

Es stellt sich die Frage, ob die Einfügung eines zusätzlichen Erkennungsbits mit Hilfe des partiellen Erkennungsmechanismus die maximal erträgliche Fehlerrate so erhöhen kann, daß die entstehende Qualitätseinbuße kompensiert oder eventuell sogar überkompensiert wird. Dazu sollen im folgenden die Maxima der für unterschiedliche k_r zu berechnenden Fehler-

kurven auf Sequenzlängen umgerechnet, wobei jedoch beachtet werden muß, daß durch das Anhängen von Erkennungsbits die Gesamtlänge entsprechend verändert wird ($\nu \rightarrow \nu + k_r$), was direkten Einfluß auf die Erkennungswahrscheinlichkeiten $1 - r_1$ und $1 - r_f$ hat. Bei nicht konstanter Sequenzlänge ändern sich Gl. (4.20) zu

$$r_1 = \frac{\nu}{\nu + k_r} \quad (4.26)$$

und Gl. (4.21) zu

$$r_f = \frac{2^\nu - 1}{2^{\nu+k_r} - 1}. \quad (4.27)$$

Abbildung 4.15 zeigt die kritischen Fehlerraten für den Fall, daß die Erkennungsbits aus einer entsprechenden Verlängerung stammen. Vor allem der lineare Modus zeigt eine merkliche Reduzierung der kritischen Fehlerrate im Vergleich zu den Werten der konstanten Sequenzlänge in Abb. 4.14. Um Rückschlüsse auf eine maximale Sequenzlänge ziehen zu können, kann die allgemeine Beziehung $R = 1 - Q = 1 - q^\nu$ zu

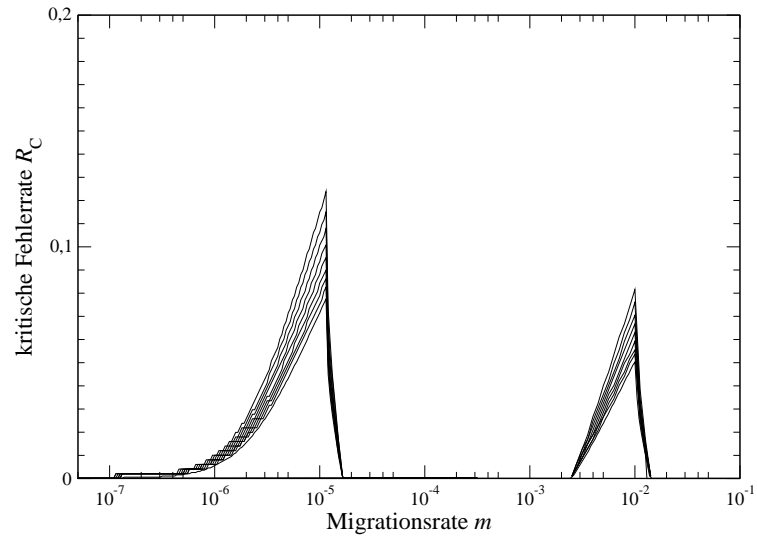
$$\nu_{k_r}^{(1/f)} = \frac{\ln(1 - R_{\max}^{(1/f, \nu+k_r)})}{\ln(q_{\min})}, \quad (4.28)$$

umgeformt werden, mit $q_{\min} = (1 - R_{\max}^{(\nu)})^{1/\nu}$, welches die minimale Monomerkopiergenauigkeit definiert, die das katalytische Modell ohne Erkennung für ν lange Sequenzen erlaubt. Es findet genau dann eine (Über)Kompensation statt, wenn die Bedingung $\nu_{k_r}^{(1/f)} \geq \nu + k_r$ erfüllt ist. Abbildung 4.16 zeigt die mit Hilfe von Gl. (4.28) aus $R_{\max}^{(1/f, \nu+k_r)}$ berechneten Sequenzlängen für die Grundsequenzlänge von $\nu = 20$. Der jeweilige Wert von $R_{\max}^{(1/f, \nu+k_r)}$ läßt sich wieder aus dem PRESS Modell berechnen, wobei diesmal jedoch die beiden Gl. (4.26) und Gl. (4.27) berücksichtigt werden müssen. Die Parametersätze korrespondieren erneut zu denen von Abb. 4.14 und Abb. 4.15.

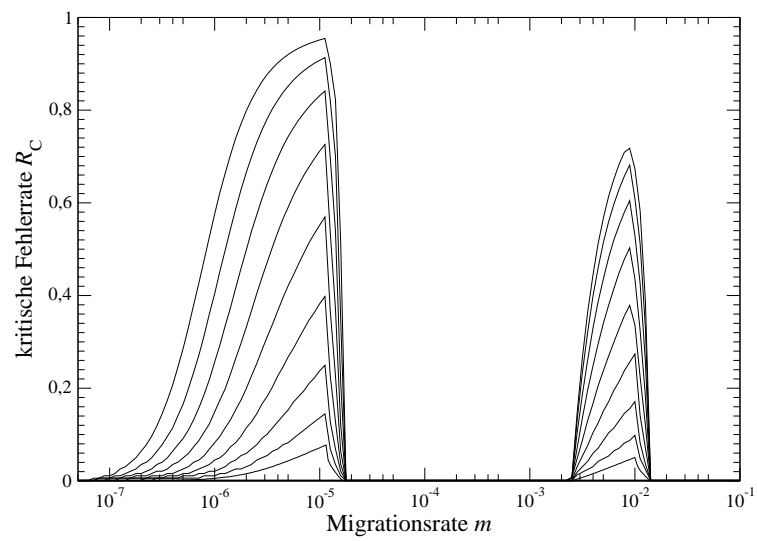
Im einzelnen zeigt sich, daß die lineare Erkennung in der Lage ist, die zusätzlichen Bits gerade zu kompensieren. Dies ist im Rahmen der Modellannahmen auch verständlich: Da die Y_1 -Spezies vernachlässigt wird, führen nur Mutation die nicht in der Erkennungsregion liegen zur Erhöhung der $Y_0 = Y$ -Population. Die Mutation in der Erkennungsregion würde hingegen einen Zufluß zu den Y_1 -Molekülen bedeuten, die jedoch als nicht-existent angenommen wurden und somit keinen Einfluß auf das System haben. Demzufolge stabilisieren die Erkennungsbits gerade sich selbst. Die Gesamtmutationsrate bleibt also konstant. Eine Überkompensation, die eine Sequenzverlängerung über die Erkennungsbits hinaus ermöglichen würde findet nicht statt. Da der störende Einfluß der Y_1 -Moleküle vernachlässigt wurde, ist eine Überkompensation in realen Systemen auch nicht zu erwarten.

Anders sieht die Situation im Fall der Faltung aus. Hier zeigt sich eine deutliche Erhöhung der tolerierbaren Gesamtlänge, die sich mit jedem zusätzlichen Erkennungsbit fast verdoppelt. Dies trägt dem Umstand Rechnung, daß r_f in guter Näherung als Sequenzlängenunabhängig angesehen werden kann. Die gewonnene Stabilität ist im Kontext der Faltung jedenfalls deutlich höher als im linearen Fall, so daß selbst in Modellen, die den vollständigen Sequenzraum berücksichtigen, eine Überkompensation zu erwarten ist. Das heißt, das eine kurz-kettige und schlecht katalysierende RNA-Sequenz aufgrund ihrer Faltungsfähigkeit im Rahmen einer partiellen Erkennung die Möglichkeit einer deutlichen Sequenzverlängerung hat. Die so zusätzlich mögliche Speicherung von Information kann im Laufe der Evolution dazu führen, die Qualität der Replikase zu erhöhen.

Damit soll das Kapitel über das einfache katalytische Modell abgeschlossen werden. Es konnte gezeigt werden, daß altruistische Replikatoren allein durch die Existenz eines separierenden Raumes innerhalb eines Migrationsbereichs stabilisiert werden können, völlig im Gegensatz zur homogenen Prognose. Dabei stellte sich weiter heraus, daß dieser Bereich insbesondere stetig von der Zerfallsrate abhängig ist, daß also langlebige Moleküle in Bereichen niedriger Diffusion und kurzlebige in höheren Diffusionsbereichen stabilisiert werden können. In einer weiteren Untersuchung zeigte sich partielle Sequenzerkennung als probates Mittel zur Erhöhung der durch den Raum induzierten Kooperation. Dort etablierte sich ebenfalls, daß die molekulare Sequenzfaltung diesen Effekt noch deutlich erhöhen kann, so daß Gesamtsequenzverlängerungen möglich sind mit oben beschriebenen Folgen für die Evolution.

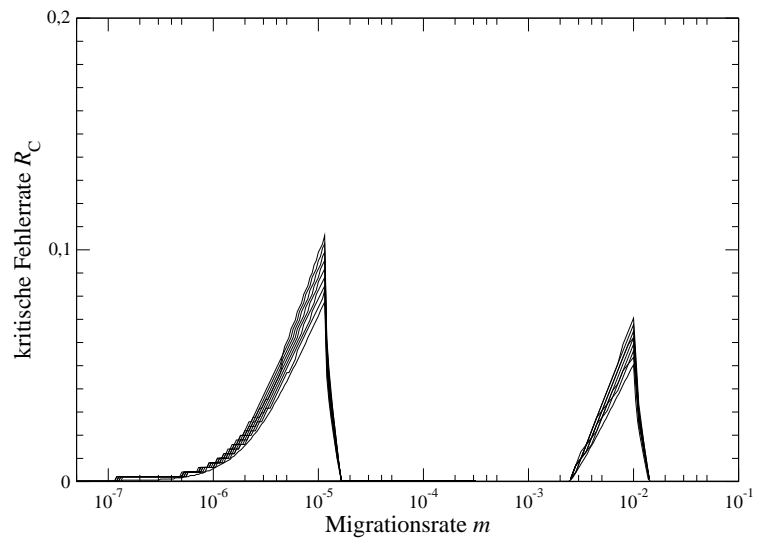


(a) linear

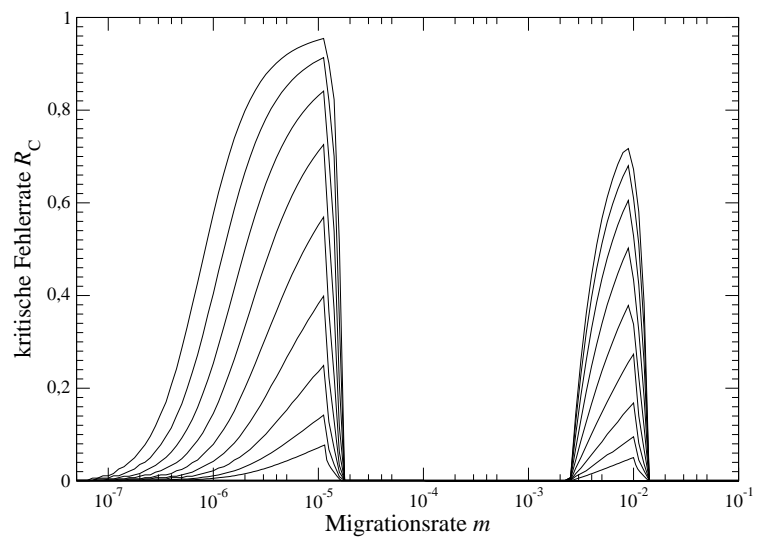


(b) gefaltet

Abbildung 4.14: Vergleich des Stabilitätszuwachs gemessen in $R_c(m)$ für $k_r = 0, \dots, 8$ (von unten nach oben) für $D = 10^{-5}$ und 10^{-2} als Funktion von m für linearen (a) und gefalteten (b) Erkennungsmodus. Dabei werden k_r Bits erkannt, ohne daß eine entsprechende Sequenzverlängerung ($\nu = 20$) stattfindet. Unterschiedliche Skalierung der y-Achse muß beachtet werden.



(a) linear



(b) gefaltet

Abbildung 4.15: Vergleich des Stabilitätswachstums gemessen in $R_c(m)$ für $k_r = 0, \dots, 8$ (von unten nach oben) für $D = 10^{-5}$ und 10^{-2} als Funktion von m für linearen (a) und gefalteten (b) Erkennungsmodus. Dabei stammen die Erkennungsbits aus einer zusätzlichen Sequenzverlängerung ($\nu \rightarrow 20 + k_r$). Unterschiedliche y-Achse zum linearen Fall beachten.

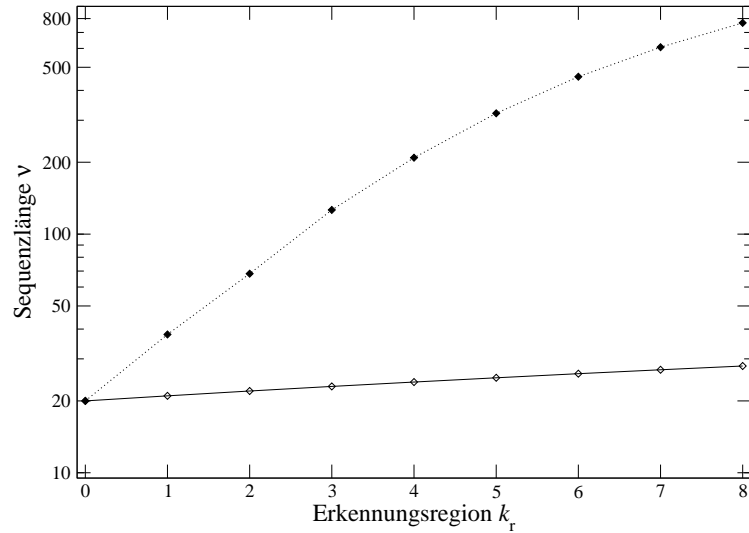
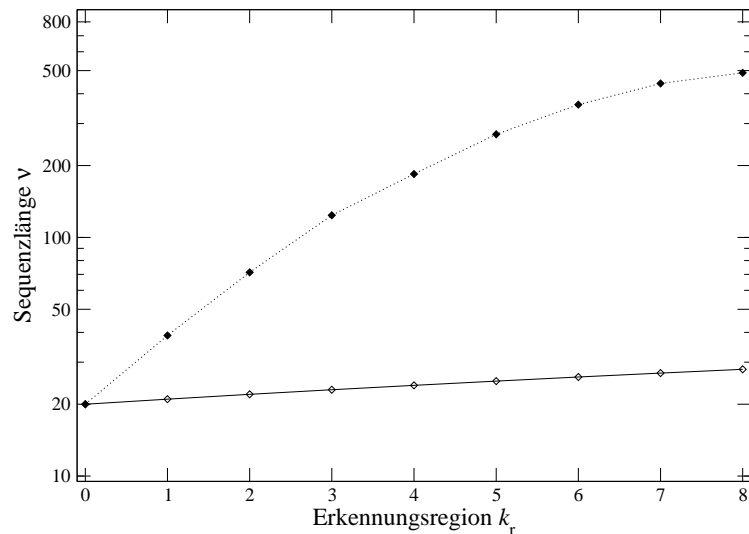
(a) $D = 0.00001$ (b) $D = 0.01$

Abbildung 4.16: Die Abbildung zeigt die tolerierbare Sequenzlänge bei der Addition von Erkennungsbits zu funktional-aktiven Sequenzen mit Grundlänge $\nu = 20$. Die durchgezogene Linie zeigt die jeweilige kritische Sequenzlänge. Die Symbole kennzeichnen die Datenpunkte, die sich bei k_r Erkennungsbits ergeben. Man erkennt, daß die lineare Erkennung nur in der Lage ist, die zusätzlichen Erkennungsbits zu kompensieren, wo hingegen eine deutliche Sequenzverlängerung bei der Faltungsmethode zu beobachten ist.

Kapitel 5

Simulation mit konfigurierbarer Hardware

5.1 Der Parallelrechner Ngen

Der Aufbau und die Funktionsweise des Parallelrechners NGEN werden erörtert, mit dessen Hilfe Reaktionen hochoptimiert simuliert werden können, was eine Untersuchung von molekularen Systemen auf sehr großen Zeitskalen erlaubt.

5.1.1 Konventionelle Rechnerarchitekturen

Die Simulation von räumlich aufgelösten biomolekularen Systemen ist mitunter ein extrem zeitintensives und ressourcenaufwendiges Problem. Je nach Komplexität können solche Simulationen schnell die Grenzen moderner Computer erreichen. Dabei ist es nicht einmal so, daß deren Leistung prinzipiell nicht ausreichend wäre - daß eigentliche Problem stellt ein enormer Overhead dar, der neben der eigentlichen Aufgabe, nämlich der puren Simulation, ebenfalls verwaltet wird. So müssen zum Einen während der ganzen Zeit Interrupts gesteuert, eventuelle Peripherie verwaltet und Betriebssysteme aufrecht erhalten werden.

Noch ineffizienter erweist sich jedoch der eigentliche Aufwand, der für die Simulation notwendig ist. Der meist in einer Hochsprache geschriebene Simulations-Quellcode wird zwar von modernen Compilern sehr effizient übersetzt, doch die für die einzelnen Programmteile notwendigen Algorithmen benötigen im allgemeinen sehr viele Prozessorentakte. Als Beispiel sei hier die Erzeugung von Zufallszahlen angeführt, die in stochastischen Simulationen *en masse* anfallen (zum Beispiel bei der Bestimmung von Reaktionen oder bei der Diffusion), jedoch einige 100 Takte für deren Berechnung notwendig sind.

Weitere Problemen dieser Art entstehen dadurch, daß der Zustandsraum für Reaktions-Diffusions Prozesse (definiert durch die zeitliche Änderung räumlich verteilter Moleküle) nur sehr schwierig und mit großem Kodierungsaufwand auf die Beschreibungsebene eines einzigen, lokalen Prozessors abgebildet werden können. Exemplarisch sei hier auf die Schwierigkeit einer effektiven Implementierung der Diffusion hingewiesen. Die parallele Bewegung der Moleküle, die neben den Reaktionen zu gewährleisten ist, bedarf eines riesigen Rechenaufwandes, da intern die unkorrelierte Bewegung jedes einzelnen Moleküls stochastisch simuliert werden muß. Dies wird umso drastischer, je größer diese Menge wird, da ab einer gewissen Grenze die Daten nicht mehr im schnellen Cache Speicher der CPU abgelegt werden können, sondern

über den im Vergleich dazu sehr langsamen Datenbus geschleust werden müssen. Anhand dieses Beispiels, bei dem der noch viel größere Aufwand der einzelnen Reaktionen außer acht gelassen wurde, wird deutlich, wie die Kodierung und Rekodierung, die mit den inkompatiblen Beschreibungsebenen der elektronischen Hardware und der molekularen Prozesse einhergehen, eine ineffektive Nutzung von Rechnerressourcen zur Folge hat.

5.1.2 FPGA basierende Architektur

Um die angegebenen Probleme zu verringern, muß demnach ein entsprechender Computer designed werden, der von seiner grundlegendsten Architektur bereits den Anforderungen molekularbiologischer Simulationen entspricht. Mit anderen Worten, es muß eine Architektur manifestiert werden, die bereits deutliche Ähnlichkeiten zu realen, molekularbiologischen Systemen aufweist, da nur so der beschriebene Overhead deutlich reduziert werden kann.

Führt man sich die Prozesse eines realen RNA/DNA *in vitro* Experiments vor Augen, so erkennt man einige generelle Eigenschaften, die solche Systeme mit sich bringen:

- Massive Parallelität der Reaktionen
- Räumliche Struktur der Reaktionsumgebung
- Vielfalt an möglichen Reaktionen
- Sequenzartigkeit der Moleküle

Einen Lösungsvorschlag, der die dargestellte Problematik zu umgehen wußte, wurde 1991 von Prof. John S. McCaskill und seinen Mitarbeitern vorgestellt. Diese präsentierten einen auf FPGA (Field Programmable Gate Arrays) basierten Minirechner, der aus vier dieser Chips konzipiert wurde. Anstatt die Idee eines Ein-Prozessoren-Rechners weiter zu verfolgen und diesen in Geschwindigkeit und Kapazität zu verbessern, gingen John McCaskill *et al.* den Weg räumlich verteilter Prozessoren, die massiv parallel arbeiten. Die programmierbaren Prozessoren, die in die von den FPGAs zu Verfügung gestellten konfigurierbaren Hardware eingebettet werden, können bezüglich der molekularen Reaktionen hochoptimiert werden. Übrige Ressourcen werden zur Speicherung von Bitstrings - den Molekülen - bereit gestellt.

Dieser Rechner ist somit aufgrund seiner Architektur bereits speziell auf die Gegebenheiten räumlich strukturierter, biomolekularer Systeme zugeschnitten. Moleküle, die als Daten auf der für Prozessoren grundlegendsten Ebene kodiert sind, entsprechen außerdem in ihrem Charakter der Vorstellung von RNA bzw. DNA ähnlichen Strukturen. Der Fluß der Daten von Prozessor zu Prozessor wird so arrangiert, daß eine einfache Implementierung von Diffusionsprozessen gewährleistet wird. Die Raumdimension, in der die Simulation entsprechender Chemien stattfinden soll, kann durch die Veränderung der Verbindungstopologie der Prozessoren bestimmt werden. Jeder Einzelne dieser Prozessoren muß trotzdem so flexibel ausgelegt sein, daß praktisch jede Chemie simuliert werden kann. Um auch ohne deutliche Erweiterung der FPGA Ressourcen mehr Daten speichern und verarbeiten zu können, kam man 1992 auf die Idee, an einer Ebene aus FPGAs schnelle Speicherbausteine, SRAMs, zu koppeln, die im Vergleich zu den FPGAs zum einen deutlich preiswerter, aber zur reinen Datenspeicherung auch sehr viel besser geeignet waren, da ihre Speicherkapazität fast beliebig gewählt werden konnte. Diese Zuteilung erlaubte es, die Ressourcen der FPGAs fast ausschließlich für die Programmierung der Reaktionsprozessoren aufzuwenden und in den SRAMs riesige Mengen

an Zeichenketten zu speichern.

Ende 1992 wurde ein Rechner namens NGEN präsentiert, bei dem das beschriebene Konzept etabliert wurde [72]. Seit kurzem ist auch das Nachfolgemodell Meregen funktionsfähig und einige kleinere Applikationen wurden bereits gestartet. Da die prinzipielle Architektur mit der von NGEN fast identisch ist, können viele der hier vorgestellten Prinzipien auch auf Meregen angewandt werden. Am Ende von Anhang A.14 wird eine Abschätzung über die zu erwartende Rechenleistung gegeben.

5.1.3 Hardwarekomponenten

Die Hardware der NGEN besteht prinzipiell aus 162 FPGA Chips und 1296 32kByte static RAM (SRAM), was insgesamt 40.5MByte entspricht. Die 162 FPGAs setzen sich aus 144 frei programmierbaren Rechen- und 18 Steuer-FPGAs zusammen. Diese wiederum sind so angebracht, daß 8 Rechen- und 1 Steuer-FPGA eine Karte bilden, was 18 Karten bzw. 9 Kartenpaare ergibt. Alle FPGAs sind untereinander mit 8Bit Datenleitungen verbunden, die auch vier der insgesamt acht zugeordneten 32kByte SRAMs ansteuern. Den Zugriff auf die restlichen vier RAM-Bausteine erfolgt exklusiv durch die einzelnen FPGAs. Ein FPGA vom benutzten Typ besteht aus einer Matrix von (18×18) konfigurierbaren logischen Blöcken (CLB), wobei jedes einzelne CLB prinzipiell aus zwei F-Funktionsgeneratoren, einem nachfolgendem, koordinierendem H-Funktionsgenerator und aus zwei Flip-Flops besteht, vergl. Abb. 5.1. Die F-Funktionsgeneratoren sind frei programmierbar und besitzen vier Eingänge, so daß jede vierwertige boolesche Funktion simuliert werden kann. Daraus ergeben sich die Hardwareressourcen, die zur Implementierung für eine Reaktions-Diffusions Simulation benutzt werden können. Da eine Programmierung auf der untersten Ebene der CLBs in der gegebenen Größenordnung praktisch unmöglich ist, werden entsprechende Softwarepakete zur Entwicklung und zur Übersetzung in die Sprache der CLBs zur Verfügung gestellt. Diese Tools erlauben eine komfortable (makroorientierte) Programmierung in der Logik von digitalen Schaltung. Nach einem Test auf logische Konsistenz werden die Schaltungen von einem Compiler in einzelne (den FPGA Ressourcen angepaßten) Komponenten zerlegt und anschließend miteinander verbunden (gerouted). Gerade der letzte Punkt ist kritisch, da zum einem das Verbinden der Komponenten selbst Ressourcen beansprucht und zum anderen das Finden einer (sub)optimalen Struktur ein sehr komplexes Problem darstellt.

Ist eine Schaltung gefunden, die die gegebenen Konditionen erfüllt, können die Schaltungskomponenten als entsprechendes Bitmuster in die CLBs geschrieben und diese wiederum entsprechend verbunden werden. Danach ist die NGEN mit dem entsprechenden Programm ausgestattet und kann dieses ausführen.

5.1.4 Datenflußarchitektur

Bei der Erstellung der NGEN Programme müssen jedoch noch mehr als die oben beschriebenen Bedingungen beachtet werden. So bedingt der Zusammenschluß von Programmier-elementen und reinen Speicherbausteinen eine besondere Art der Datenorganisation.

Gerade für das hier beschriebene Ziel einer dreidimensionalen Reaktions-Diffusions Umgebung, bietet sich das Konzept des zyklischen Datenfluß an. Dabei werden die 144 FPGAs so angeordnet, daß sie ein rechtwinkliges (in x-y-Richtung) (12×12) Array darstellen, bei dem die gegenüberliegenden Seiten toroidal miteinander verbunden sind. Auf diesem Array ste-

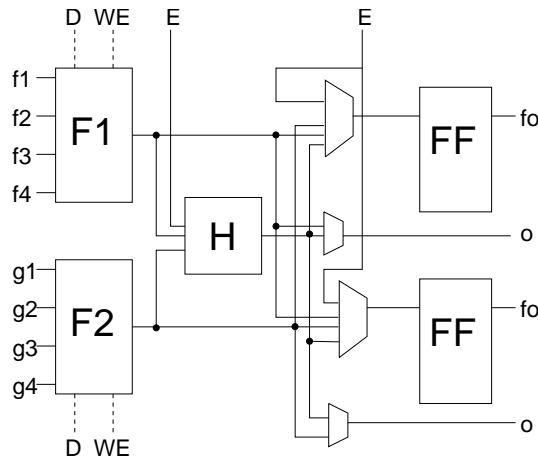


Abbildung 5.1: Die Struktur eines konfigurierbaren logischen Blocks (CLB). Die beiden F-Funktionsgeneratoren können jede vierwertige boolesche Funktion simulieren, die nachfolgenden Flip-Flops dienen als lokaler Speicher.

(Mit freundlicher Genehmigung von Dr. Jens Breyer.)

hen orthogonal (in z-Richtung) die SRAM Speicherbausteine. Stellt man sich anschaulich die Speicherkapazitäten für die benutzten Bitstrings einer bestimmten Länge als einen Einheitskubus vor, so können die Instanzen der Datenspeicherung und -verarbeitung innerhalb der NGEN wie in Abb. 5.2 dargestellt werden. Dort wird anhand des Beispiels eines (2×2) Arrays - anstatt der tatsächlich vorhandenen (12×12) FPGAs - der Zusammenschluß der SRAMs (weiß), und der FPGAs (hell grau/grau) gezeigt. Dabei wird die dreistöckige FPGA-Ebene anschaulich in die Prozessorebene (hell grau) und der zweistöckigen Diffusionsebene aufgeteilt. Die Anzahl der möglichen parallel verarbeitbaren Daten (Anzahl der Kubensäulen) pro FPGA - im Beispiel (4×4) - hängt prinzipiell von der Komplexität der verwendeten Logik ab. Je komplexer die Schaltungen, desto höher ist der zur Realisation notwendige Hardwareaufwand, was mehr Ressourcen beansprucht und die Zahl der möglichen parallelen Prozesse verkleinert. Da die Datenverarbeitung nur in den FPGA Ebenen möglich ist - die nebenbei auch die zur Verarbeitung notwendigen Speicherkapazität zur Verfügung stellen - werden die in den SRAMs gespeicherten Daten mit Hilfe zyklischer Adresszähler in z-Richtung durch die FPGAs geschleust. Zwischen der Prozessoren- und der Diffusionsebene erfahren sie zusätzlich noch eine Verschiebung in x-y-Richtung, siehe Abb. 5.2. Bei der in Kapitel 5.2.1 gegebenen Beschreibung des dreidimensionalen Diffusionsalgorithmus wird diese Verschiebung nochmals wichtig.

Es muß noch erwähnt werden, daß die Datenflußarchitektur eine Relation zwischen Volumen (Gesamtzahl der Datenelemente) und Simulationszeit (Zeit für eine vollständige Volumenberechnung) liefert. Wird das Volumen durch einen symmetrischen Kubus mit Kantenlänge l aufgespannt, so werden die Daten - aufgrund der parallelen Verarbeitung in der Ebene - nur in einer Richtung seriell evaluiert. Mit $l = \sqrt[3]{V}$ ergibt sich so die Beziehung

$$t \propto \sqrt[3]{V}. \quad (5.1)$$

Dabei muß natürlich beachtet werden, daß die Ressourcen der NGEN nicht immer eine exakte Symmetrisierung des Volumens erlauben, so daß Gl. (5.1) nur als Approximation angesehen werden kann.

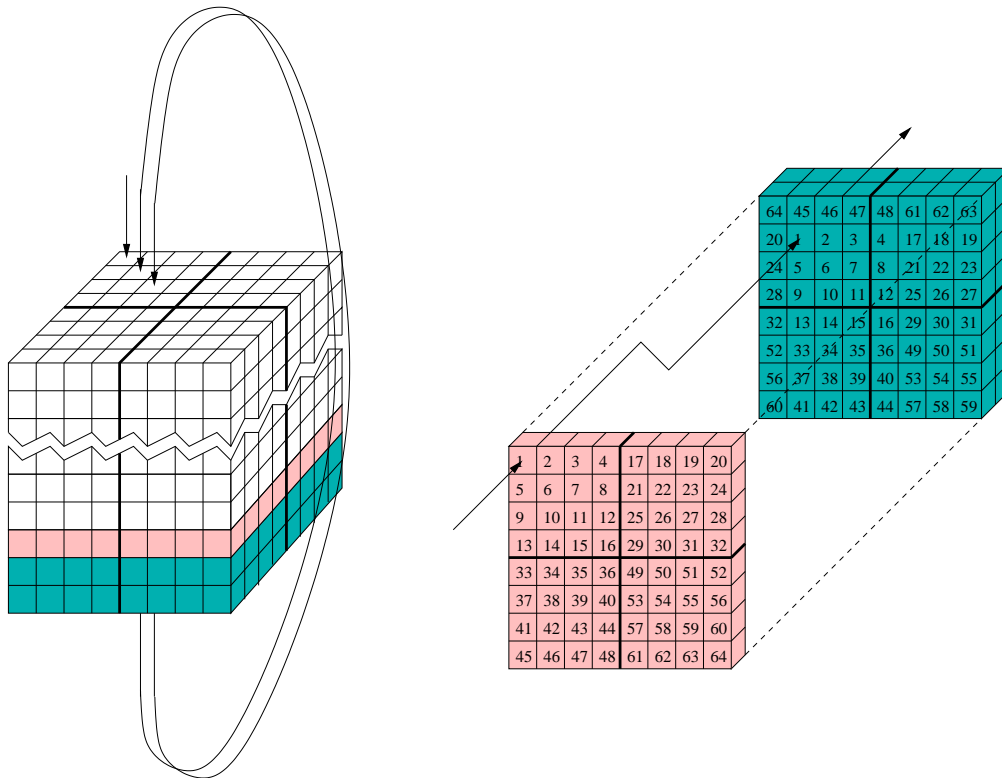


Abbildung 5.2: Der beschriebene Datenfluß durch das Volumen der NGEN soll bildlich dargestellt werden. Dabei beschreibt der linke Teil des Bildes einen (2×2) Ausschnitt aus den vorhandenen (12×12) FPGAs. Jedes FPGA ist dabei nochmals in ein (4×4) Grid unterteilt (Kubensäule), in dem parallel Daten verarbeitet werden können. Die Anzahl der Kubensäulen pro FPGA ist von der Komplexität der Logik abhängig. Die z-Achse beschreibt (von unten nach oben) die Diffusions- und die Reaktionsebene sowie die SRAMS, die der reinen Datenspeicherung dienen. Der rechte Bildausschnitt verdeutlicht sowohl den Fluß in z-, als auch in x-y-Richtung, was für die elegante Implementierung des Diffusionsalgorithmus nötig ist.

Für eine umfangreichere Beschreibung der NGEN soll die Arbeit von Breyer erwähnt werden [73], in der ein CATCH-ähnliches [69] Amplifikationssystem in zwei Dimensionen simuliert wurde. Für weiterführende Informationen zum Aufbau einer dreidimensionalen Topologie siehe [74] und [75].

5.2 Parallele, dreidimensionale Diffusion

Ein in drei Dimensionen parallel arbeitender Diffusionsalgorithmus wird vorgestellt, der die Diffusionsgleichung erfüllt.

5.2.1 Algorithmische Struktur

Die Möglichkeit, in drei Dimensionen Parallelverarbeitung zu betreiben, bedarf eines effektiven Diffusionsalgorithmus, welcher die Forderungen erfüllt, die paralleles Rechnen mit sich bringt. Für ein Verständnis der folgenden Prozeduren empfiehlt sich die Lektüre von Anhang A.12, in dem der zweidimensionale Toffoli-Margolus Algorithmus [76] vorgestellt wird.

Der kleinste (symmetrische) Kubus, der mehr als aus einer Elementarzelle besteht ist der $(2 \times 2 \times 2)$ Würfel, mit der ein Volumen ausgefüllt werden soll. Zwei unterschiedliche Füllmethoden unterscheiden sich insofern, daß die $(2 \times 2 \times 2)$ Würfel um je einen Gitterpunkt in x-, y- und z-Richtung verschoben werden. Dabei kann die Datenflußarchitektur der NGEN benutzt werden, um ohne größeren Aufwand die z-Verschiebung zu etablieren.

Die Würfel haben die Möglichkeit, sich um jeweils 90° um die x-, y-, oder z-Achse mit der Wahrscheinlichkeit $p_x = p_y = p_z = p$ zu drehen, wobei schon drei der sechs möglichen Rotationen ausreichend sind, um die Diffusion zu etablieren. Die Umkehroperationen zu den einzelnen Rotationen sind also nicht erforderlich. Der Parameter p kann benutzt werden, um die Diffusionskonstante des Algorithmus zu justieren. Die Rotation der Kuben geschieht zeitgleich und unkorreliert. Die vier möglichen Operationen werden in Abbildung 5.3 verdeutlicht. Anschließend wird das Volumen durch die alternative Füllmethode überdeckt und die Kuben erneut rotiert. Dieses algorithmische Verfahren erlaubt die parallele Diffusion des gesamten Volumens.

5.2.2 Erfüllung der Diffusionsgleichung

Wie die folgenden Untersuchungen zeigen werden, erfüllt der Algorithmus die Diffusionsgleichung

$$\frac{\partial}{\partial t} P(\mathbf{r}, t) = D \Delta P(\mathbf{r}, t), \quad (5.2)$$

wobei Δ den Laplace-Operator darstellt und D die Diffusionsrate (im Gegensatz zur sonst verwendeten PRESS Migrationsrate m). Die Fundamentallösung von Gl. (5.2) ist eine Gauß-Verteilung (Normalverteilung),

$$P(\mathbf{r}, t) = \frac{1}{(4\pi D(t - t_0))^{d/2}} \exp\left(-\frac{(\mathbf{r} - \mathbf{r}_0)^2}{4D(t - t_0)}\right). \quad (5.3)$$

Dabei gibt d die geometrische Dimension des Systems an; im gegebenen Fall also $d = 3$. Gl. (5.2) erlaubt die Faktorisierung ihrer Lösung, so daß

$$P(\mathbf{r}, t) = P(x, y, z, t) = P(x, t) P(y, t) P(z, t) \quad (5.4)$$

geschrieben werden kann, was anschaulich bedeutet, daß keinerlei Korrelationen zwischen den Teilchenbewegungen in den drei räumlichen Dimensionen besteht.

Um den Bezug zwischen dem dreidimensionalen Raum (der wieder durch ein orthogonales, diskretes Gitter mit äquidistanten Elementarzellen aufgespannt ist) und dem Diffusionsalgorithmus herzustellen, ist es von Interesse, das Verhalten eines einzelnen Kubus zu untersuchen. Dabei sollen seine 8 Elementarzellen wie in Abb. 5.4 numeriert werden, wobei ohne

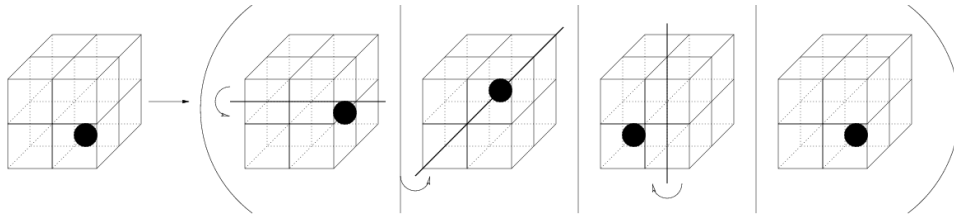


Abbildung 5.3: Die vier Operationen des dreidimensionalen Diffusionsalgorithmus, die anhand der Position eines einzelnen Teilchens erklärt werden. Die letzte Operation ist die Einheitsoperation, die keine Veränderung der Teilchenposition mit sich bringt und mit der Wahrscheinlichkeit $1-p$ stattfindet.

Beschränkung der Allgemeinheit solch eine Volumenfüllung benutzt werden soll, daß Zelle 5 eines gegebenen Kubus zum Zeitpunkt $t = 0$ im Ursprung des Koordinatensystems liegt. Im Anhang A.13 zeigt Tabelle A.1 die resultierende Koordinatenzuordnung zu den einzelnen acht Zellen. Da die Numerierung für alle Kuben gleich ist, und sie die Position eines Teilchens innerhalb eines Kubus eindeutig definiert, soll der Begriff der internen Koordinate oder des internen Zustands eingeführt werden. Die Rotation eines Kubus um eine der drei Hauptachsen hat eine Änderung des internen Zustands zur Folge und somit auch eine Änderung der absoluten Koordinaten. Der Zusammenhang zwischen den drei Rotationen und den dazugehörigen Koordinatentransformationen ist im Detail im Anhang A.13, Tabelle A.3 zu sehen. Der Vorteil des Gebrauchs des internen Zustands zeigt sich auch bei der Beschreibung der alternierenden Volumenfüllung. Die im Anhang A.13 gezeigte Tabelle A.2 veranschaulicht, wie die internen Zustände wechseln, ohne dabei jedoch eine Koordinatenverschiebung mit sich zu ziehen, da die virtuelle Umstrukturierung keine Veränderung der realen Teilchenposition zur Folge hat. Die Benutzung der obigen Zuordnungen erlaubt eine effiziente Untersuchung des Algorithmus. Um dies mit einer hinreichend genauen Statistik durchzuführen, wird für jeden Parametersatz (Zeit und Rotationswahrscheinlichkeit) über eine Million Trajektorien gemittelt, die jeweils vom Ursprung starten und zu einem Gitterpunkt führen, der durch die externen Koordinaten definiert ist. Die Simulationszeit - die diskret ist - wird in Einheiten eines vollständigen Updates, also einer parallelen Rotationsoperation gemessen. Die Umgruppierung der Kuben benötigt aufgrund des virtuellen Charakters keine Zeit und wird als instantan modelliert. Anschließend werden - separat in x , y und z Richtung - alle Trajektorien, die zum gleichen Gitterpunkt führen zusammengezählt und durch die Anzahl aller Trajektorien - im gegebenen Fall eine Million - dividiert. Das daraus resultierende normierte Histogramm beschreibt die Aufenthaltswahrscheinlichkeit eines Teilchens nach der Zeit t . Abbildung 5.5 zeigt so ein zeitliches Verhalten für $p = 0.3$. Ist $p_x = p_y = p_z = p$, wird also symmetrisch rotiert, zeigen sich keine Unterschiede der Histogramme für die drei Hauptachsen. Die Diffusion erfolgt also vollständig symmetrisch und isotrop. Allerdings ist es möglich, durch eine Änderung der einzelnen Rotationswahrscheinlichkeiten eine gezielte Asymmetrie zu manifestieren.

Die aus numerischen Berechnungen gewonnenen Datensätze erlauben im weiteren, die Diffusionskonstante in Abhängigkeit von p zu bestimmen. Dazu wird das FWHM (Full-Width-Half-Maximum) der Gauß-Kurven für ausgesuchte Zeiten und Wahrscheinlichkeiten ermittelt. Das Kriterium eignet sich deshalb als Kenngröße, da es zum einen nur die Meßpunkte berücksichtigt, die eine hohe statistische Sicherheit haben (hohe Ereignissrate in der Nähe des Maximums) und zudem eine Kreuzkorrelation zwischen diesen (nämlich Höhe und Breite) mit sich bringt, was die statistische Sicherheit weiter erhöht. Da bei einheitlichem p keine

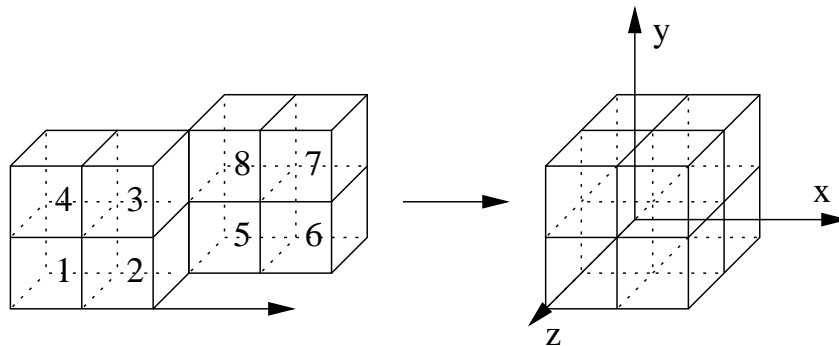


Abbildung 5.4: Der rechte Bildteil zeigt einen auseinandergeschobenen Kubus, auf dem eine (beliebige) Numerierung der Elementarzellen zu sehen ist. Zusammengeschoben füllen eine Vielzahl solcher Kuben das Gittervolumen aus. Für die weiteren Untersuchungen soll so ein einzelner Kubus benutzt werden, dessen 5. Zelle zu Beginn der Simulation auf dem Ursprung des dreidimensionalen Gitters liegt.

Asymmetrie in den Dimensionen zu sehen sind, reicht eine eindimensionale Untersuchung der Ergebnisse aus. Für eine Normalverteilung definiert sich das FWHM durch

$$b_{\text{FWHM}} = 4\sqrt{D \ln 2} \sqrt{t}, \quad (5.5)$$

wobei - wie erwähnt - die Dimension d explizit auf eins und t_0 auf Null gesetzt wurde.

Gl. (5.5) sagt - unter der Voraussetzung, daß das normierte Histogramm einer Normalverteilung entspricht - eine lineare Beziehung zwischen b_{FWHM} und \sqrt{t} voraus. Die tatsächlich hervorragende Übereinstimmung, die in Abb. 5.6 zu sehen ist, zeigt somit zumindest prinzipiell, daß der Algorithmus die Diffusionsgleichung erfüllt. Die Auswertung der FWHM-Beziehung in Abhängigkeit von p erlaubt die Bestimmung des Diffusionskoeffizienten (ebenfalls in Abhängigkeit von p). Ein Plot dieser Beziehung, siehe Abb. 5.7, zeigt eine erneute lineare Abhängigkeit, die mit Hilfe einer weiteren Regressionsanalyse verifiziert werden kann. Diese liefert die Beziehung

$$D = 0.520(\pm 0.002) p, \quad (5.6)$$

wobei D in Einheiten der Einheitslänge (Gitterkonstante) und der Simulationszeit angegeben wird. Die lineare Abhängigkeit ist insofern verständlich, als daß die Diffusionsaktivität proportional zur Rotationswahrscheinlichkeit ist und somit auch den Diffusionskoeffizienten linear verändern sollte. Der theoretisch maximale Koeffizient von $D \approx 0.17$ (bei $p_{\text{max}} = 1/3$) kann in der Praxis durch hintereinanderschalten vieler Diffusionsprozesse beliebig erhöht werden.

Schließlich kann aufgrund der Ergebnisse, die zwar keinen Beweis im strengen Sinne, jedoch ausreichende Evidenz präsentieren, behauptet werden, daß der Algorithmus ein effizientes Werkzeug ist, um parallele und effiziente Diffusion zu simulieren. Im Vergleich zum konventionellen Teilchenaustausch, erweist sich der Algorithmus in der Praxis zudem als deutlich schneller, da er mit einer erheblich reduzierten Anzahl an Zufallszahlen auskommt. Somit sollte die Benutzung der parallelen Variante für Simulationen in drei Dimensionen auch dann erwägt werden, wenn die übrigen Berechnungen nicht parallel ausgeführt werden. Im Anhang A.14 wird abschließend gezeigt, wie die Algorithmen in die Hardware der NGEN implementiert werden.

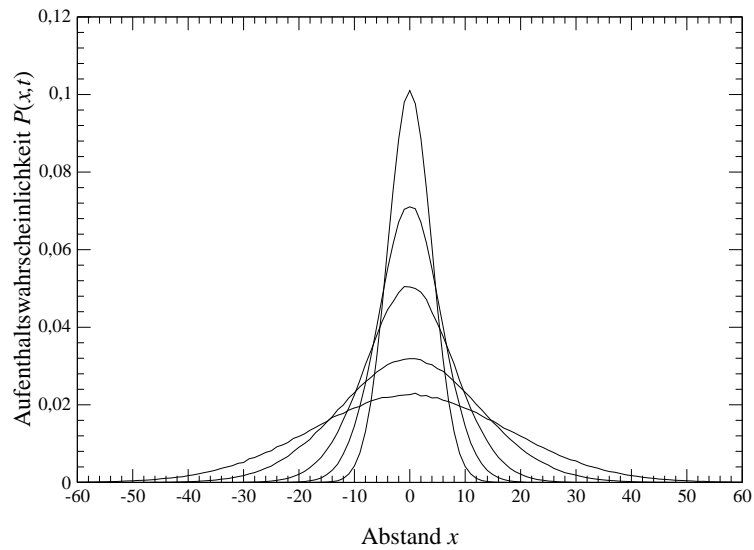


Abbildung 5.5: Die normierten Histogramme zur Rotationswahrscheinlichkeit $p = 0.3$ zeigen die gewünschte Form einer Normalverteilung. Auch der zeitliche Verlauf (für $t = 50, 100, 200, 500, 1000$) bestätigt die Vermutung.

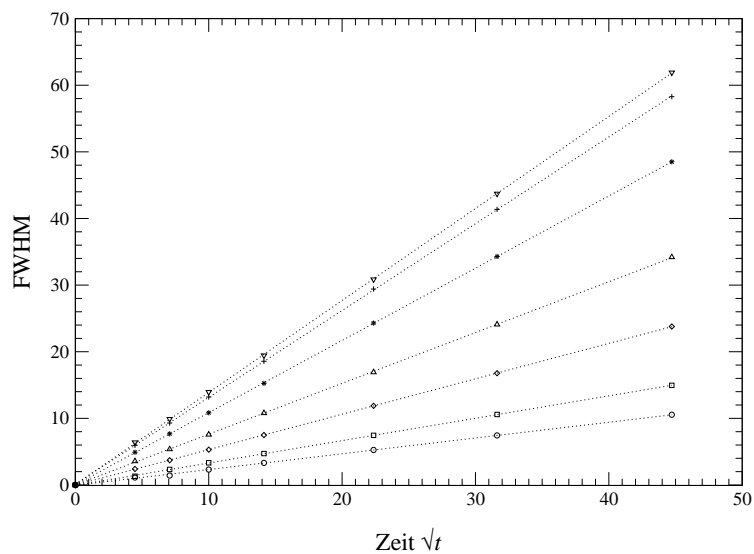


Abbildung 5.6: Das FWHM in Abhängigkeit der Wurzel der Simulationszeit, \sqrt{t} , wird für unterschiedliche Rotationswahrscheinlichkeiten, $p = 0.01, 0.02, 0.05, 0.1, 0.2, 0.3, 1/3$ (von unten nach oben) aufgetragen. Die theoretisch vorhergesagte lineare Beziehung wird - bis auf Fehler im Promille Bereich - bestätigt.

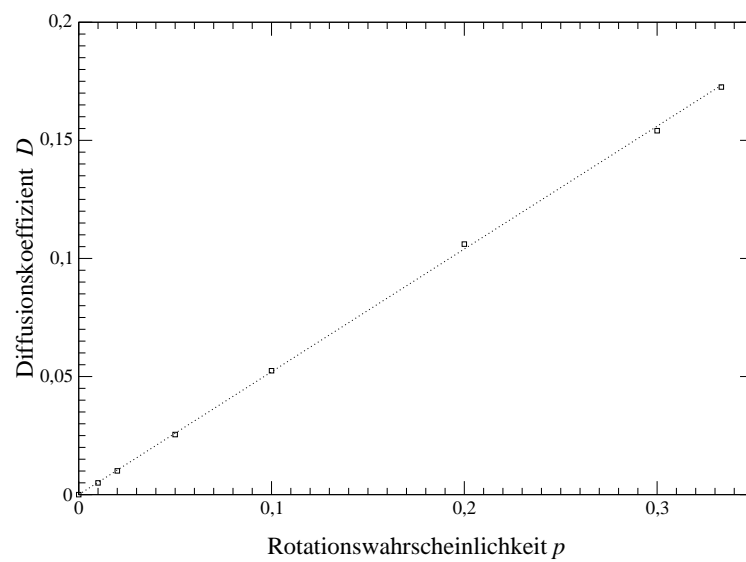


Abbildung 5.7: Die Bestimmung des Diffusionskoeffizienten D aus den mit Hilfe der linearen Regression gewonnenen Daten, zeigt eine lineare Abhängigkeit zu den und Rotationswahrscheinlichkeiten p auf.

Kapitel 6

Diskussion und Ausblick

6.1 Diskussion der Ergebnisse

Die Ergebnisse der Arbeit werden zusammengefaßt und vor dem Hintergrund möglicher genereller Erweiterungen diskutiert.

6.1.1 Zusammenfassung und Schlußfolgerung

Die Arbeit hat sich mit der Frage beschäftigt, inwiefern der Einfluß räumlicher Trennung sich auf die biomolekulare Evolution auswirkt und ob bzw. inwiefern damit molekulare Kooperation induziert werden kann. Gerade die Frage nach Stabilisierung von molekularer Kooperation hat auch innerhalb der Arbeitsgruppe von McCaskill einen besonderen Stellenwert erlangt und wurde intensiv untersucht. So kann man einen Bogen von mittlerweile über einer Dekade spannen. Dabei wurden Fragen gestellt und mögliche Lösungsvorschläge gegeben, die sich auf theoretischer Seite mit Parasitenabwehr durch Selbstreplizierenden Spots [30] und auch mit der Simulation von evolvierbaren, molekularen Systemen auf konfigurierbarer Hardware [77, 73] befaßten. Gleichzeitig konnten auch theoretische Modelle *in vitro* umgesetzt werden, wie das CATCH-System [69] oder weitere Veröffentlichungen zur Musterbildung [29].

Wie beschrieben, versucht die vorliegende Arbeit ebenfalls die Prinzipien zur Stabilisierung molekularer Kooperation im Raum zu erforschen, doch steht hier deutlich der analytische Aspekt im Vordergrund, um eine Allgemeingültigkeit der Ergebnisse zu erzielen. Dafür wurde eine Plattform konzipiert - erst im Kontext einfacher Systeme [56], später in allgemeiner Form [61], die es nicht nur erlaubt, den räumlichen Einfluß auf beliebige biomolekulare Systeme analytisch zu beschreiben, sondern darüber hinaus auch die diskrete Natur molekularer Reaktionen berücksichtigt. Die Arbeiten von Durrett und Levin haben in diesem Zusammenhang gezeigt, daß besonders für den Fall niedriger Konzentrationen eine Modellierung räumlicher Effekte mit partiellen Differentialgleichungen falsche Ergebnisse liefern kann [32]. Führt man sich jedoch die experimentellen Konditionen vor Augen, so erkennt man die Notwendigkeit von sehr geringen Konzentrationen, um eine entsprechende Separation ähnlich wie im PRESS Modell zu erzielen. Dabei geht man davon aus, daß in wässrigen Lösungen im dreidimensionalen Raum typische Diffusionsraten, D , (zur Verdeutlichung wird nicht die Parzellenmigrationsrate m benutzt) für RNA-Sequenzen mit einer Länge von *ca.* 100 Basen ungefähr $D \approx 10^{-11} m^2/s$ sind. Auf der Zeitskala τ von wenigen Sekunden für *in vitro* Experimente mit RNA [78] erhält man eine charakteristische Längenskala von $L = \sqrt{D\tau} = 10^{-5} m$.

Ein Kubus - der gerade einer Parzelle im PRESS Modell entsprechen soll - dieser Dimension hat ungefähr ein Volumen von $1\mu\text{l}$. Bei einer typischen Parzellenpopulation der Größenordnung $10^0 - 10^1$ erhält man eine Konzentration von *ca.* 10pM . Dies ändert sich allerdings deutlich, wenn man Gele oder porösen Medien benutzt, bei denen die Diffusionsrate reduziert ist. Auch in zwei Dimensionen ändert sich die Konzentration je nach Medium um bis zu 5 Größenordnungen. Trotzdem läßt sich die Notwendigkeit nach geringen Konzentrationen erkennen, um der PRESS Modellierung zu entsprechen. Die Parzellenstruktur des Systems zeigt sich darüber hinaus als ein intermediäres Modell zwischen den kontinuierlichen [22, 28, 23] und den streng kompartimentierenden, diskreten, räumlichen Ansätzen [79, 21] und dem Modell der stochastischen Korrektur von Szathmáry und Demeter [80, 81]. In seiner Konzeption ähnelt das PRESS System dem Inselmodell von Kimura [40], jedoch für den Fall einer geringen Inselpopulation und einer vollständig exakten Berechnung ohne Modellnäherungen.

Bei der Behandlung des PRESS Modells zeigte sich weiter, daß die Komplexität der Beschreibung - sprich die Anzahl von Wahrscheinlichkeitszuständen - schnell mit ansteigender Teilchen- und Spezieszahl zunimmt, so daß bereits das katalytische Modell mit 10 Zuständen nicht mehr vollständig analytisch lösbar war. Allerdings resultiert die PRESS Modellierung immer in ein System von gewöhnlichen Differentialgleichungen, die numerisch sehr gut zu handhaben sind. So konnte schließlich das Verhalten des katalytischen Modells mit einer Parzellenpopulation von 10 - was insgesamt 66 internen Zuständen entspricht - problemlos numerisch berechnet werden. Auch die unendliche Dimensionalität des Simplex zeigte sich als geringe Einschränkung, da Simulationen in physikalischen Raumdimensionen die prinzipiellen Eigenschaften der Ergebnisse jeweils bestätigten. Wie niedere Dimensionen direkt in das PRESS Modell aufgenommen werden können, ist zur Zeit noch nicht geklärt. Erste Ansätze einer Kopplung von stochastischer Diffusionsmodellierung und analytischer PRESS Beschreibung haben sich bereits bewährt (Füchslin).

Die Anwendung des PRESS Systems auf die einzelnen Modelle zeigte ein prinzipiell unterschiedliches Verhalten von Selbstreplikatoren und wechselwirkenden Systemen in räumlichen Umgebungen. Resultierte der Effekt räumlicher Korrelationen im Fall der Quasispezies in einen universell negativen Einfluß auf die Menge von Informationen, die in einfachen nicht-wechselwirkenden Populationen unter dem Druck von Mutation und Selektion generiert werden kann, so zeigten die katalytischen Molekültypen entgegen der homogene Berechnung eine Stabilisierung in bestimmten Migrationsbereichen. Diese Stabilisierung wurde dabei völlig ohne artifizielle Verbindung der katalytischen Spezies erzielt. Vielmehr noch reichte es aus, nur eine einzelne katalytische Sequenz zu haben, die eine vollständige Population erhalten konnte. Diese Erkenntnis ist im Hinblick auf eine mögliche Grundstufe evolvierbarer Systeme sehr wertvoll, da in diesem Kontext das gleichzeitige Auftreten mehrerer unterschiedlicher katalytischer Spezies beliebig unwahrscheinlich ist. Genau diese Forderung stellen aber viele der bisherigen Ansätze, wie beispielsweise das Hyperzyklusmodell [4, 17, 18] oder die autokatalytischen Netzwerke [60].

Eine genauere Analyse der Eigenschaften brachte zu Tage, daß im einfachen wechselwirkenden System in Abhängigkeit der Migration unterschiedlich stabile Zustände generiert werden konnten. So zeigten sich in diesem Zusammenhang Grenzyklen, die als Ergebnisse einer Hopf Bifurkation bei kritischen Migrationsraten auftraten. Das reduzierte Auftreten der Grenzyklen mit ansteigender Parzellenpopulation konnte als Ergebnis einer Entkopplung von parasitärer und katalytischer Spezies verstanden werden.

Eine Übersicht der wichtigsten Eigenschaften zeigt modellübergreifend Tabelle 6.1.1. Es

soll betont werden, daß das Modell eine Berechnung sowohl der kritischen Migrationsbereiche (ähnlich wie die obige Konzentrationsabschätzung) als auch weiterer, beliebiger Systemeigenschaften erlaubt, so daß die theoretischen Vorhersagen im Rahmen experimenteller Untersuchungen verifiziert werden können.

Anhand der in den einzelnen Kapiteln vorgestellten Ergebnisse werden zusammenfassend folgende Aussagen getroffen, die - wie oben angegeben - zum Teil dem bisherigen *Status Quo* widersprechen:

1. Zur Stabilisierung von Kooperation sind keine Kopplungen - weder schwache noch starke - zwischen den Spezies notwendig.
2. Eine Population kann allein durch eine einzige funktionelle und altruistische Spezies aufrecht erhalten werden.
3. Die evolutive Stabilisierung benötigt keine globalen Muster; einfache Raumseparation ist ausreichend.
4. Partielle Sequenzerkennung kann die Stabilität einfacher, katalytischer Systeme erhöhen.
5. Faltungsmechanismen erlauben darüber hinaus eine deutliche Stabilitätserhöhung und damit eine mögliche Sequenzverlängerung.

Daraus läßt sich schlußfolgern:

Nicht der Hyperzyklus ist die nächst höhere Organisationsstufe zu den Darwinschen Systemen (Selbstreplikatoren). Es sind die einzelnen, einfachen (altruistischen) Katalysatoren!

6.1.2 Quo vadis?

Nach der Zusammenfassung der Ergebnisse dieser Arbeit soll ein Ausblick auf weitere Untersuchungen und damit der Fokus auf mögliche Erweiterungen des katalytischen Systems gelenkt und Bezug zu komplexeren Systemen genommen werden. Dabei ist es - im Hinblick auf die Evolution - von Bedeutung, ob die gefundene Stabilisierung des minimalen, katalytischen Systems tatsächlich als eine mögliche Grundstufe einer eventuellen RNA-Welt angesehen werden kann. Dies ist nur dann der Fall, wenn das System auch erweiterbar ist, daß heißt, wenn neue, funktionelle Moleküle in das System eingebaut werden können und es trotzdem gegen Parasitenbefall resistent bleibt. Nur diese Erweiterungen könnten auch erklären, wie es von dem Wechsel einer möglichen frühzeitigen RNA-Welt zu der deutlich komplexeren DNA/Protein Welt kam, die man heute kennt.

Beide Welten unterscheiden sich vor allem dadurch, daß im Gegensatz zur RNA-Welt, bei der die informationstragenden Sequenzen letztlich nur durch eine Faltungsoperation von der funktionellen Form getrennt wird, die Information und die Funktion in der DNA/Protein-Welt separiert werden. Dort sorgen die Prozesse der Transkription und der Translation dafür, daß entsprechend der in der DNA gespeicherten Informationen die zugehörigen Aminosäuren gebildet werden, aus denen sich schließlich die Proteine zusammensetzen. Die Einführung der Translation erlaubt eine mögliche Änderung der Beschreibungsebene evolutiver Prozesse, auf der unter Umständen leichter Optima evaluiert werden können. Daß das genetische

Modelleigenschaft	einfache QS	QS mit Zerfall	Katalysten	Kat. mit Erk.
Wechselwirkung	–	–	×	×
Funktionserweiterung	–	–	×	×
Sequenzenerweiterung	–	–	–	×
krit. Migration	–	m_1, m_2	m_a, m_b, m_c	m_a, m_b, m_c
Effekt des Raums	destabil. $\forall m$	destabil. $\forall m$	stabil. für $m_a \leq m \leq m_b$	stabil. für $m_a \leq m \leq m_b$
asympt. Stabilität	$\forall m$	$\forall m$	$m_a \leq m < m_c$	$m_a \leq m < m_c$
Grenzzyklen	–	–	bei $m_c < m$ für $n = 3, (4)$	bei $m_c < m$ für $n = 3, (4)$
Wertebereich Q_c	$1 > Q_c \geq 1/\sigma$	$1 \geq Q_c \geq 1/\sigma$	$1 \geq Q_c \geq 0.9$	$1 \geq Q_c > 0$
Populationskollaps	–	für $m < m_1$	$m < m_a \vee m_b < m$	$m < m_a \vee m_b < m$

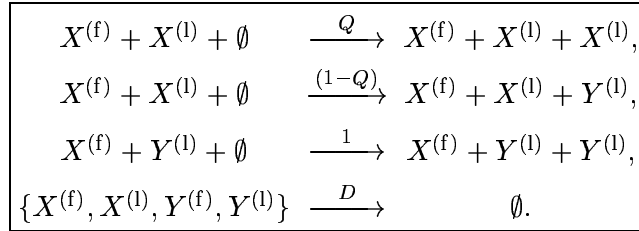
Tabelle 6.1: Die Tabelle vergleicht die Haupteigenschaften der behandelten Modelle. Quasispezies ohne (Kapitel 3.2) und mit explizitem Zerfall (Kapitel 3.3). Kooperatives System ohne (Kapitel 4.1) und mit Teilerkennung (Kapitel 4.2).

Kodierungsverfahren sich im Laufe der Zeit tatsächlich verändert hat, kann an bestimmten Mitochondrien nachgewiesen werden [82]. Für eine Einführung in die Theorie des genetischen Codes siehe [83]. Die Funktionsweise der Translation kann in [84] nachgelesen werden.

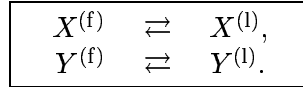
Die Abkopplung der Funktion von der Information wird durch ein Kodierungsverfahren erreicht, wobei noch nicht vollständig geklärt ist, warum die Natur diesen Weg beschritten hat, wieso diese Art der Informationsverarbeitung stabil ist und warum sich, aus der Menge aller möglichen stabilen Kodierungsverfahren, genau eins durchgesetzt hat. Eine frühe Arbeit, die sich mit der Anzahl von Codons und Aminosäuren befasste, konnte zeigen, daß sich deren Zahl mit Hilfe der Aspekte der Koevolution erklären [85] lassen, deren Wirkung ja bereits zu Beginn der Arbeit im Zusammenhang mit parasitären Effekten angesprochen wurde. Wills befaßte sich später mit der Frage der Eindeutigkeit des Kodierungssystem und zeigte, daß stochastische Schwankungen in der frühen Entwicklungsphase einer DNA/Protein-Welt dafür sorgen könnten, daß nur ein Kodierschema übrigbleibt [86].

Neben diesen sehr allgemeinen Fragestellungen bleibt die Frage, ob und wie das bisher benutzte katalytische Modell erweitert werden kann, ohne daß es seine evolutive Stabilität verliert. Damit könnte nicht nur gezeigt werden, daß eine RNA-Evolution prinzipiell möglich wäre, man hätte auch gleichzeitig eine Realisierung dieser Evolution präsentiert. Dabei ist eine naheliegende Erweiterung die Modellierung der ungefalteten und gefalteten Struktur der RNA. Das Konformationsmodell (Altmeyer und McCaskill) berücksichtigt diese Faltungsfähigkeit der RNA explizit. Dazu werden die Moleküle in zwei Gruppen unterteilt, un-

gefaltete, informationstragende Sequenzen, $X_i^{(l)}$, und gefaltete, katalytisch aktive Sequenzen, $X_i^{(f)}$. Die Faltung wird als Prozeß modelliert, der mit vorgegebenen Raten beide Konformationen ineinander überführt. Die möglichen Funktionen werden erneut auf die katalytische Replikation reduziert, wobei jedoch nur gefaltete, katalytisch aktive Sequenzen ungefaltete Moleküle replizieren können. Zusammengefaßt sehen die möglichen Operationen im 2-Spezies Modell wie folgt aus:



Neben diesen Reaktions- und Zerfallsgleichungen kommen beim Konformationsmodell ergänzend die Konformationsänderungen hinzu, die wie folgt aussehen:



In ersten Ergebnissen zeigt sich, daß auch dieses komplexere System im Raum stabilisiert werden kann. Für den Limes schneller Konformationsraten geht es zudem in das einfache katalytische Modell mit reduzierten Reaktionsraten über. Seien k_1 und k_{-1} die Raten der Hin- bzw. Rückkonformation und seien diese zusätzlich deutlich größer als die Replikationsrate des Katalysten, $k_1, k_{-1} \gg 1$. Für diesen Fall ist das Verhältnis von gefalteten zu ungefalteten Molekülen durch $x^{(f)}/x^{(l)} = k_1/k_{-1}$ gegeben. Im einfachen Modell ohne explizite Faltung wird die Rate der Replikation durch den Term x^2 betimmt. Die entsprechende Replikationsrate im Konformationsmodell ist $x^{(f)}x^{(l)}$, wobei die Verbindung zwischen den Modellen mit Hilfe der Beziehung $x = x^{(f)} + x^{(l)}$ geknüpft wird. Damit können die Konzentrationen $x^{(l)}$ und $x^{(f)}$ relativ zu x bestimmt werden. Eine kurze Rechnung führt zu

$$x^{(f)}x^{(l)} = \frac{k_1 k_{-1}}{(k_1 + k_{-1})^2} x^2, \quad (6.1)$$

dabei gelten die obigen Beziehung sowohl für die katalytisch aktiven als auch für die parasitären Sequenzen. Die Reduzierung der katalytischen Aktivität wird somit durch den Faktor $\frac{k_1 k_{-1}}{(k_1 + k_{-1})^2}$ beschrieben. Weitere Erweiterungen und Untersuchungen zum Konformationsmodell werden folgen.

In diesem Zusammenhang soll auf ein weiteres Modell hingewiesen werden, das von McCaskill und Füchslin veröffentlicht wurde [75] und sich mit dem Kodierungsproblem von DNA zu Proteinen befaßt. Die ähnliche Modellierung zum katalytischen Modell läßt jedoch auch Rückschlüsse zu. Die Autoren konnten zeigen, daß ein System, bei dem neben der Replikase zusätzlich noch eine Translatasefunktion gefunden werden muß, daß sowohl die Kodierung der Replikase, als auch der Translatase selbst bestimmt, ebenfalls im Raum stabilisiert werden kann. Damit konnte nicht nur anhand eines existierenden Modells belegt werden, daß der Prozeß der genetischen Kodierung stabil ist, sondern daß auch ein Übergang von einer RNA ähnlichen Formulierung zu DNA/Protein basierten Ansatz möglich ist, was sich aus der bereits formulierten Ähnlichkeit der beiden Modelle ableiten läßt.

Das Kapitel über die konfigurierbare Hardware zeigte eine effektive Simulationsmöglichkeit auf, die zwar *in vitro* Experimente nicht ersetzen, aufgrund der Rechenleistung und ihrer Architektur jedoch Voraussagen leisten kann, die mit konventioneller Hardware nicht möglich sind. Neben dem rein simulativen Aspekt der konfigurierbaren Hardware kann aber auch in diesem Zusammenhang der Gedanke möglicher Erweiterungen verfolgt werden. So sind komplexere Architekturen denkbar und auch bereits realisiert [87], welche die Möglichkeit der selbstständigen Veränderung ihrer Hardware erlauben. Dies kann dazu genutzt werden, daß ein solcher Rechner sich erst auf das gegebene Problem optimiert um dann mit entsprechenden Algorithmen die eigentliche Aufgabe zu lösen. So sind eigenständige Substrukturen denkbar, die sich als effiziente Löser bestimmter Probleme evolutionär ergeben und die mit anderen Strukturen kommunizieren und dabei Daten austauschen. Solche evolvierbare Hardware wäre das silikonbasierte Pendant zu biomolekularen Systemen. In diesen evolvierbaren, wechselwirkenden Systemen entsteht dann aber erneut das Problem der Parasiten, womit sich der Kreis zum Anfang der Arbeit schließt.

Anhang A

Beweise, Ergänzungen und Vertiefungen

A.1 Beweis der Rekursionsgleichung

Zu beweisen ist die Behauptung, daß in einer linearen Kette von Zuständen, wie in Kapitel 2.1.5 beschrieben, auch ohne expliziter Annahme des detaillierten Gleichgewichts die Wahrscheinlichkeit P_k durch

$$P_k = P_0 \prod_{i=0}^{k-1} \frac{w_{i,i+1}}{w_{i+1,i}} \quad (\text{A.1})$$

beschrieben werden kann.

Der Induktionsanfang für P_1 wurde im entsprechenden Abschnitt mit Hilfe der dort beschriebenen Randbedingungen bereits gezeigt, so daß unter der Voraussetzung $P_k = P_0 \prod_{i=0}^{k-1} \frac{w_{i,i+1}}{w_{i+1,i}}$ nur noch der Induktionsschluß von P_k auf P_{k+1} zu zeigen ist. Aus der Master Gleichung ergibt sich, vergl. Gl. (2.7),

$$P_{k+1} = \frac{(w_{k,k+1} + w_{k,k-1})P_k - w_{k-1,k}P_{k-1}}{w_{k+1,k}}. \quad (\text{A.2})$$

Unter Verwendung der Induktionsvoraussetzung für P_k und P_{k+1} folgt

$$P_{k+1} = P_0 \left(\frac{\prod_{i=0}^{k-1} \frac{w_{i,i+1}}{w_{i+1,i}} (w_{k,k+1} + w_{k,k-1}) - \prod_{i=0}^{k-2} \frac{w_{i,i+1}}{w_{i+1,i}} w_{k-1,k}}{w_{k+1,k}} \right) \quad (\text{A.3})$$

$$= P_0 \prod_{i=0}^{k-2} \frac{w_{i,i+1}}{w_{i+1,i}} \left(\frac{\frac{w_{k-1,k}}{w_{k,k-1}} (w_{k,k+1} + w_{k,k-1}) - w_{k-1,k}}{w_{k+1,k}} \right) \quad (\text{A.4})$$

$$= P_0 \prod_{i=0}^{k-2} \frac{w_{i,i+1}}{w_{i+1,i}} \left(\frac{w_{k-1,k} w_{k,k+1}}{w_{k,k-1} w_{k+1,k}} \right), \quad (\text{A.5})$$

was schließlich zu

$$P_{k+1} = P_0 \prod_{i=0}^k \frac{w_{i,i+1}}{w_{i+1,i}} \quad (\text{A.6})$$

umgeformt werden kann und die Aussage beweist. \square

A.2 Rückmutationswahrscheinlichkeit

Für den speziellen Fall der Swetina-Schuster Fitneßlandschaft, die für eine Sequenz k die Replikationsraten

$$A_k = \begin{cases} \sigma & : k = i \\ 1 & : k \neq i \end{cases} \quad (\text{A.7})$$

definiert, wird unter der Annahme einer Gleichverteilung der selbstreplizierenden Sequenzen die Wahrscheinlichkeit der Rückmutation von einer beliebigen Sequenz j zur Mastersequenz i berechnet. Bei einer Sequenzlänge ν können alle Sequenzen entsprechend ihres Hamming Abstands $d(i, j)$ zur Mastersequenz in $\nu + 1$ Fehlerklassen zusammengefaßt werden. Sei nun q die Wahrscheinlichkeit einer korrekten Replikation pro Monomer, das aus einem Alphabet der Größe λ stammt, so ergibt sich die Wahrscheinlichkeit während dieses Prozesses genau n Mutation zu haben, p_n , zu

$$p_n^{(\text{mut})} = (1 - q)^n q^{\nu - n}. \quad (\text{A.8})$$

Die Wahrscheinlichkeit, daß von einer Sequenz mit Hamming Abstand n zur Mastersequenz alle n Mutationen so ausfallen, daß diese Sequenz zur Mastersequenz mutiert, ist

$$p_n^{(\text{ziel})} = \frac{n! (\nu - n)!}{\nu!}. \quad (\text{A.9})$$

Schließlich fehlt noch die Wahrscheinlichkeit, aus einer Gleichverteilten Population eine Sequenz auszuwählen, die genau den Hamming Abstand n aufweist,

$$p_n^{(\text{freq})} = \frac{\nu!}{\lambda^\nu n! (\nu - n)!}. \quad (\text{A.10})$$

Das Produkt von $p_n^{(\text{mut})}$, $p_n^{(\text{ziel})}$ und $p_n^{(\text{freq})}$ ergibt nun die Gesamtwahrscheinlichkeit $p_n^{(\text{rm})}$, daß bei einer zufälligen Auswahl aus einer gleichverteilten Menge, eine den Hamming Abstand n zur Mastersequenz hat und alle Mutationen so sind, daß diese letztlich zur Mastersequenz führen. Daraus folgt

$$p_n^{(\text{rm})} = n_n^{(\text{mut})} p_n^{(\text{ziel})} p_n^{(\text{freq})} = \frac{(1 - q)^n q^{\nu - n}}{\lambda^\nu}. \quad (\text{A.11})$$

Summiert über alle möglichen n ,

$$p^{(\text{rm})} = \sum_{n=1}^{\nu} p_n^{(\text{rm})} \quad (\text{A.12})$$

$$= \frac{\left(\left(\frac{1-q}{q}\right)^\nu - 1\right)(q-1)q^\nu}{\lambda^\nu(2q-1)}, \quad (\text{A.13})$$

und mit $Q = q^\nu$ ergibt sich schließlich

$$p^{(\text{rm})} = \frac{\left((1 - Q^{1/\nu})^\nu - Q\right)(Q^{1/\nu} - 1)}{\lambda^\nu(2Q^{1/\nu} - 1)}. \quad (\text{A.14})$$

Da im allgemeinen nur kleine Mutationswahrscheinlichkeiten interessant sind, kann Gl. (A.14) an der Stelle $Q = 1$ entwickelt werden und man erhält

$$p^{(\text{rm})} \approx p^{(\text{rm})}|_{Q=1} + \frac{\partial p^{(\text{rm})}}{\partial Q}|_{Q=1}(1 - Q) \quad (\text{A.15})$$

$$\approx \frac{1 - Q}{\lambda^\nu \nu}, \quad (\text{A.16})$$

was die Aussage, daß die Gesamtrückmutationswahrscheinlichkeit proportional zu $\nu^{-1}\lambda^{-\nu}$ ist, bestätigt. \square

A.3 Genauigkeit von Zufallszahlen

Da gerade die Erzeugung von Zufallszahlen eine sehr rechenintensive Angelegenheit ist, bietet sich eine Aufspaltung von Zahlen und eine damit resultierende Abnahme der Genauigkeit an. Insbesondere bitweise Mutationen von Sequenzen der Länge ν benötigen gerade ν unabhängige Zufallszahlen pro Bitstring. Dabei stellt sich die Frage, wie genau solche Zufallszahlen sein sollten, um das Ergebnis nicht zu verfälschen.

Sei nun x , mit $0 \leq x \leq 1$, eine Zufallszahl, die mit einem absoluten Fehler von δx berechnet werden kann. Der relative Fehler von x ergibt sich aus $r_x = \delta x/x$. Sei f eine von x abhängige Funktion, so ergibt sich der absolute Fehler von $f(x)$ aus der Gaußschen Fehlerfortpflanzung zu

$$\delta f = \frac{\partial f(x)}{\partial x} \delta x. \quad (\text{A.17})$$

Für den Fall von bitweisen Mutationen eines Bitstrings der Länge ν , definiert sich die Funktion $f(x)$ als $f(x) = x^\nu$, was einen absoluten Fehler von $\delta f = \nu x^{\nu-1} \delta x$ und den relativen Fehler $r_f = \delta f/f$

$$\begin{aligned} r_f &= \frac{\delta f}{f} = \frac{\nu x^{\nu-1} \delta x}{x^\nu} \\ &= \nu \frac{\delta x}{x}, \end{aligned} \quad (\text{A.18})$$

zur Folge hat.

Im folgenden sollen nun Abschätzungen für die Genauigkeit von Zufallszahlen hergeleitet werden, wie sie beispielsweise für die Verwendung von Fehlerschwellen-Beziehungen notwendig sind. Dabei besteht generell die Beziehung $q = \sqrt[\nu]{Q}$, wobei q die Wahrscheinlichkeit einer einzigen Bitmutation (Bitflip) und Q den Qualitätsfaktor beschreiben. Setzt man $x = q$ und $\delta x = 1/2^n$, wobei n die Genauigkeit der Zufallszahl in Bits angibt, so wird Gl. (A.18) zu

$$r_f = \nu \frac{2^{-n}}{\sqrt[\nu]{Q}}. \quad (\text{A.19})$$

Mit $Q \approx 1$ aber $Q < 1$, kann man den Term $\sqrt[\nu]{Q}$ als $\sqrt[1-\epsilon]{1-\epsilon}$ um $\epsilon \approx 0$ entwickeln,

$$\begin{aligned} \sqrt[\nu]{Q} &= \sqrt[1-\epsilon]{1-\epsilon} \\ &\approx 1 - \frac{\epsilon}{\nu} \\ &\approx \frac{\nu - (1 - Q)}{\nu}, \end{aligned} \quad (\text{A.20})$$

womit Gl. (A.19) unter der Annahme $\nu > 1 - Q$ zu

$$r_f \approx \nu 2^{-n} \quad (\text{A.21})$$

wird. Soll der relative Fehler r_f nun kleiner p sein, so ergibt sich für die Anzahl zu benutzender Bits die Abschätzung

$$n > \frac{\ln \nu - \ln p}{\ln 2}. \quad (\text{A.22})$$

Beispiel

Bei einer gewünschten Genauigkeit von $p = 1\%$ und einer Stringlänge von $\nu = 20$ erhält man $n > (3 + 4,6)/0.7 \approx 11$. Dies besagt, daß für die angegebenen Parameter mindestens Zufallszahlen mit einer Genauigkeit von 11 Bits generiert werden müssen. \square

A.4 Eigenschaften der Matrix \mathbf{L}

Die pure Definition der Matrix $L_{ij} = L_{i,j} = w_{i,j} - \delta_{i,j} \sum_k w_{k,j}$ läßt bereits ohne die explizite Kenntnis der Komponenten zwei Schlußfolgerungen zu:

1. Die Matrix ist gesamtwahrscheinlichkeitserhaltend,
 $\sum_i P_i(t) = \text{konstant}$.
2. Die Matrix ist singulär.

Die erste Aussage kann auf

$$\frac{d \sum_i P_i(t)}{dt} = \sum_i \dot{P}_i(t) = 0 \quad (\text{A.23})$$

reduziert werden. Mit $\dot{P}_i(t) = \sum_j L_{ij} P_j(t)$ ergibt sich mit der Definition von L_{ij}

$$\sum_{i,j} (w_{i,j} - \delta_{i,j} \sum_k w_{k,j}). \quad (\text{A.24})$$

Die Ausführung der Summation über i ergibt

$$\sum_{i,j} w_{i,j} - \sum_{j,k} w_{k,j} = 0, \quad (\text{A.25})$$

was Aussage 1 beweist. \square

Die Singularität ist eine direkte Konsequenz der Gesamtwahrscheinlichkeitserhaltung, da die Summe des Gleichungssystems nur dann für alle P_j Null sein kann, wenn eine Zeile als Linearkombination der anderen geschrieben werden kann. \square

A.5 Erzeugung der Matrix \mathbf{L}

Im folgenden soll gezeigt werden, wie die Matrix \mathbf{L} , deren Komponenten den Wahrscheinlichkeitsflüssen zu den kombinierten Zuständen entsprechen, aus den Übergangswahrscheinlichkeiten zu den einzelnen Populationsvariablen (in diesem Fall k_X und k_Y) erzeugt wird. Dazu werden die beiden Vektoren \mathcal{X} und \mathcal{Y} benötigt, deren k te Komponenten die entsprechenden Werte der k_X und k_Y enthalten. Im Allgemeinen gibt es ebenensoviele Vektoren wie unabhängige Populationsvariablen. Die Komponenten $w_{k,k'}$, die letztlich die Matrix bestimmen, werden aus den

$$w_{k_X, k'_X}^{(X,1)}, w_{k_X, k'_X}^{(X,2)}, \dots, w_{k_X, k'_X}^{(X,\alpha)}$$

bei konstantem k_Y , den

$$w_{k_Y, k'_Y}^{(Y,1)}, w_{k_Y, k'_Y}^{(Y,2)}, \dots, w_{k_Y, k'_Y}^{(Y,\beta)}$$

bei konstantem k_X und den

$$w_{k_X, k'_X, k_Y, k'_Y}^{(XY,1)}, w_{k_X, k'_X, k_Y, k'_Y}^{(XY,2)}, \dots, w_{k_X, k'_X, k_Y, k'_Y}^{(XY,\eta)}$$

wie folgt zusammengesetzt: Zu einem Paar k, k' gehören die Populationsvariablen $k_X = \mathcal{X}_k$, $k'_X = \mathcal{X}_{k'}$ und $k_Y = \mathcal{Y}_k$, $k'_Y = \mathcal{Y}_{k'}$, sowie weitere Größen für entsprechend mehr Spezies. Für den Fall, daß z.B. ein $w_{k_X, k'_X}^{(X,i)}$ nur für den Übergang $k'_X = k_X + 1$ definiert ist, muß für das weitere Vorgehen dies explizit durch Multiplikation mit dem Kroneckersymbol $\delta_{k, k'-1}$ berücksichtigt werden. Nach dieser Modifikation ist $w_{k, k'}$ schließlich die Summe über alle Übergangswahrscheinlichkeiten, also

$$w_{k, k'} = \sum_{i=1}^{\alpha} w_{\mathcal{X}_k, \mathcal{X}_{k'}}^{(X,i)} + \sum_{j=1}^{\beta} w_{\mathcal{Y}_k, \mathcal{Y}_{k'}}^{(Y,j)} + \sum_{k=1}^{\eta} w_{\mathcal{X}_k, \mathcal{X}_{k'}, \mathcal{Y}_k, \mathcal{Y}_{k'}}^{(XY,k)}. \quad (\text{A.26})$$

Sind alle $w_{k, k'}$ bestimmt, ergibt sich die Matrix direkt aus der Definition $\mathbf{L} = L_{ij} = L_{i,j} = w_{i,j} - \delta_{i,j} \sum_k w_{k,j}$.

A.6 Matrix \mathbf{L} für Quasispezies mit Zerfall

Die Matrix \mathbf{L} , die sich aus den Kompositionen der Übergangsraten zusammensetzt, sieht für die Quasispezies mit explizitem Zerfall wie folgt aus

$$\mathbf{L} = \begin{pmatrix} -\frac{\bar{x}+\bar{y}}{2} & \frac{2D+m(2-\bar{x}-\bar{y})}{4} & \frac{2D+m(2-\bar{x}-\bar{y})}{4} & 0 & 0 & 0 \\ \frac{m\bar{x}}{2} & -\frac{2(D+\sigma)-m(2+\bar{y})}{4} & \frac{m\bar{x}}{4} & 0 & \frac{2D+m(2-\bar{x}-\bar{y})}{4} & \frac{2D+m(2-\bar{x}-\bar{y})}{2} \\ \frac{m\bar{y}}{2} & \frac{m\bar{y}}{4} & -\frac{2(1+D+m)-m\bar{x}}{4} & \frac{2D+m(2-\bar{x}-\bar{y})}{2} & \frac{m\bar{y}}{4} & 0 \\ 0 & 0 & \frac{1}{2} + \frac{m\bar{y}}{4} & -\frac{2D+m(2-\bar{x}-\bar{y})}{2} & \frac{m\bar{y}}{4} & 0 \\ 0 & \frac{m\bar{y}}{4} + \frac{\sigma(1-Q)}{2} & \frac{m\bar{x}}{4} & \frac{m\bar{x}}{2} & -\frac{4D+m(4-\bar{x}-\bar{y})}{4} & \frac{m\bar{y}}{2} \\ 0 & \frac{m\bar{x}}{4} + \frac{\sigma Q}{2} & 0 & 0 & \frac{m\bar{x}}{4} & -\frac{2D+m(2-\bar{x})}{2} \end{pmatrix}$$

A.7 Transformation von \mathbf{L} zu \mathbf{L}'

Die singuläre Struktur der Matrix \mathbf{L} erlaubt den einfachen Einbau der Forderung $\sum_i P_i(t) = 1$, indem jede Komponente einer Zeile γ der Matrix \mathbf{L} durch 1 ersetzt wird, $L_{\gamma i} = 1$, $\forall i$ und sich daraus die Matrix $\mathbf{L}^{(\gamma)}$ ergibt. Dabei kann γ beliebig gewählt werden. Ersetzt man die entsprechende Komponente $\dot{P}_\gamma(t)$ des im homogenen Fall dem Nullvektor entsprechenden $\dot{\mathbf{P}}(t)$ ebenfalls mit eins, hat man die Forderung $\sum_i P_i(t) = 1$ erfüllt. Ein entsprechendes Gleichungssystem sähe allgemein für N_Z Zustände wie folgt aus,

$$\begin{pmatrix} 0 \\ \vdots \\ 0 \\ 1 \\ 0 \\ \vdots \\ 0 \end{pmatrix} = \begin{pmatrix} a_{1,1} & a_{1,2} & \dots & a_{1,N_Z-1} & a_{1,N_Z} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ a_{\gamma-1,1} & a_{\gamma-1,2} & \dots & a_{\gamma-1,N_Z-1} & a_{\gamma-1,N_Z} \\ 1 & 1 & \dots & 1 & 1 \\ a_{\gamma+1,1} & a_{\gamma+1,2} & \dots & a_{\gamma+1,N_Z-1} & a_{\gamma+1,N_Z} \\ \vdots & \vdots & \dots & \vdots & \vdots \\ a_{N_Z,1} & a_{N_Z,2} & \dots & a_{N_Z,N_Z-1} & a_{N_Z,N_Z} \end{pmatrix} \begin{pmatrix} P_1 \\ \vdots \\ P_{\gamma-1} \\ P_\gamma \\ P_{\gamma+1} \\ \vdots \\ P_{N_Z} \end{pmatrix}, \quad (\text{A.28})$$

wobei die $a_{i,j} = w_{i,j} - \delta_{i,j} \sum_k w_{k,j}$ als Platzhalter für die Komponenten dienen.

A.8 Homogene Quasispezies mit Zerfall

Um den Effekt des expliziten Zerfalls auf die homogenen Quasispezies zu studieren, werden erneut die vereinfachenden Annahmen aus Kapitel 3.1.2 benutzt. Damit ein zeit- und konzentrationsabhängiger Ausfluß nicht die Effekte des Zerfalls kompensiert, wird $\Phi(t) = 0$ gesetzt. Ausgehend von Gl. (3.30) und Gl. (3.31),

$$\dot{x} = \sigma Qx(1 - x - y) - Dx, \quad (3.30)$$

$$\dot{y} = \sigma(1 - Q)x(1 - x - y) + y(1 - x - y) - Dy, \quad (3.31)$$

ergibt sich im stationären Zustand, definiert durch $\dot{x}(t) = \dot{y}(t) = 0$, aus Gl. (3.30), $y(t) = y = 1 - x - D/(\sigma Q)$. Substituiert in Gl. (3.31) resultiert für $Q = Q(x, D)$

$$Q(x, D) = \frac{1 + D + (\sigma - 1)x + \sqrt{(1 + D + (\sigma - 1)x)^2 - 4D}}{2\sigma}. \quad (A.29)$$

Mit $Q_c(D) = Q(x \rightarrow 0, D)$ erhält man

$$Q_c(D) = \frac{1 + D + |1 - D|}{2\sigma}, \quad (A.30)$$

wobei der Betragsterm $|1 - D|$ Gl. (A.30) in zwei Regionen zerlegt,

$$Q_c(D) = \begin{cases} 1/\sigma & : D \leq 1 \\ D/\sigma & : D > 1 \end{cases}. \quad (A.31)$$

Anschaulich gesprochen, reduziert der explizite Tod die Stabilität der Mastersequenz erst dann, wenn die Zerfallsrate D größer als die Replikationsrate der Nicht-Mastersequenz ist. Für $D > 1$, wird die kritische Fehlerrate proportional zu D erniedrigt, bis sie bei $D = \sigma$ den Wert Null erreicht. Dies ist leicht zu verstehen, da Moleküle schneller zerfallen als sie entstehen und das ganze System somit aussterben muß.

A.9 Homogenes Kooperationsmodell

Es wird für das einfache Kooperationsmodell in einer nicht-räumlich strukturierten Umgebung gezeigt, daß bereits bei infinitesimalen Mutation die funktionalen Elemente, X , nicht mehr aufrecht erhalten werden können und aus dem System verschwinden. Als Folge besteht für einen gewissen Zeitraum das System nur noch aus Y -Sequenzen, die jedoch - aufgrund ihrer mangelnden Fähigkeit zur Replikation - ebenfalls stetig zerfallen und aussterben. Ausgehend von Gl. (4.1) und Gl. (4.2),

$$\dot{x} = Qx^2(1 - x - y) - Dx, \quad (4.1)$$

$$\dot{y} = (1 - Q)x^2(1 - x - y) + xy(1 - x - y) - Dy, \quad (4.2)$$

soll wieder der stationäre Fall untersucht werden. Löst man Gl. (4.2) nach y auf und setzt die sich ergebenden zwei Lösungen in Gl. (4.1) ein, erhält man das Lösungspaar

$$x = 0 \quad \wedge \quad y = 0, \quad (A.32)$$

$$x = \frac{D}{Q} \quad \wedge \quad y = 1 - \frac{D}{Q} - \frac{D}{Qx} = -\frac{D}{Q}. \quad (A.33)$$

Mit $D, Q > 0$ ist y in Gl. (A.33) jedoch negativ und damit unphysikalisch, so daß nur die triviale Lösung $x = 0, y = 0$ in Frage kommt. Dies jedoch heißt, daß das kooperative Modell für kein $Q < 1$ zu stabilisieren ist. \square

A.10 Erzeugung der Matrix \mathbf{R}

Im folgenden soll gezeigt werden, wie die singuläre ($N_Z \times N_Z$) Matrix \mathbf{L} auf eine nicht-singuläre Matrix ($(N_Z - 1) \times (N_Z - 1)$), \mathbf{R} , reduziert werden kann und dabei die Zeitentwicklungsgleichung weiterhin homogen bleibt. Dies soll am instruktiven Beispiel einer 4×4 Matrix demonstriert werden, wobei die Verallgemeinerung evident ist. Sei nun

$$\mathbf{L} = \begin{pmatrix} a_{1,1} & a_{1,2} & a_{1,3} & a_{1,4} \\ a_{2,1} & a_{2,2} & a_{2,3} & a_{2,4} \\ a_{3,1} & a_{3,2} & a_{3,3} & a_{3,4} \\ a_{4,1} & a_{4,2} & a_{4,3} & a_{4,4} \end{pmatrix}, \quad (\text{A.34})$$

wobei die $a_{ij} = w_{i,j} - \delta_{i,j} \sum_k w_{k,j}$ sind. Die Singularität erlaubt nun das Streichen einer beliebigen Zeile. Für das weitere Vorgehen ist es nun wichtig, die Zeile auszuwählen, die das zeitliche Verhalten des Vakuums festlegt. In diesem Beispiel sei dies \dot{P}_1 . Damit ist gesichert, daß die Rücktransformationsvektoren einen Nulleintrag in der ersten Komponente haben, $\mathcal{X}_1 = \mathcal{Y}_1 = \dots = 0$. Nach Streichung der ersten Zeile erhält man

$$\begin{pmatrix} \dot{P}_2 \\ \dot{P}_3 \\ \dot{P}_4 \end{pmatrix} = \begin{pmatrix} a_{2,1} & a_{2,2} & a_{2,3} & a_{2,4} \\ a_{3,1} & a_{3,2} & a_{3,3} & a_{3,4} \\ a_{4,1} & a_{4,2} & a_{4,3} & a_{4,4} \end{pmatrix} \begin{pmatrix} P_1 \\ P_2 \\ P_3 \\ P_4 \end{pmatrix}. \quad (\text{A.35})$$

Um auch den rechten Vektor um eine Komponente zu reduzieren, werden die drei Werte $a_{2,1}P_1$, $a_{3,1}P_1$ und $a_{4,1}P_1$ separat hingeschrieben,

$$\begin{pmatrix} \dot{P}_2 \\ \dot{P}_3 \\ \dot{P}_4 \end{pmatrix} = \begin{pmatrix} a_{2,1}P_1 \\ a_{3,1}P_1 \\ a_{4,1}P_1 \end{pmatrix} + \begin{pmatrix} a_{2,2} & a_{2,3} & a_{2,4} \\ a_{3,2} & a_{3,3} & a_{3,4} \\ a_{4,2} & a_{4,3} & a_{4,4} \end{pmatrix} \begin{pmatrix} P_2 \\ P_3 \\ P_4 \end{pmatrix}. \quad (\text{A.36})$$

Aufgrund der Wahrscheinlichkeitserhaltung $\sum_{i=1}^{N_Z} P_i = 1$ kann P_1 als $1 - \sum_{i=2}^{N_Z} P_i$ geschrieben werden, so daß sich

$$\begin{pmatrix} a_{2,1}P_1 \\ a_{3,1}P_1 \\ a_{4,1}P_1 \end{pmatrix} = \begin{pmatrix} a_{2,1} - a_{2,1}(P_2 + P_3 + P_4) \\ a_{3,1} - a_{3,1}(P_2 + P_3 + P_4) \\ a_{4,1} - a_{4,1}(P_2 + P_3 + P_4) \end{pmatrix} \quad (\text{A.37})$$

ergibt, wobei der zweite Teil nun als Komponenten der Matrix geschrieben werden kann, also

$$\begin{pmatrix} \dot{P}_2 \\ \dot{P}_3 \\ \dot{P}_4 \end{pmatrix} = \begin{pmatrix} a_{2,1} \\ a_{3,1} \\ a_{4,1} \end{pmatrix} + \begin{pmatrix} a_{2,2} - a_{2,1} & a_{2,3} - a_{2,1} & a_{2,4} - a_{2,1} \\ a_{3,2} - a_{3,1} & a_{3,3} - a_{3,1} & a_{3,4} - a_{3,1} \\ a_{4,2} - a_{4,1} & a_{4,3} - a_{4,1} & a_{4,4} - a_{4,1} \end{pmatrix} \begin{pmatrix} P_2 \\ P_3 \\ P_4 \end{pmatrix}. \quad (\text{A.38})$$

Bisher waren es nur Standardumformungen, die nicht von der expliziten Form der Matrixeinträge abhängig waren und die daher auch mit jeder beliebigen singulären Matrix möglich sind. Die spezielle Struktur der Matrix \mathbf{L} erlaubt es nun, auch die restlichen Einträge des Vektors in die Matrix einzubauen. Da die erste Zeile gestrichen wurde, bestimmen die drei Komponenten $a_{2,1}$, $a_{3,1}$ und $a_{4,1}$ die Raten, mit der der Vakuumszustand in einen teilbesetzten Zustand übergeht. Da gemäß der benutzten Modellierung keine Vakuumsfluktuationen - was in diesem Zusammenhang bedeutet, daß aus dem Nichts per Reaktion nichts entstehen kann - möglich

sind, können die Einträge nur Migrationsterme, also von der Form $a_{i,1} = c_i m \bar{k}_{X,Y,\dots}$, oder gleich Null sein, die eine Veränderung des Vakuumszustands erlauben. Wobei die beliebige Aufstellung der Zustände bestimmt, an welcher Stelle sich ein X, Y, \dots befindet. Jede Molekülsorte kommt dabei genau einmal vor. Aus der Überlegung ist ferner abzuleiten, daß der Faktor c_i für alle i gleich ist und dem Wert $1/n$ entspricht. Um eine explizite Matrix genauer zu studieren, sei nochmals auf Anhang A.6 hingewiesen, wo die unreduzierte (6×6) Matrix zu den Quasispezies mit Zerfall abgebildet ist. Die Werte der $\bar{k}_X, \bar{k}_Y, \dots$ ergeben sich erneut aus der Selbstkonsistenzgleichung, die im folgenden partiell in die Matrix eingebaut wird. Dazu benutzt man für die Mittelwerte \bar{k}_X und \bar{k}_Y die bekannte Beziehung

$$\bar{k}_X = \kappa = \sum_{i=1}^{N_Z} \mathcal{X}_i P_i, \quad (3.45)$$

$$\bar{k}_Y = \zeta = \sum_{i=1}^{N_Z} \mathcal{Y}_i P_i. \quad (3.46)$$

Da angenommen wurde, daß die jeweils ersten Komponenten von \mathcal{X} und \mathcal{Y} Null entsprechen, kann die Summe auch von $i = 2$ starten, so daß die Populationsvariablen nur von den P_2 bis P_{N_Z} abhängig sind. Das erlaubt eine weitere Vereinfachung, wobei sich beispielsweise die Form

$$\begin{pmatrix} \dot{P}_2 \\ \dot{P}_3 \\ \dot{P}_4 \end{pmatrix} = \begin{pmatrix} a_{2,2} - a_{2,1} + m/n\mathcal{X}_2 & a_{2,3} - a_{2,1} + m/n\mathcal{X}_3 & a_{2,4} - a_{2,1} + m/n\mathcal{X}_4 \\ a_{3,2} - a_{3,1} + m/n\mathcal{Y}_2 & a_{3,3} - a_{3,1} + m/n\mathcal{Y}_3 & a_{3,4} - a_{3,1} + m/n\mathcal{Y}_4 \\ a_{4,2} & a_{4,3} & a_{4,4} \end{pmatrix} \begin{pmatrix} P_2 \\ P_3 \\ P_4 \end{pmatrix}, \quad (A.39)$$

ergibt, wenn die Komponente $a_{2,1}$ zu \bar{k}_X , $a_{3,1}$ zu \bar{k}_Y und $a_{4,1}$ zu Null korrespondiert. Die resultierende Matrix soll fortan mit \mathbf{R} bezeichnet werden. Eine Verallgemeinerung auf N_Z Zuständen und s Spezies ist evident. Schließlich erlaubt das Verfahren allgemein eine Reduktion von $\mathbf{L} \rightarrow \mathbf{R}$, wobei weiterhin die homogene Zeitentwicklung

$$\dot{\mathbf{P}} = \mathbf{R}\mathbf{P} \quad (A.40)$$

gilt (mit dimensionsangepaßten \mathbf{P} und $\dot{\mathbf{P}}$). Da die Matrix \mathbf{R} nun nicht mehr singulär ist, können beispielsweise Umkehrmatrizen oder Determinanten berechnet werden.

A.11 Lineare Stabilitätsanalyse

A.11.1 Allgemeine Begriffe

In diesem Kapitel wird kurz auf die Theorie der linearen Stabilitätsanalyse eingegangen und gezeigt, wie diese benutzt werden kann, um Aussagen über die stationären Lösungen des PRESS Modells zu erzielen. Für weitere Informationen siehe [88, 89].

Die Menge aller dynamischen Systeme kann grob in zwei Teile zerlegt werden: Diskrete und kontinuierliche Systeme, wobei sich beide Adjektive auf die Zeitentwicklung beziehen. Sowohl die diskreten, als auch die kontinuierlichen Systeme können wiederum in autonome und nicht-autonome Systeme unterschieden werden. Nicht-autonome Systeme zeichnen sich dadurch aus, daß sie im Gegensatz zu den autonomen eine (zeitabhängigen) Wechselwirkung mit ihrer Umwelt haben. Durch Aufnahme der Zeit in den Phasenraum ist es jedoch möglich,

nicht-autonome Systeme in autonome zu überführen. Zur Vollständigkeit soll erwähnt werden, daß es bei den diskreten Modellen auch noch die Klasse der stochastischen Systeme gibt, bei der der Determinismus zwischen den einzelnen Zuständen verloren geht. Trotz der Vielzahl an unterschiedlichen Klassen, reduziert sich die Untersuchung auf die der autonomen, kontinuierlichen Systeme, die alle behandelten PRESS Modelle beinhaltet.

Dynamische Systeme werden durch ihren Zustand - im allgemeinen ein Zustandsvektor \mathbf{u} - und ihre Zeitentwicklung (Dynamik) - $\dot{\mathbf{u}} = \mathbf{F}(\mathbf{u})$ - eindeutig definiert. Die Anzahl der Komponenten von \mathbf{u} bzw. $\mathbf{F}(\mathbf{u})$ werden durch die Dimension des Phasenraums bestimmt. Ist dieser kleiner als vier, kann der Phasenraum durch ein sogenanntes Flußdiagramm anschaulich dargestellt werden, wobei die Achsen die Komponenten des Zustandsvektors sind. Die Kurven im Phasenraum, auf denen sich das System bewegt, werden als Trajektorien bezeichnet. Die Linien, bei denen die Zeitableitung einer Komponente Null ist werden als Nullklinen bezeichnet. Ein Schnittpunkt aller Nullklinen wird als stationärer Zustand bzw. Fixpunkt bezeichnet, da dort alle Ableitungskomponenten null sind, siehe Abb. A.1 Diese Fixpunkte können nun stabil oder instabil sein. Die Frage nach der Stabilität und dem Verhalten solcher Fixpunkte ist Forschungsgegenstand der Stabilitätsanalyse.

A.11.2 Stabilitätskriterien

Bereits 1892 entwickelte Ljapunow seinen Stabilitätssatz, der in schwacher Form besagt (Stabilität), daß ein Orbit, der in der Nachbarschaft eines Fixpunktes startet, in dieser Nachbarschaft verweilt. Zusätzlich zu dieser Form existiert auch noch der Begriff der asymptotische Stabilität. Diese starke Form des Stabilitätsbegriffs gibt an, daß jeder Orbit, der in der Nachbarschaft eines Fixpunktes startet, diesen asymptotisch erreicht. Asymptotisch stabile Fixpunkte werden auch als Attraktoren bezeichnet. Die asymptotische Stabilität, ist equivalent zu der Forderung, daß eine hinreichend kleine (infinitesimale) Störung der Zustände aus dem Fixpunkt mit der Zeit verschwinden.

Sei nun \mathbf{u}_0 ein Fixpunkt und $\delta\mathbf{u}(t)$ eine Störung, so daß $\mathbf{u}(t) = \mathbf{u}_0 + \delta\mathbf{u}(t)$ ist und für die Zeitentwicklung

$$\frac{d(\mathbf{u}_0 + \delta\mathbf{u}(t))}{dt} = \frac{d\delta\mathbf{u}(t)}{dt} = \mathbf{F}(\mathbf{u}_0 + \delta\mathbf{u}(t)) \quad (\text{A.41})$$

gilt. Für die lineare Stabilitätsanalyse wird der Term $\mathbf{F}(\mathbf{u}_0 + \delta\mathbf{u}(t))$ mit Hilfe der Fréchet-Ableitung linear um \mathbf{u}_0 entwickelt. Die Ableitungsmatrix

$$\mathbf{F}'(\mathbf{u}) = \left(\frac{\partial F_i(\mathbf{u})}{\partial u_j} \right) \quad (\text{A.42})$$

wird auch als Jacobi Matrix bezeichnet. Der Gleichgewichtspunkt \mathbf{u}_0 ist dann asymptotisch stabil, wenn alle Eigenwerte der Matrix $\mathbf{F}'(\mathbf{u})$ in der linken, offenen (komplexen) Halbebene liegen, also die Realteile der Eigenwerte echt kleiner Null sind. Da die Berechnung der Eigenwerte λ_i aus dem charakteristischen Polynom folgt, welches sich wiederum aus der Bedingung $\det(\mathbf{F}'(\mathbf{u}) - \lambda\mathbf{I}) = 0$ ergibt, ist in der Praxis für Probleme deren Phasenraumdimension größer als fünf ist im allgemeinen nicht mehr explizit lösbar. Für die Entscheidung, ob alle Lösungen auf der linken offenen Halbebene liegen, ohne diese zu berechnen, bietet sich das Kriterium von Routh-Hurwitz an [89]. Es besagt, daß alle Nullstellen des charakteristischen Polynoms

$$a_n \lambda^n + a_{n-1} \lambda^{n-1} + \dots + a_1 \lambda + a_0 = 0 \quad (\text{A.43})$$

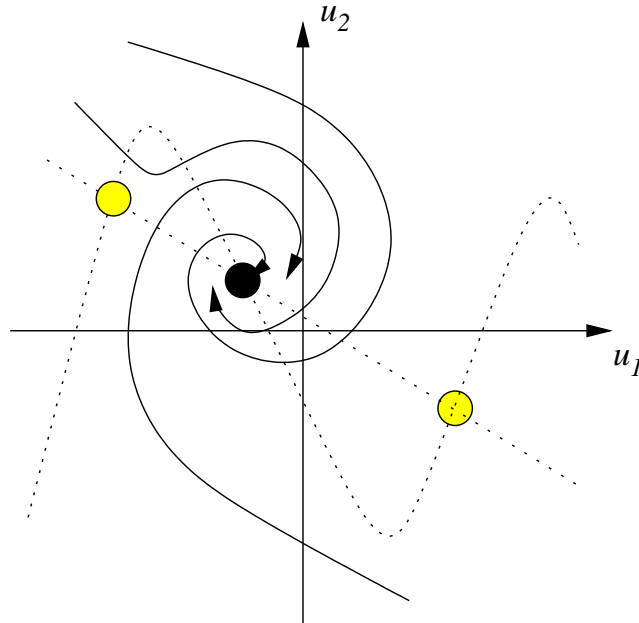


Abbildung A.1: Die Abbildung zeigt einen Ausschnitt aus einem zweidimensionalen Phasenraum, wobei die Komponenten u_1 und u_2 als Achsen dienen. Aufgetragen als Charakteristikum sind die Nullklinen (gepunktete Linien), die an den Schnittpunkten liegenden Fixpunkte (Kreise) und einige ausgewählte Trajektorien (durchgezogene Linien). Die Fixpunkte unterscheiden sich in zwei instabile (hell) und einem stabilen Fixpunkt (dunkel), zu dem sich die Trajektorien asymptotisch bewegen.

genau dann die Lagebedingung erfüllen, wenn alle Determinanten

$$a_1, \begin{vmatrix} a_1 & a_0 \\ a_3 & a_2 \end{vmatrix}, \begin{vmatrix} a_1 & a_0 & 0 \\ a_3 & a_2 & a_1 \\ a_5 & a_4 & a_3 \end{vmatrix}, \dots, \begin{vmatrix} a_1 & a_0 & 0 & \dots & 0 \\ a_3 & a_2 & a_1 & \dots & 0 \\ \dots & \dots & \dots & \dots & \dots \\ a_{2n-1} & a_{2n-2} & a_{2n-3} & \dots & a_n \end{vmatrix} \quad (\text{A.44})$$

positiv sind.

Wie bereits gesagt, sind die Fixpunkte nicht mehr asymptotisch stabil, wenn mindestens ein Realteil der Nullstellen des charakteristischen Polynoms positiv ist. Sei der Aufenthalt der komplexen Eigenwerte λ_i mit $\Re(\lambda_i) = \eta_i$ und $\Im(\lambda_i) = \xi$ von einem Kontrollparameter μ (Parameter der Funktion $\mathbf{F} = \mathbf{F}(\mu)$) abhängig - $\lambda_i = \lambda_i(\mu)$. Eine Variation von μ kann nun dazu führen, daß mindestens eine der Nullstellen die stabile linke offene Halbebene verläßt und nach rechts wandert, was eine Veränderung der Anzahl der Attraktoren zur Folge hat. Dieser Übergang soll o.B.d.A. bei $\mu = 0$ stattfinden. Je nach Anzahl und Art dieser wandernden Nullstellen, werden mehrere Fälle unterschieden. Der einfachste ist, daß mindestens der Realteil eines Eigenwertes größer Null wird, was dazu führt, daß ein asymptotischer Fixpunkt instabil wird. Interessanter ist der Fall, bei dem gleichzeitig zwei komplex konjugierte Nullstellen, $\lambda_{i_1} = \lambda_{i_2}^*$, die rechte Halbebene erreichen. Dieser spezielle Übergang, bei dem ein asymptotisch stabiler Attraktor verschwindet, wird als Hopf Bifurkation bezeichnet. In kontinuierlichen Systemen äußert sich der Übergang als ein stabiler (nicht asymptotisch stabiler) Grenzzyklus (periodische Trajektorien), sofern die zwei weiteren Bedingungen

$$\eta_{i_{1,2}}(\mu = 0) = \omega_0 \neq 0 \text{ und } \left. \frac{d\xi_{i_{1,2}}(\mu)}{d\mu} \right|_{\mu=0} \neq 0 \quad (\text{A.45})$$

erfüllt sind. Dabei bestimmt η die Periode des Grenzyklus. In Abbildung A.2(a) ist diese

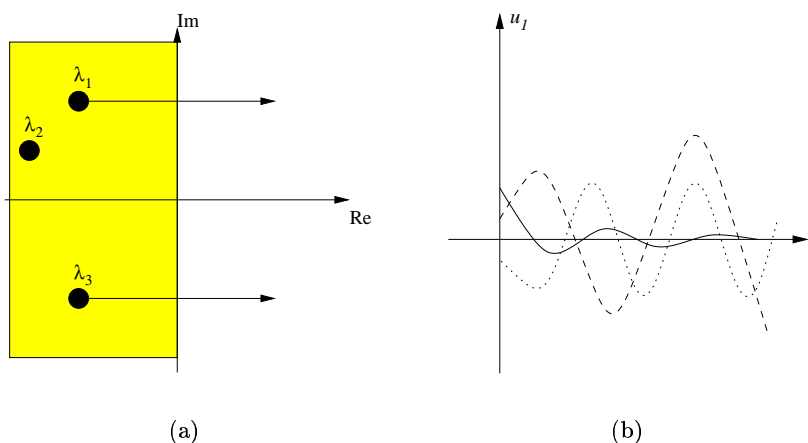


Abbildung A.2: Abbildung (a) zeigt die komplexe Ebene, auf der sich alle Nullstellen des charakteristischen Polynoms befinden. Liegen diese in der linken offenen Halbebene (grau), liegt ein asymptotisch stabiler Zustand vor. In Abhängigkeit eines Kontrollparameters können einige Nullstellen zur rechten Halbebene wandern. In dem Fall, daß zwei konjugiert komplexe Nullstellen dies tun, spricht man von einer Hopf Bifurkation. Das rechte Bild zeigt das Verhalten einer Komponente für asymptotisch stabile (durchgezogene Linie), instabile (gepunktete Linie) und dem Spezialfall der stabilen (nicht asymptotisch stabilen) Zustände, die sich unter weiteren Bedingungen in kontinuierlichen Systemen als periodische Grenzyklen äußern.

Bewegung zweier konjugiert komplexen Nullstellen in Abhängigkeit eines Kontrollparameters zu sehen. Abb. A.2(b) zeigt das Verhalten einer Komponente für einen asymptotisch stabilen, einen stabilen und einen instabilen Fall.

A.11.3 Anwendung auf das Simplex Modell

Im Simplex Modell werden die Zustände im Phasenraum durch den Wahrscheinlichkeitsvektor $\mathbf{P}(t)$ gegeben. Die Dynamik ergibt sich aus der Beziehung $\dot{\mathbf{P}}(t) = \mathbf{R}\mathbf{P}(t)$. Ersetzt man die κ, ζ, \dots Abhängigkeiten in \mathbf{R} mit Hilfe der jeweiligen Selbstkonsistenzgleichungen durch die $\mathbf{P}(t)$, also $\mathbf{R}(\kappa, \zeta, \dots) \rightarrow \mathbf{R}(\mathbf{P}(t))$, so erlaubt die Abbildung

$$\mathbf{u}(t) \rightarrow \mathbf{P}(t), \quad (\text{A.46})$$

$$\mathbf{F}(\mathbf{u}) \rightarrow \mathbf{R}(\mathbf{P})\mathbf{P} \quad (\text{A.47})$$

die Nutzung aller beschriebenen Methoden. So berechnet sich beispielsweise die Jacobi Matrix aus Gl. A.42, welche als Ausgangspunkt für die Stabilitätsuntersuchungen gilt, als

$$\mathbf{F}'(\mathbf{P}) = \left(\frac{\partial(\mathbf{R}(\mathbf{P})\mathbf{P})_i}{\partial P_j} \right), \quad (\text{A.48})$$

und mit entsprechend weiteren Anpassungen.

A.12 Zweidimensionaler Margolus-Algorithmus

Die Realisation eines dreidimensionalen Diffusionsalgorithmus für Parallelrechner steht in enger Beziehung mit der zweidimensionalen Version, die hier beschrieben werden. Die Idee für einen effektiven Diffusionsalgorithmus, der die mit parallel arbeitenden Rechner einhergehenden Problemen löste, wurde erstmal von Margolus und Toffoli im Zusammenhang mit den Parallelrechnern der CAM-Reihe (Cellular Automata Machine) vorgestellt [76]. Dabei wird ein zweidimensionales (in x- und y-Richtung) toroidales Gitter in identische Parzellen aufgeteilt, auf denen jeweils genau ein Datenelement untergebracht wird. Die Grundidee ist es, dieses Gitter mit (2×2) Clustern zu parkettieren, so daß vier Parzellen in einem dieser quadratischen Cluster liegen. Alle Cluster können nun unabhängig voneinander mit einer Wahrscheinlichkeit $p_+ = p_- = p$ um 90° im bzw. gegen den Uhrzeigersinn rotiert werden. Die Wahrscheinlichkeit $1 - p$ gibt dabei an, wie häufig keine Rotation durchgeführt und somit keine Diffusion stattfindet. Mit p kann demnach die Diffusionskonstante justiert werden. Im nachfolgenden Schritt wird die Parkettierung so umorganisiert, daß die Cluster jeweils um genau einen Gitterpunkt in x- und y-Richtung verschoben sind. Mit dieser neu entstandenen Parkettierung werden die Rotationsoperationen wie schon oben beschrieben durchgeführt. Anschließend wird die alte Parkettierung wieder hergestellt und der ganze Prozeß wiederholt. Eine anschauliche Darstellung des Algorithmus findet sich in Abb. A.3 für ein (8×8) Gitter wieder. Das Ergebnis

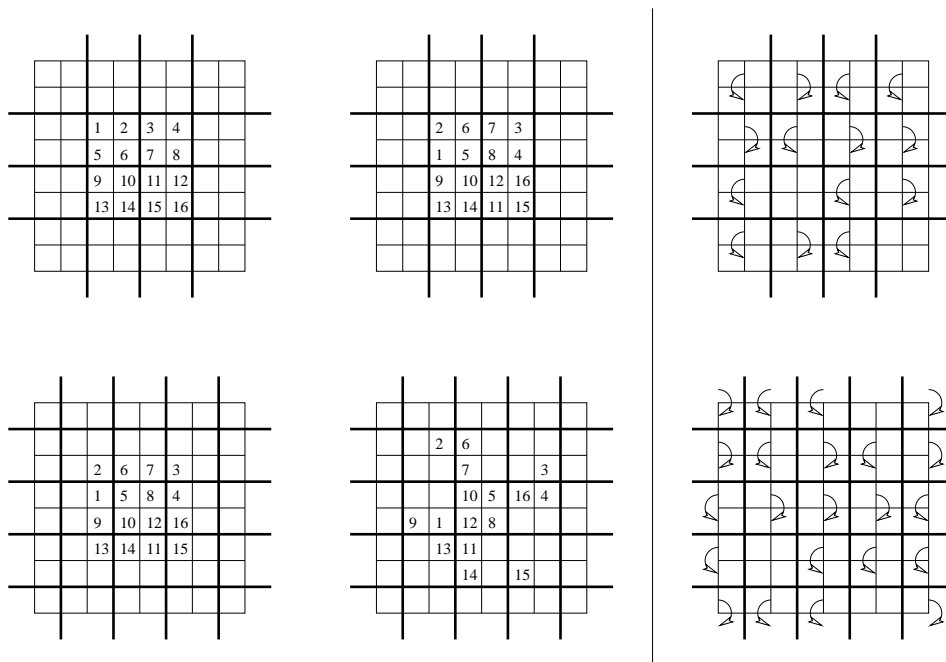


Abbildung A.3: Der linke Bildausschnitt zeigt die zeitliche Entwicklung (von links oben nach rechts unten) von Daten, die dem Margolus-Algorithmus unterliegen. Zur besseren Darstellung wurden die Daten, die alle gleich behandelt werden, durchnummeriert. Nach dem Parkettieren der Ebene in (2×2) Cluster, werden diese korrelationsfrei rotiert (siehe entsprechenden rechten Ausschnitt). Danach ergibt sich die um ein Gitterpunkt in x- und y-Richtung verschobene Parkettierung, auf der erneut rotiert wird. Man erkennt, wie sich die Daten allmählich über die Ebene verteilen.

ist, daß sich selbst für kurze Abstände eine effektive, parallele Diffusion ergibt, bei der sicher gestellt wird, daß zwei Daten nicht auf einen Gitterpunkt springen können.

A.13 Internen Zustände und externe Koordinaten

Z	x	y	z
1	0	0	1
2	1	0	1
3	1	1	1
4	0	1	1

Z	x	y	z
5	0	0	0
6	1	0	0
7	1	1	0
8	0	1	0

Tabelle A.1: Zusammenhang zwischen internen Zuständen, Z und den zugehörigen kartesischen Koordinaten.

Z_1	Z_2
1	7
2	8
3	5
4	6

Z_1	Z_2
5	3
6	4
7	1
8	2

Tabelle A.2: Die Übergänge der internen Zustände bei einem Wechsel der Volumenfüllung. Aufgrund der Geometrie tauschen die Zustände paarweise. Eine Veränderung der externen Koordinaten findet nicht statt.

Z_1	Z_2	Δx	Δy	Δz
1	5	0	0	-1
2	6	0	0	-1
3	2	0	-1	0
4	1	0	-1	0
5	8	0	1	0
6	7	0	1	0
7	3	0	0	1
8	4	0	0	1

Z_1	Z_2	Δx	Δy	Δz
1	2	1	0	0
2	6	0	0	-1
3	7	0	0	-1
4	3	1	0	0
5	1	0	0	1
6	5	-1	0	0
7	8	-1	0	0
8	4	0	0	1

Z_1	Z_2	Δx	Δy	Δz
1	2	1	0	0
2	3	0	1	0
3	4	-1	0	0
4	1	0	-1	0
5	6	1	0	0
6	7	0	1	0
7	8	-1	0	0
8	5	0	-1	0

Z_1	Z_2	Δx	Δy	Δz
1	4	0	1	0
2	3	0	1	0
3	7	0	0	-1
4	8	0	0	-1
5	1	0	0	1
6	2	0	0	1
7	6	0	-1	0
8	5	0	-1	0

Z_1	Z_2	Δx	Δy	Δz
1	5	0	0	-1
2	1	-1	0	0
3	4	-1	0	0
4	8	0	0	-1
5	6	1	0	0
6	2	0	0	1
7	3	0	0	1
8	7	1	0	0

Z_1	Z_2	Δx	Δy	Δz
1	4	0	1	0
2	1	-1	0	0
3	2	0	-1	0
4	3	1	0	0
5	8	0	1	0
6	5	-1	0	0
7	6	0	-1	0
8	7	1	0	0

Tabelle A.3: Die Übergänge der internen Zustände von Z_1 zu Z_2 bei x-, y- und z-Rotation (von links nach rechts) und die entsprechende Veränderung der externen Koordinaten für die positive (oben) und negative (unten) Drehorientierung.

A.14 Implementierung in Hardware

Die Benutzung des Diffusionsalgorithmus (und auch der speziellen Reaktionen) auf der NGEN bedarf deren Übersetzung in die Sprache der Hardware. Um dies effizient tun zu können, benutzt man sehr häufig entsprechende „schematic-entry“ Softwarepakete, die eine komfortable Eingabe der Schaltpläne am PC ermöglichen. Der Versuch, die vollständige Implementierung angeben zu wollen, würde den Rahmen deutlich sprengen. Um trotzdem einen Einblick geben zu können, sollen zwei Abbildungen die unterschiedlichen hierarchischen Ebenen der Logik eines FPGAs verdeutlichen, die die Gesamtfunktionalität der Programme bestimmen. Die Verwendung von Ebenen wird benutzt, um neue Symbole verwenden zu können, die aus elementaren Logikbausteinen (AND, NOT, XOR, Multiplexer, etc.) bestehen. Die Symbole können recht gut mit den Funktionen bzw. den Prozeduren in den höheren Programmiersprachen verglichen werden. Im allgemeinen ist es so, daß je höher man sich in der Ebenenhierarchie befindet, desto mehr Symbole sind vorzufinden. Wandert man sukzessiv tiefer (wobei dies geschieht, indem man virtuell in ein Symbol eintaucht), trifft man zwangsläufig auf eine Ebene, in der alle Symbole aufgelöst und nur noch elementare Logikbausteine anwesend sind. Abbildung A.4 zeigt die dritthöchste Ebene in einem speziellen NGEN Programm, die das vollständige Diffusionsmodul eines FPGAs darstellt. Neben den vielen Datenbussen (breite Linien) und den Einzelverbindungen (dünne Linien), die sowohl für den Transfer der Sequenzen (beispielsweise von Molekülen) als auch von Parametern und Steuerbits notwendig sind, sieht man im rechten, oberen Teil vier gleiche Symbolanordnungen. Das jeweils größte Viereck in den Anordnungen stellt symbolisch einen $(2 \times 2 \times 2)$ Elementarwürfel des dreidimensionalen Diffusionsalgorithmus dar, dessen Dateninhalt in Abhängigkeit der entsprechenden Steuerbits rotiert wird. Begibt man sich virtuell in ein solches Viereck und wandert dort einige Hierarchien hinunter, trifft man auf die Ebene, die nur noch elementare Logikbausteine enthält. Diese wird in Abb. A.5 gezeigt. Ohne detailliert auf die Funktionsweise einzugehen, vermittelt es schließlich doch eine Idee über die recht komplexe Hardwareprogrammierung. Entsprechendes gilt auch für die Einbettung der Reaktionen.

Sind alle notwendigen Schaltpläne erstellt, müssen sie auf ihre logische Konsistenz überprüft werden, anschließend kompiliert und mit entsprechenden Routinen in die Hardwarebausteine der NGEN geladen werden. Dabei wird jeder Gesamtentwurf bis auf wenige Veränderungen fast identisch in alle vorhandenen FPGAs geladen, so daß die Schaltpläne letztlich nur einmal ausgearbeitet werden müssen. Danach ist die NGEN bereit, das Programm abzuarbeiten. Als Lohn der recht komplexen Programmierung überzeugt die bereits zehn Jahre alte NGEN mit immer noch sehenswerten Simulationsleistungen. So wurde beispielsweise ein Reaktions-Diffusions Programm erstellt, welches eine Population von *ca.* 1Million Strings (jeweils von 32Bit Länge) beinhaltet, wobei das vollständige Volumen in einer Sekunde ungefähr 230mal geupdatet wird (Generationen). Der zur Zeit im Bau befindliche Nachfolger der NGEN, die Meregen, überragt diese in der reinen Taktfrequenz um den Faktor 16 und bietet darüberhinaus auch noch ungefähr 16 mal mehr Speicher (ohne zwischen den einzelnen Speichersorten zu unterscheiden). Auch für Meregen gilt für symmetrische (kubische) dreidimensionale Volumen Gl. (5.1), also die Relation $t \propto \sqrt[3]{V}$, so daß man die sehr grobe Abschätzung wagen, daß Meregen eine Population von 16 Millionen Strings (ebenfalls von der Länge 32Bit) beherbergen kann und das gesamte Volumen in einer Sekunde *ca.* 1470mal durchläuft - technisch gesehen unabhängig von der Komplexität der Reaktionen.

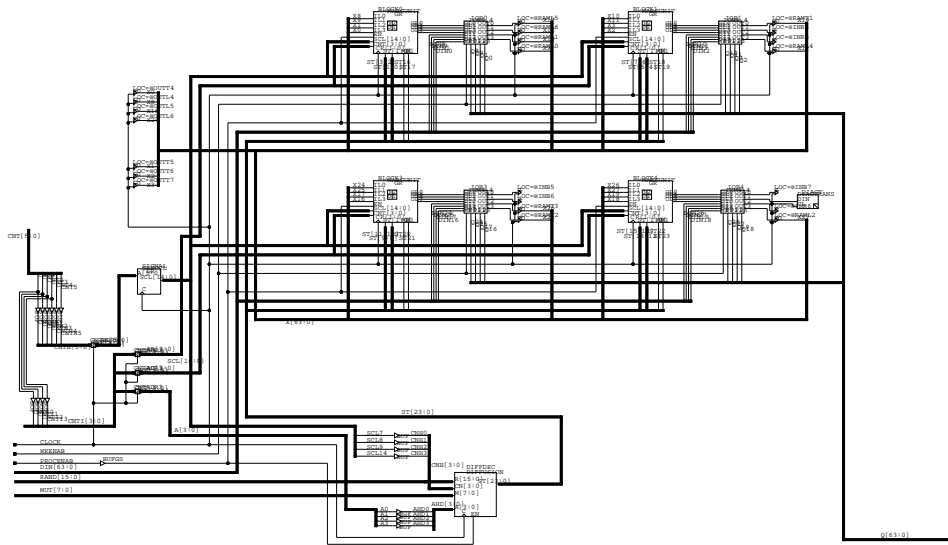


Abbildung A.4: Die Abbildung zeigt einen Schaltplan, der symbolisch alle Elemente beinhaltet, die für die Diffusion auf einem FPGA notwendig sind.
(Mit freundlicher Genehmigung von Dr. Rudolf Füchslin.)

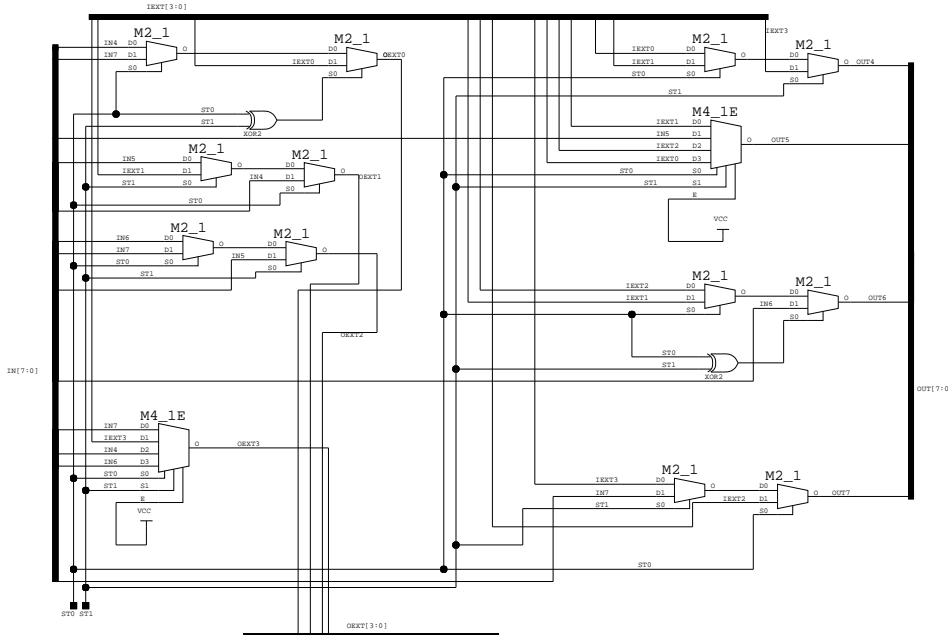


Abbildung A.5: Der Schaltplan stellt eine unterste Ebene innerhalb der vollständigen Diffusionslogik, Abb. A.4 dar. Hier finden sich nur noch elementare Logikbausteine wieder (Multiplexer, XOR, Datenbusse, etc.).
(Mit freundlicher Genehmigung von Dr. Rudolf Füchslin.)

Literaturverzeichnis

- [1] C. Darwin. *On the origin of species by means of natural selection*. John Murray, 1859.
- [2] E. Schrödinger. *What is Life?* Cambridge University Press, 1948.
- [3] M. Eigen. Selforganization of matter and the evolution of biological macromolecules. *Naturwissenschaften*, 58:465–523, 1972.
- [4] M. Eigen and P. Schuster. The hypercycle: A principle of natural self-organization, Part A. *Naturwissenschaften*, 64:541–565, 1977.
- [5] I. Leuthäusser. An exact correspondence of between Eigen's evolution model and a two-dimensional Ising system. *J. Chem. Phys.*, 84:1884–1885, 1985.
- [6] E. Baake, M. Baake, and H. Wagner. Ising quantum chain is equivalent to a model of biological evolution. *Phys. Rev. Lett.*, 78:559–562, 1997.
- [7] R. Pastor-Satorra and R. Sole. Field theory for a reaction-diffusion model of quasispecies dynamics. Santa Fe Report, 2001.
- [8] D. C. Mattis and M. L. Glasser. The use of quantum field theory in diffusion-limited reactions. *Rev Mod Phys*, 70:979–1001, 1998.
- [9] M. Eigen, J. S. McCaskill, and P. Schuster. Molecular Quasi-Species. *J. Phys. Chem.*, 92:6881–6891, 1988.
- [10] M. Eigen, J. S. McCaskill, and P. Schuster. The Molecular Quasi-Species. *Adv. Chem. Phys.*, 75:149–263, 1989.
- [11] P. W. Anderson. Suggested model for prebiotic evolution: The use of chaos. *Proc. Natl. Acad. Sci. USA*, 80:3386–3390, 1983.
- [12] D. S. Rokhsar, P. W. Anderson, and D. L. Stein. Self-organization in prebiological systems: Simulations of a model for the origin of genetic information. *J. Mol. Evol.*, 23:119–126, 1986.
- [13] F. Dyson. A model for the origin of life. *J. Mol. Evol.*, 18:344–350, 1982.
- [14] J. Hofbauer and K. Sigmund. *The Theory of evolution and dynamical Systems*. Cambridge University Press, 1988.
- [15] D. A. Rand, H. B. Wilson, and J. M. McGlade. Dynamics and evolution: evolutionary stable attractors, invasion exponents and phenotype dynamics. *Phil. Trans. R. Soc. Lond. B.*, 343:261–283, 1994.

- [16] S. A. Kauffman. *The origins of order*. Oxford University Press, 1993.
- [17] M. Eigen and P. Schuster. The hypercycle: A principle of natural self-organization, Part B. *Naturwissenschaften*, 65:7–41, 1978.
- [18] M. Eigen and P. Schuster. The hypercycle: A principle of natural self-organization, Part C. *Naturwissenschaften*, 65:341–369, 1978.
- [19] E. Weinberger. Spatial stability analysis of Eigen’s quasispecies model and the less than 5 membered hypercycle under global population regulation. *Bull. Math. Biol.*, 53:623–638, 1991.
- [20] U. Müller-Herold. What is a hypercycle? *J. theor. Biol.*, 182:569–584, 1983.
- [21] M. Eigen, P. Schuster, W. Gardiner, and R. Winkler-Oswatitsch. The origin of genetic information. *Scientific American*, 244:78–94, 1981.
- [22] M. Boerlijst M. and P. Hogeweg. Spiral wave structure in pre biotic evolution: Hypercycles stable against parasites. *Physica D*, 48:17–28, 1991.
- [23] M. B. Cronhjort. Hypercycles versus parasites in the origin of life: Model dependencies in spatial hypercycles systems. *Origins Life Evol B*, 25:227–233, 1995.
- [24] S. Altmeyer, C. Wilke, and T. Martinetz. How fast do structures emerge in hypercycle systems? In C. Wilke, S. Altmeyer, and T. Martinetz, editors, *Third German Workshop on Artificial Life*. Verlag Harri Deutsch, 1998.
- [25] B. P. Belousov. Eine periodische reaktion und ihr mechanismus. In L. Kuhnert, editor, *Selbstorganisation chemischer Strukturen*. Verlag Harri Klein, 1959.
- [26] A. N. Zaikin and A. M. Zhabotinsky. Concentration wave propagation in two-dimensional liquid-phase self-oscillating system. *Nature*, 225:535–538, 1970.
- [27] J. J. Tyson and J. P. Keener. Singular perturbation theory of travelling waves in excitable media. *Physica D*, 32:327–361, 1988.
- [28] M. Boerlijst, M. Lamers, and P. Hogeweg. Evolutionary consequences of spiral waves in a host-parasitoid system. *Proc. Roy. Soc. (London) B*, 253:15–18, 1993.
- [29] J. Ackermann and T. Kirner. Parasites and pattern formation. *Z. Naturforsch.*, 54 A:146–152, 1999.
- [30] J. S. McCaskill. Ursprünge der molekularen Kooperation: Theorie und Experiment. In P. Schuster, editor, *Antrittsvorlesung an der Friedrich Schiller Universität, Jena*, pages 27–40.
- [31] M. B. Cronhjort and C. Blomberg. Cluster compartmentalization may provide resistance to parasites for catalytic networks. *Physica D*, 101:289–298, 1997.
- [32] R. Durrett and S. Levin. The importance of being discrete (and spatial). *Theor. Pop. Biol.*, 46:363–394, 1994.

- [33] R. Durrett and S. Levin. Stochastic spatial models: A users guide to ecological applications. *Phil. Trans. R. Soc. Lond. B*, 343:329–350, 1994.
- [34] Hamilton. The evolution of altruistic behavior. *Am. Nat.*, 97, 1963.
- [35] W. J. Alonso. The role of kin selection theory on the explanation of biological altruism: A critical review. *J. Comp. Biol.*, 3:1–14, 1998.
- [36] S. Wright. Evolution in Mendelian population. *Genetics*, 16:97–159, 1931.
- [37] S. Wright. *Evolution and the genetics of population, Vol. 3*. Univ. of Chicago Press, 1977.
- [38] T. Maruyama. Effective number of alleles in a subdivided population. *Theor. Popul. Biol.*, 1:273–306, 1970.
- [39] B. D. H. Latter. The island model of population differentiation: A general solution. *Genetics*, 73:147–157, 1973.
- [40] M. Kimura. Diffusion model of intergroup selection, with special reference to evolution of an altruistic character. *Proc. Natl. Acad. Sci.*, 80:6317–6321, 1983.
- [41] G. Vojta and M. Vojta. *Taschenbuch der statistischen Physik*. Teubner, 2000.
- [42] H. Haken. *Synergetics*. Springer, 1983.
- [43] D. T. Gillespie. Exact stochastic simulation of coupled chemical reactions. *J. Phys. Chem.*, 81:2340–2361, 1977.
- [44] I. Leuthäusser. Statistical mechanics of Eigen’s evolution model. *J. Stat. Phys.*, 48:343–360, 1987.
- [45] G. Paun, G. Rozenberg, and A. Salomaa. *DNA Computing*. Springer, 1998.
- [46] C. R. Cantor and P. R. Schimmel. *Biophysical Chemistry, part I-III*. W. H. Freeman and Company, 1997.
- [47] S. A. Kauffman and S. Levin. Towards a general theory of adaptive walks on rugged landscapes. *J. Theor. Biol.*, 128:11–45.
- [48] E. Weinberger. Correlated and uncorrelated fitness landscapes and how to tell the difference. *Biol. Cybern.*, 63:325–336, 1990.
- [49] M. Nilsson and N. Snoad. Error threshold on dynamic fitness-landscape. *Phys. Rev. Lett.*, 84:191–194, 2000.
- [50] C. O. Wilke, C. Ronnewinkel, and T. Martinetz. Dynamic fitness landscapes in molecular evolution. *Phys. Rep.*, 349:395–446, 2001.
- [51] B. L. Jones, R. H. Enns, and S. S. Rangnekar. On the theory of selection of coupled macromolecular systems. *Bull. Math. Biol.*, 38:15–28, 1976.
- [52] D. S. Rumschitzki. Spectral properties of eigen’s evolution matrices. *J. Math. Biol.*, 24:667–680, 1987.

- [53] O. Perron. Zur Theorie der Matrices. *Math. Ann.*, 64:248–263, 1907.
- [54] J. Swetina and P. Schuster. Selfreplication with errors. A model for polynucleotide replication. *Biophys. Chem.*, 16:329–353, 1982.
- [55] J. Maynard-Smith. Models of evolution. *Proc. Roy. Soc. (London) B*, 219:315–325, 1983.
- [56] S. Altmeyer and J. S. McCaskill. Error threshold for spatially resolved evolution in the quasispecies model. *Phys. Rev. Lett.*, 86:5819–5822, 2001.
- [57] J. F. Crow and M. Kimura. *Introduction to population genetics theory*. Burgess Publishing Company, Edina, MN, 1970.
- [58] M. Nowak M. and P. Schuster. Error thresholds of replication in finite populations mutation frequencies and the onset of Mullers ratchet. *J. Theor. Biol.*, 137:375–395, 1989.
- [59] D. Alves and J. F. Fontanari. Error threshold in finite populations. *Phys. Rev. E*, 57:7008–7013, 1998.
- [60] P. F. Stadler and P. Schuster. Dynamics of small autocatalytic reaction networks: Bifurcation, permanence and excluision. *Bull. Math. Biol.*, 52:485–508, 1990.
- [61] J. S. McCaskill, R. M. Fuchslin, and S. Altmeyer. The stochastic evolution of catalysts in spatially resolved systems. *Biol. Chem.*, 382:N. N., 2001. in press.
- [62] W. K. Johnston, P. J Unrau, M. S. Lawrence, M. E. Glasner, and D. Bartel. RNA-catalyzed RNA polymerization: Accurate and general RNA-template primer extension. *Science*, 292:1319–1325, 2001.
- [63] J. E. Satulovsky and T. Tome. Spatial instabilities and local oscillations in a lattice gas Lotka-Volterra model. *J. Math. Biol.*, 35:344–358, 1997.
- [64] J. Maynard-Smith. The logic of animal conflict. *Nature*, 246:15–18, 1973.
- [65] G. T. Vickers, V. C. L. Hutson, and C. J. Budd. Spatial patterns in population conflicts. *J. Math. Biol.*, 31:411–430, 1993.
- [66] J. Maynard-Smith. Group selection and kin selection: a rejoinder. *Nature*, 201:1145–1147, 1964.
- [67] M. Kimura. *The neutral theory of molecular evolution*. Cambridge University Press, 1983.
- [68] J. S. McCaskill. The equilibrium partition function and base pair binding probabilities for RNA secondary structure. *Biopolymers*, 29:1105–1119, 1990.
- [69] T. Ellinger, R. Ehricht, and J. S. McCaskill. In vitro evolution of molecular cooperation in CATCH, a cooperatively coupled amplification system. *Chemistry and Biology*, 5:729–741, 1988.
- [70] P. Schuster, W. Fontana, P. F. Stadler, and I. L. Hofacker. From sequences to shapes and back - A case study in RNA secondary structures. *Proc. Roy. Soc. (London) B*, 255:279–284, 1994.

- [71] C. M. Reidys, P. F. Stadler, and P. Schuster. Generic properties of combinatorial maps - Neutral networks of RNA secondary structures. *Bull. Math. Biol.*, 59:339–397, 1997.
- [72] J. S. McCaskill, T. Maeke, U. Gemm, L. Schulte, and U. Tangen. NGEN: A massively parallel reconfigurable computer for biological simulation: Toward a self-organizing computer. In *Lecture notes in computer science*. Springer, 1997.
- [73] J. Breyer. *Anwendung eines in der Hardware konfigurierbaren Parallelrechners auf die Simulation evolutiver in vitro Amplifikationssysteme*. PhD thesis, Friedrich Schiller University, 1998.
- [74] S. Altmeyer. Entwicklung und Programmierung einer dreidimensionalen Architektur für NGEN mit nachfolgender Simulation einer typischen Chemie in drei Raumdimensionen. IMB internal report, 1999.
- [75] R. M. Füchslin, S. Altmeyer, and J. S. McCaskill. Simulation of evolving systems using FPGA technology. In *Caesar Proceedings*. Springer, 2001. in press.
- [76] T. Toffoli and N. Margolus. *Cellular automata machines*. MIT Press, 1987.
- [77] B. Böddeker. *Physikalische Modelle für Koevolution und Simulation auf einem hardwareprogrammierbaren Prozessor / Untersuchung von räumlich und haftungsinduzierter Korrelation auf stabile Symbiose von Funktionen*. Diploma thesis, University of Göttingen, Germany, 1995.
- [78] J. S. McCaskill and G. J. Bauer. Images of evolution: Origin of spontaneous RNA replication waves. *Proc. Natl. Acad. Sci. USA*, 90:4191–4195, 1993.
- [79] C. Bresch, U. Niesert, and D. Harnasch. Hypercycles, parasites and packages. *J. Theor. Biol.*, 85:399–405, 1980.
- [80] E. Szathmary. The eukaryotic cell as an information integrator. *End. Cell. Res.*, 3:113–132, 1986.
- [81] E. Szathmary and L. Demeter. Group selection of early replicators and the origin of life. *J. Theor. Biol.*, 128:463–486, 1987.
- [82] T. Ohama, S. Osawa, S. Watanabe, and K. T. H Jukes. Evolution of the mitochondrial genetic code IV. AAA is an asparagine codon in some animal mitochondria. *J. Mol. Evol.*, 129:2715–2726, 1990.
- [83] J. Maynard Smith and E. Szathmary. *Evolution*. Spektrum, 1996.
- [84] B. Lewin. *Genes VI*. Oxford University Press, 1997.
- [85] J. T. Wong. The evolution of a universal code. *Proc. Natl. Acad. Sci. USA*, 73:2336–2340, 1976.
- [86] P. R. Wills. Self-organization of genetic coding. *J. Theor. Biol.*, 162:267–287, 1993.
- [87] U. Tangen. Self-Organisation in micro-configurable hardware. In M. Bedau, J. S. McCaskill, N. H. Packard, and S. Rasmussen, editors, *Artificial Life VII*. MIT Press, 2000.

- [88] G. Iooss and D. D. Joseph. *Elementary stability and bifurcation theory*. Springer, 1980.
- [89] I. N. Bronstein and K. A. Semendjajew. *Taschenbuch der Mathematik*. Harri Deutsch, 2000.

Curriculum Vitae

Name Stephan Wolfgang Altmeyer
Geboren 21.10.1970 in Bottrop
email stephan.altmeyer@web.de
Zivilstand ledig

Ausbildung

15.05.1990 Abitur am Vestischen Gymnasium, Bottrop
23.12.1996 Diplom der Physik an der Universität Dortmund
26.09.2001 Aufnahme des Promotionsverfahrens

Promotion

02.97-06.99 Forschungstätigkeit bei Prof. Martinetz am Institut für Neuroinformatik, Universität Bochum
01.99-04.99 Forschungsaufenthalt am Institut für Molekulare Biotechnologie, Jena
06.99-12.01 Wechsel zum GMD Forschungszentrum in Sankt Augustin bzw. Universität Köln
09.01-12.01 Promotionsverfahren an der Universität Köln

Deutsche Zusammenfassung

Die Arbeit widmet sich der Frage, wie sich der Einfluß des Raums auf die Stabilisierung biomolekularer, evolvierbarer Systeme auswirkt.

Dazu wird eine Modellplattform - das Press Modell - hergeleitet, die es erlaubt, analytische Aussagen zu treffen, ohne molekulare Systeme durch den Gebrauch von kontinuierlichen Variablen zu approximieren wobei der diskrete Charakter berücksichtigt wird.

Die Anwendung des Press Modells auf Selbstreplikatoren zeigt einen durchweg negativen Einfluß auf die Stabilität (Reduzierung der kritischen Fehlerschwelle) der Population im Vergleich zu den homogenen Quasispezies. Dabei ist die Erniedrigung umso stärker, je deutlicher der Einfluß des Raums wird. Anhand von stochastischen Simulationen, die die analytischen Ergebnisse verifizieren, wird die Destabilisierung auch in physikalischen Dimensionen bestätigt.

Das Studium des einfachsten, wechselwirkenden Systems von altruistischen Katalysatoren stellt sich hingegen in der Gegenwart des Raums völlig anders dar. Zeigen Rechnungen mit gewöhnlichen Differentialgleichungen im homogenen Fall unter dem Einfluß von Mutationen keine Resistenz des Gesamtsystems gegen Parasiten, so belegt das Press Modell eine Stabilisierung für gewisse Parameterbereiche. Die Aussagen können mit Simulationen erneut überprüft und auf physikalische Raumdimensionen extrapoliert werden, wobei sich der gezeigte Effekt bestätigt.

Als Methode, die Stabilität katalytischer Systeme zu erhöhen, wird die partielle Sequenzerkennung untersucht. Es stellt sich heraus, daß in diesem Kontext die Sequenzfaltung eine wichtige Rolle spielt und in der Lage ist, die Erkennung im Vergleich zum linearen Modus (keine Faltung) deutlich zu verstärken. Anschließende Rechnungen zeigen, daß mit Hilfe der Faltungsabbildung eine signifikante Anhebung der Fehlerschwelle und somit eine Sequenzverlängerung möglich ist.

Das letzte Kapitel widmet sich dem Parallelrechner Ngen, der eine effiziente Simulation molekularer Reaktionen erlaubt und in der Lage ist, sehr große Datenmengen in kurzer Zeit abzuarbeiten. In diesem Zusammenhang wird ein Diffusionsalgorithmus vorgestellt, der zur parallelen Diffusion benutzt werden kann und die Diffusionsgleichung erfüllt.

Eine Zusammenfassung der Ergebnisse führt zu der Schlußfolgerung, daß nicht der Hyperzyklus die nächst höhere Organisationstufe zu den Darwinschen Systemen bildet, sondern der einfache (altruistische) Katalysator.

English abstract

The thesis deals with the question in how far spatial resolution influences the stability of biomolecular, evolvable systems.

Therefore a model is derived - the Press model - that allows for analytical calculations of molecular systems while taking the discrete character of biomolecules into account.

The application of the Press model onto molecular selfreplicators shows a monotonic decrease in the maximal sustainable error probability with decreasing diffusion coefficient. Stochastic simulations verify these results and prove them to be independent of the spatial dimension.

The study of the simplest interacting system of altruistic catalysts show a completely different behaviour in the presence of space. In contrast to homogeneous calculations with ordinary differential equations where no resistance against parasites can be found the Press model predicts stability of the systems for parameter ranges. This stability is again verified with stochastic simulations which also allow extrapolations to physical dimensions.

For the catalytic model partial sequence recognition is studied. In this context sequence folding turns out to be an efficient method to enhance the stability compared to the moderate increasement that linear recognition (without folding) provides. Subsequent calculations show that with the use of partial sequence recognition of folded molecules sequence extensions are possible.

The last chapter gives an introduction to the parallel computer Ngen with which efficient and fast simulations of a large number of molecules can be performed. Furthermore, a parallel diffusion algorithm is presented that fulfills the diffusion equation.

A discussion is given that resumes the results and compares both the models. Summarized, the conclusion can be given that not the hypercycle is the next higher level of organization of Darwinian systems. It is the simple (altruistic) catalyst.

Danksagungen

An dieser Stelle soll all denen gedankt werden, die zum Gelingen dieser Arbeit beigetragen haben.

An erster Stelle möchte ich meinen Eltern danken, deren durchgehende Unterstützung mich in die Lage versetzt haben, diese Arbeit zu schreiben. Ebenso meiner Freundin Indira Duda, deren liebevolle Pflege mir über die harte Zeit des Zusammenschreibens geholfen hat.

Ein besonders herzlicher Dank gilt meinem Betreuer Prof. John McCaskill. Er hat mir nicht nur die Möglichkeit der Themenbearbeitung sondern darüber hinaus auch immer das Gefühl gegeben, gute und sinnvolle Arbeit zu leisten. Sein Ideenreichtum, die intensive Unterstützung und eine enge Zusammenarbeit haben mir diese Arbeit überhaupt erst ermöglicht. Darüber hinaus gilt ihm mein Dank, daß er mir die Freude an der Forschung wieder vermittelt und mich zu einem ernsthaften Wissenschaftler ausgebildet hat.

Weiterer Dank gilt meinem weiteren Betreuer Prof. Michael Lässig, der sich sofort bereit erklärt hat, diese Arbeit an der Universität Köln zu vertreten. Zu jedem Zeitpunkt konnte ich mich auf ihn verlassen und er stand mir helfend mit Rat und Tat zur Seite.

Nicht vergessen möchte ich Dr. Ruedi Füchslin, der wohl beinahe jede Frage dieser Welt beantworten kann und der mir sehr viel über die Physik beigebracht hat. Die vielen anregenden und erkenntnisreichen Diskussionen haben viel zum Gelingen dieser Arbeit beigetragen.

Ebenso möchte ich Thomas Maeke, dessen Zauber-Perl-Skripts so manchen Nervenzusammenbruch vereitelt haben und dessen Ruhe ich mir oft zum Vorsatz genommen habe und Thomas Rücker danken, der viele der physikalisch motivierten Ideen kritisch hinterfragt hat und mir bei unzähligen Fragen zur Molekularbiologie weiterhelfen konnte. Das Gleiche gilt auch für Jörg Ackermann und Thomas Kirner, die beide für Fragen jeglicher Art offen waren.

Desweiteren sei der ganzen BioMIP-Gruppe für die entspannte Atmosphäre und der guten und freundschaftlichen Kooperation gedankt.

Last but not least sei auch die ehemalige Arbeitsgruppe komplexe Systeme genannt, wobei ich im einzelnen Prof. Thomas Martinetz für das freundschaftliche Klima, Dr. Claus Wilke für die gemeinsame Planung des GWALs und Christopher Ronnewinkel für die heiteren Diskussionen danken möchte.

Erklärung

Ich versichere, daß ich die von mir vorgelegte Dissertation selbständig angefertigt, die benutzten Quellen und Hilfsmittel vollständig angegeben und die Stellen der Arbeit - einschließlich Tabellen, Karten und Abbildungen -, die anderen Werken im Wortlaut oder dem Sinn nach entnommen sind, in jedem Einzelfall als Entlehnung kenntlich gemacht habe; daß diese Dissertation noch keiner anderen Fakultät oder Universität zur Prüfung vorgelegen hat; daß sie - abgesehen von unten angegebenen Teilpublikationen - noch nicht veröffentlicht worden ist sowie, daß ich eine solche Veröffentlichung vor Abschluß des Promotionsverfahrens nicht vornehmen werde. Die Bestimmungen dieser Promotionsordnung sind mir bekannt. Die von mir vorgelegte Dissertation ist von Prof. Dr. Michael Lässig und Prof. Dr. John McCaskill betreut worden.

Teilpublikationen: [24], [56], [61], [75].