# THERMODYNAMIC PATTERNS OF LIFE
## Emergent Phenomena in Reaction Networks

JAKOB FISCHER


seit 1558

Dissertation zur Erlangung des Akademischen Grades
doctor rerum naturalium (Dr. rer. nat.)

vorgelegt dem Rat der Fakultät für Mathematik und Informatik
der Friedrich Schiller Universität Jena

von Dipl. Phys. Jakob Fischer
geboren am 25. Feb. 1986 in Saulgau, jetzt Bad Saulgau
Juni 2017

BETREUER: Prof. Dr. Peter Dittrich

GUTACHTER:

1. Prof. Dr. Peter Dittrich (Friedrich-Schiller-Universität Jena)
2. Prof. Dr. Daniel Merkle (University of Southern Denmark, Odense)

TAG DER ÖFFENTLICHEN VERTEIDIGUNG: 20. August 2018

## ABSTRACT

Reaction networks are an important tool for the analysis of complex chemical reaction systems. They help us understand systems ranging from specific metabolisms to planetary atmospheres. This thesis develops methods for the analysis of living systems by using reaction networks with a focus on the inclusion of thermodynamic properties. Existing methods for reaction pathway analysis are extended and the thermodynamics of reaction pathways is analysed to gain insight into thermodynamic structures. New methods for more realistic artificial chemistries are developed using thermodynamic constraints. A model of evolvable artificial ecosystems is created to understand the effect of evolution and life on the flow of matter and energy through the system.

To investigate general thermodynamic properties of large-scale reaction networks, artificial reaction networks are created with a simple scheme for deriving thermodynamically consistent reaction rates. Linear and nonlinear networks using four different complex network models are simulated to their non-equilibrium steady state for various boundary fluxes. The distribution of entropy production of the individual reactions of the network follows a power law in the intermediate region with an exponent of circa $-1.5$ for linear and $-1.66$ for nonlinear networks. An elevated entropy production rate is found in reactions associated with weakly connected species. This effect is stronger in nonlinear networks than in linear ones. Increasing the flow through those nonlinear networks also increases the number of cycles and leads to a narrower distribution of chemical potentials.

In the context of finding signs of life by detecting chemical disequilibrium, a photochemical model of the modern atmosphere and a model of the Archean atmosphere are compared. Calculating the reaction pathways that are most relevant for explaining their reaction network's steady state with a new method allows for the detection of topological differences between the two models. Pathways of the modern Earth atmosphere are simpler (less reactions) and contain fewer cycles than their Archean counterparts. To compare thermodynamic properties of the found pathways, values for chemical potentials of most of the species are estimated. The Archean atmosphere is shown to be driven more by radiation, generally increasing the chemical energy of the matter it exchanges with its environment. In contrast, the matter fluxes of the modern atmosphere have a negative net energy balance. An analysis of methane consuming pathways shows their importance in both atmospheres and, in agreement with previous work, allows for a quantification of the thermodynamics of methane oxidation for the modern atmosphere.

A simple artificial chemistry model that incorporates thermodynamic constraints and mass conservation is build. To emphasize the analogy to planetary systems, the system is not driven to disequilib-

rium by matter flow but by photoreactions. This technique, together with pathway analysis, is demonstrated for small example systems and is discussed in a broader context. Increasing the strength of the force that drives the system to (thermodynamic) disequilibrium concentrates the mass of the system in a smaller number of chemical species. To model the influence of life on pathways, an artificial ecosystem model is developed. An artificial ecosystem is an artificial chemistry build from a fixed anorganic part and interfacing organisms. Organisms are evolved by successively replacing the organism that is least active, measured by its mass and its reaction rates. Evolution of the reaction networks entails an evolution of reaction pathways towards simplicity, thus indicating that the presence of pronounced, relatively simple pathways in real systems is a consequence of an evolutionary mechanism.

All in all, this thesis is meant to show how to analyse and characterise complex reaction systems using a network-oriented view on thermodynamics, thereby informing on the effect of life on chemical disequilibrium and the complexity of associated reaction networks.

## ZUSAMMENFASSUNG

Reaktionsnetzwerke sind ein wichtiges Werkzeug zur Analyse von komplexen chemischen Reaktionssystemen. Sie helfen beim Verständnis von Systemen der Größenordnung eines spezifischen Metabolismus bis hin zu Atmosphärenchemien. In dieser Abschlussarbeit werden Methoden zur Modellierung und Analyse lebender Systeme durch Verwendung von Reaktionsnetzwerken und die Einbeziehung thermodynamischer Eigenschaften entwickelt. Existierende Methoden zur Analyse von Reaktionspfaden werden erweitert und die Thermodynamik von Reaktionspfaden analysiert, um Einblick in thermodynamische Strukturen zu gewinnen. Es werden neue Modelle für künstliche Chemien unter Einbeziehung thermodynamischer Randbedingungen entwickelt. Das Modell eines künstlichen Ökosystems wird entwickelt, um den Einfluss von Evolution und Leben auf chemische Fließstrukturen besser zu verstehen.

Zur Untersuchung allgemeiner thermodynamischer Eigenschaften in großskaligen Reaktionsnetzwerken werden künstliche Reaktionsnetzwerke mit thermodynamisch konsistenten Reaktionsraten erzeugt. Lineare und nichtlineare Reaktionsnetzwerke mit vier verschiedenen Modellen komplexer Netzwerke werden für verschiedene Randbedingungen in das Fließgleichgewicht simuliert. Die Verteilung der Dissipation der Einzelreaktionen des Netzwerks folgt einem Potenzgesetz mit einem ungefähren Koeffizienten von $-1.5$ für lineare und $-1.66$ für nichtlineare Netzwerke. Für schwach verknüpfte chemische Spezies wird eine vergleichsweise erhöhte Dissipation festgestellt. Der Effekt ist bei nichtlinearen Netzwerken stärker als bei linearen Netzwerken. Ein Erhöhen des Flusses durch nichtlineare Netzwerke erhöht die Anzahl der Zyklen und führt zu einer engeren Verteilung der chemischen Potentiale.

Im Kontext der Detektion von Leben durch chemisches Nichtgleichgewicht wird ein photochemisches Modell der modernen Atmosphäre mit einem entsprechenden Modell der archaischen Atmosphäre der Erde verglichen. Durch das Berechnen der Reaktionspfade für das spezifische Fließgleichgewicht können Unterschiede zwischen den Modellen gefunden werden. Reaktionspfade der modernen Atmosphäre sind einfacher (weniger Reaktionen) und enthalten weniger Zyklen als ihre archaischen Gegenstücke. Eine Abschätzung der chemischen Potentiale für die chemischen Spezies erlaubt eine genauere thermodynamische Charakterisierung der Reaktionspfade. Es zeigt sich, dass die archaische Atmosphäre stärker von Strahlungsenergie getrieben ist, welche in ihr in chemische Energie umwandelt wird, während die Materieflüsse der modernen Atmosphäre eine negative Energiebilanz haben. Die Analyse der Methan konsumierenden Reaktionspfade zeigt deren Bedeutung in beiden Atmosphären und erlaubt es, die Thermodynamik der Methanoxidation für die moderne Atmosphäre in Übereinstimmung mit existierenden Publikationen zu quantifizieren.

Ein Modell mit thermodynamischen Nebenbedingungen und Massenerhaltung wird erstellt. Um die Analogie zu planetaren Systemen hervorzuheben, wird die Chemie nicht durch Massenfluss, sondern durch Photochemie mit Energie versorgt. Dieses Modell wird zusammen mit einer Reaktionspfadanalyse an kleinen Beispielsystemen präsentiert und für den weiterreichenden Kontext diskutiert. Eine stärkere treibende Kraft in das thermodynamische Nichtgleichgewicht führt zu einer Konzentration der Masse des Systems bei weniger chemischen Spezies. Um den biologischen Einfluss auf Reaktionspfade zu modellieren, wird ein Modell für künstliche Ökosysteme entwickelt. Das bezeichnet eine künstliche Chemie, die aus organischen und anorganischen Modulen assembliert wird. Die organischen Komponenten (Organismen) werden durch ein einfaches Evolutionsschema evolviert. Dies induziert eine Evolution auf Ebene der Reaktionspfade hin zu einfacheren Reaktionspfaden. Ein Auftreten von herausgehobenen, einfachen Reaktionspfaden in realen Netzwerken kann dementsprechend als Folge eines evolutionären Prozesses verstanden werden.

Insgesamt zeigt diese Arbeit also, wie komplexe Reaktionssysteme mit einem netzwerkorientierten Blick auf ihre Thermodynamik unterschieden werden können. Dies ergibt einen neuartigen Blick auf den Effekt des Lebens auf das thermodynamische Nichtgleichgewicht und die Komplexität der assoziierten Reaktionsnetzwerke.

# PUBLICATIONS

Some ideas and figures presented in this thesis have appeared previously in the following publication:

[FKD15] J Fischer, A Kleidon, and P Dittrich. "Thermodynamics of Random Reaction Networks". In: *PLOS ONE* 10.2 (2015), e0117312. DOI: 10.1371/journal.pone.0117312.

Parts of the software used and presented in this work has been explicitly released previously:

[Fis16a] J Fischer. *jrnf_R_tools: A tool for handling of reaction network analysis and simulation*. Version 1.0.0. Dec. 2016. DOI: 10.5281/zenodo.197056.

[Fis16b] J Fischer. *jrnf_int: Integrator for reaction networks in jrnf format*. Version 1.0.0. Dec. 2016. DOI: 10.5281/zenodo.197058.

[Fis16c] J Fischer. *jrnf_tools: Command line tool for generating large random reaction networks*. Version 1.0.0. Dec. 2016. DOI: 10.5281/zenodo.197060.

Besides those releases and the current version of the software contained in the digital appendix of this thesis, the development chronology can also be followed on GitHub through the repositories *jrnf_int*, *jrnf_R_tools* and *jrnf_tools* on the page of the author[1]. The source code of tools and scripts used are also contained in the digital appendix (data medium) of the thesis (Appendix C).

[1] http://www.github.com/jakob-fischer/

*One in ten thousand of us can make a technological breakthrough capable of supporting all the rest. The youth of today are absolutely right in recognizing this nonsense of earning a wage.*

— R Buckminster Fuller

## ACKNOWLEDGMENTS

I would like to thank many people, who have generously contributed ideas, supervision and support to the work presented in this thesis.

First and foremost I want to thank my enthusiastic supervisor, Peter Dittrich. My PhD has been an amazing experience and I thank Peter wholeheartedly, not only for his tremendous academic support, but also for giving me so many wonderful opportunities.

My second supervisor Axel Kleidon deserves credit for awakening my interest in investigating a thermodynamic perspective to Earth system science. Parts of my better understanding of atmospheric chemistry presented in this thesis are thanks to discussions with people at ELSI in Tokyo. In this context I thank especially Nathaniel Virgo and Matthieu Laneuville for fruitful discussions. For helping me with calculating chemical potentials of atmospheric species I thank Eugenio Simoncini. The work presented in the chapter on artificial ecosystems has been developed in a one year stay in the lab of Takashi Ikegami. I thank him for giving me this great chance and all members of his lab for making me feel welcome.

For making science an overall social experience a big thanks goes to all my (temporary) labmates and officemates: Bashar Ibrahim, Gabi Escuela, Gerd Grünert, Jan Huwald, Julien Hubert, Kanjin Yoneda, Konrad Giżyński, Lana Sinapayen, Martin Biehl, Masanori Miyake, Matthew Egbert, Nelly Mostajio, Norihiro Maruyama, Olaf Witkowski, Peter Kreyßig, Richard Henze, Stephan Richter, Yhoichi Mototake .... Similar credit also goes to all the kind and interesting people I met at conferences, winter schools, summer schools and scientific workshops.

Last but not least I want to thank all those people that have helped me in areas besides science and make this planet a mostly pleasant experience: friends, family, strangers.

**Financial support:** The research presented in this thesis was financially supported by the International Max-Planck Research School for global Biogeochemical Cycles (IMPRS-gBGC), by a "Jahresstipendium für Doktoranden" of the German Academic Exchange Service (DAAD) and the Friedrich Schiller University Jena.

**Tools / software:** This document was typeset using the typographical look-and-feel `classicthesis` developed by André Miede. The

style was inspired by Robert Bringhurst's seminal book on typography "*The Elements of Typographic Style*". `classicthesis` is available for both LaTeX and LyX.[1]

[1] https://bitbucket.org/amiede/classicthesis/

Much of the work presented here would not have been possible without a wide variety of open source software: Systemd (init system), Git (version control), Gcc (compiler), R (data analysis), Inkscape (vector graphics), igraph (graph library) [CNo6b], . . . . It seems impossible to name all software that was utilized for this work. Many thanks to all those people that create this free and open infrastructure.

# CONTENTS

# LIST OF FIGURES

## LIST OF TABLES



## ACRONYMS

AL artificial life

AC atmospheric chemistry

AC artificial chemistry

BIF banded iron formation

CS computer science

DNA deoxyribonucleic acid

EA evolutionary algorithm

ESS Earth system science

GPP gross primary production

LUCA last universal common ancestor

MEPP maximum entropy production principle

MIF mass independent fractionation

MIF-S mass independent fractionation of sulfur

NPP net primary production

GOE great oxidation event

ODE ordinary differential equation

OoL origins of life

PAL present atmospheric level

ppm parts per million

RN reaction network

RNA ribonucleic acid

rTCA c. reductive Tricarboxyl acid cycle

TCA c. Tricarboxyl acid cycle / citric acid cycle

TD thermodynamics

# 1 INTRODUCTION

This thesis aims to bring forward methods for analysis and modelling of living and planetary systems. Approaches from different fields to life, its origins, and its definition are recapitulated in the introduction. An overview of open problems in these areas is given to motivate the approach of this thesis to thermodynamics, reaction networks and reaction pathways.
**keywords:** origins of life, history of life, Earth, atmosphere, life detection, self organisation

## 1.1 FOREWORD AND MOTIVATION

The last 50 years have brought great progress in our understanding of Earth's climate, chemistry and history. Isotope measurements have greatly enhanced our possibilities to look back in time while satellite observation allows us to obtain highly resolved data on the status quo. The growing amount of diverse data, together with the growing computational capabilities, has led to models of planetary dynamics with continuously increasing complexity. The topic of anthropogenic global warming is the most urgent case, but it is now undisputed that life has modified the planet long before [Kno03]. Thus, understanding planetary dynamics and its connection with the biosphere is not only important to solve the biggest current challenge of human civilisation, climate change, it also might be essential for understanding how life itself originated and influenced our planet's history.

Models currently used for climate simulations consist of many submodels that are developed by different groups throughout many years of work. Different computer languages and data interfaces render a consistent analysis of connected processes spanning multiple submodels difficult. At the same time, increases in data size and computing power lead to even more complex submodels. Up to now, complex networks have been used frequently to analyse model output in Earth system science; these networks, however, mainly work on a statistical level and do not directly resolve flow of matter and energy [Don+09].

The aim of this work is to bring forward methods for the analysis and modelling of living systems. The focus lies on methods that work on the conceptual level of process models[1]. In comparison to purely data-oriented models, this includes implicit constraints of thermodynamics (energy, matter) on all levels. To achieve this, data and models are formalised into reaction networks in which emergent structures can be defined and analysed (Fig. 1.1). This approach can be called unconventional in the sense that it takes methods that are com-

[1] *Process models, in contrast to statistical model, describe dynamics in terms of physical processes.*

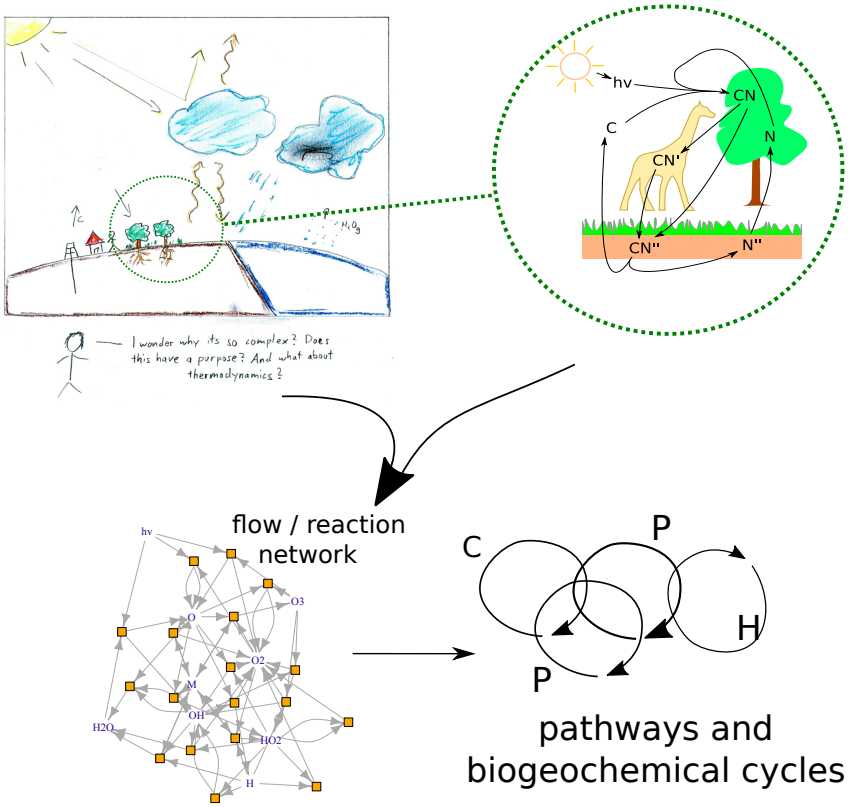



Figure 1.1: **Biosphere and its environment form complex flow patterns that can be analysed by reaction networks.** Planetary systems consist of complex and versatile dynamics and flow patterns. Even more if the planet contains a biosphere **(top)**. All flows can be understood as parts of one big reaction network **(bottom-left)**. This thesis proposes to understand the emergence of living structures as emergent phenomena in such networks that can be formally analysed by methods like pathway analysis **(bottom-right)**.

monly used in bottom-up approaches (artificial chemistries, reaction networks) and uses them on a larger scale (top-down).

This thesis touches many different fields of research. The methods of modelling and analysis stem from thermodynamics, complex networks, and bioinformatics. The general approach, however, can probably best be attributed to the field of artificial life [Lan97; Bed+00; Bed03; Ras+04]. In contrast to modelling techniques in the fields of chemistry, ecology or biology, models in artificial life are qualitative by choice. This means that models are conceptual in nature, and thus not directly meant to explain existing data but to give a better understanding of the basic concepts of life. This goes as far as the famous claim of investigating not only "life-as-it-is", but also "life-as-it-could-be" [Lan97].

*An example for artificial life's potential can be seen in the work of Adami, Ofria, and Collier [AOC00] on simulated artificial ecosystems.*

The many fields and approaches relating to phenomena of life and the planets that support it will shortly be summarized in the following. Necessarily, some of the remarks might seem incomplete or repetitive, considering the extent of the fields of research connected to life as well as the interdisciplinary approach of this thesis.

## 1.2 LIFE

### 1.2.1 *Complexity and Life*

With his short book titled "What is life?" [Sch43] Schrödinger arose awareness for the relationship between physics and life. He used first principles and statistical arguments to theorise on the relevance of an information bearing molecule, the DNA discovered years later.

Yet, there is still no widely accepted answer to the question in the title of Schrödinger's book. Life is hard to define and for many possible definitions there exist striking counterexamples [CC02; Ben10]. The common biological definitions see an organism as alive if it can metabolize, grow, react to stimuli, and is able to reproduce [Fis01]. NASA's working definition for astrobiological purposes is driven by the type of primitive life that might be found by its own missions on other planets. It states: "Life is a self-sustained chemical system capable of undergoing Darwinian evolution." [RMPM04] The problem in both cases is, that one has to look at the dynamics of the system to verify its evolution. Also, a biological species might need a specific environment and also a mate to reproduce, which would mean, that by this definition one lion in a cage would not be alive.

A popular definition in artificial life is to think of life as an autopoietic system. An *autopoietic system* is a network of processes of construction and destruction that continuously regenerates its components [VMU74; MV87]. The abstract nature of this definition is its strength and weakness at the same time. It is less biased from life forms we know, but it is hard to apply operationally. Ruiz-Mirazo, Peretó, and Moreno [RMPM04] see the autopoietic approach as an alternative (metabolism focused) to the biological (evolution focused) and try to unify them. They define a universally living being as an autonomous system that is capable of open ended evolution.

Life is often said to be complex. And as for life, there is no unique definition for complexity. It is undoubted, that there are way more complex chemical structures (size of molecules) in biological species than in non-living systems [WI99]. Very complex macromolecules are synthesized in biology. Molecules process information and regulate the metabolism to allow life to exist far from thermodynamic equilibrium. In artificially evolving systems complexity has been measured by the amount of information that the genome contains about the environment of the organism. These simulations evolved towards more complexity, if the environmental conditions were fixed [AOC00; Ada02].[2]

[2] *For coevolving species complexity could theoretically decrease.*

Complexity of living structures can also be defined from an ecosystems perspective. The species diversity and the network of feeding relationships between species has been shown to be an interesting property of ecosystems [BPW06; Duf+07]. From a philosophical standpoint, the emergence of new phenomena, which need new theories for explanation, also is an indicator for complexity [Sol16]. All biological matter follows laws of physics and chemistry, but current

physical and chemical theories are in no way adequate to explain the functioning of life.

*The connections between information (computer science) and fundamental questions of life have existed early on, on remarkable example being Von Neumann's work on self-reproducing automata [VNB66].*

Statistical aspects of Schrödinger's work are extended by recent information theoretical approaches for the understanding of life. They focus on finding measures that distinguish the causal structure of living beings. This is connected with finding how computation happens in dynamical systems [WD13] and the concept of agents [Fri10]. The emergence of computation has been connected to life before. Langton showed that computation emerges in dynamical systems close to chaos and related this to the concept of phase transition in thermodynamics [Lan90].

### 1.2.2 *Origins of Life*

[3] *Plural is used as it is unclear if there are multiple ways for life to emerge. The process is also called* abiogenesis.

Even more than the definition of life, explaining its origins[3] is a common challenge in many disciplines [Sch+15]. A good review of important controversies in this area has been written by Peretó [Per05], a more extensive summary is given by Ruiz-Mirazo, Briones, and Escosura [RMBE14].

One main problem with explaining the origin of life lies in the relationship between replication and metabolism. The metabolism of cells relies on proteins, large macromolecules that act as specific catalysts and thus allow cells to persist and grow far from thermodynamic equilibrium. Proteins are biologically constructed using information encoded in genes. For the genetic information to generate a metabolic network requires evolution, which requires replicating cells with an existing metabolism. This has lead to different hypothesis on which process came first, consequently named *replication-first* and *metabolism-first*.

One hypothesis that might resolve this dilemma is called *RNA world*. It assumes that instead of the DNA found in modern life, early life used RNA as information carrier [CSM07; Cec12]. This is possible because RNA can not only encode digital information, but also can operate as catalyst directly. The disadvantages of RNA are that it is harder to synthesize and less stable than DNA [WD13].

In artificial life there is numerous work that models artificial chemistry for a better understanding the origins of life. Much effort has gone into trying to explain how simple chemistry develops towards self maintaining subsets [Per14; JK98; BF92; Kau86]. But there have also been efforts to approach the question of how primitive forms of evolution might evolve and interact with basic chemistry [ES78; CHS17].

Another interesting hypothesis from Smith and Morowitz argues that life started with the core anabolic metabolism of the reductive tricarboxyl acid cycle (rTCA c.) that was forced into existence by the reducing atmospheric conditions of the early Earth [SM16; SM04a; MS07].[4] The same set of reactions operating in reverse direction are common in heterotrophic organisms to break down organic

[4] *This implies a certain relation to biogeochemical cycles (Sect. 1.4.2).*

molecules for energy. The latter process is then called Krebs cycle or tricarboxyl acid cycle (TCA c.).

Through fossils there is numerous evidence for life on early Earth [Scho6]. Earliest signs of life on land have been found in hot spring deposits. They are estimated to have lived 3.5 million years ago [Djo+17]. Life in oceans reaches even further back. The oldest signs of life on Earth to date have been found in hydrothermal vent precipitates, estimated 4.28 million years old [Dod+17]. Hydrothermal systems are a popular theory for the origin of life because of the high availability of energy in these places [SRK11]. A recent study has found the 355 probable genes of the last universal common ancestor (LUCA)[5] from genetic information of living species. The corresponding proteins indicate a metabolism that could have evolved in hydrothermal vents [Wei+16].

[5] *This refers to the last common ancestor of all organism alive today.*

Even if the probable mechanism for the emergence of life on Earth will be found, this will not answer questions of determinism to life. Was Earth determined to develop life or was this a happy coincidence? Could there have been other planetary trajectories, maybe even alternative trajectories that include a different mechanism of abiogenesis [Sch+15]? For answering these question not only the transition from inorganic chemistry has to be better understood. It also requires a good understanding of planet formation and evolution.

*The effect of possible planetary trajectories in feedback with biogenesis is discussed in Chopra and Lineweaver [CL16]. (Cf. Sect. 1.3.2)*

## 1.3 ATMOSPHERE AND DAISYWORLD

### 1.3.1 *Earth's Atmosphere*

An important factor for life on Earth is the composition and chemistry of its atmosphere [YD98]. The modern atmosphere is mainly composed of nitrogen (78.09%), oxygen (20.95%), argon (0.93%), and, water vapour ($\approx$ 1%), with minor traces carbon dioxide (0.039%). The troposphere is the lowest part of Earths atmosphere and contains about 80% of the mass of the atmosphere. The troposphere extends to around 10 to 17km height. It is characterised by decreasing temperature and strong convective mixing. In the stratosphere, which is above the troposphere from around 15 to 50km altitude, vertical mixing is reduced because of temperature increasing with height. Reduced density and more radiation with higher energies leads to formation of the ozone layer.

*Both Mars and Venus have atmospheres that consist of more than 90% of $CO_2$. It is removed on Earth by photosynthesis.*

Progress in geology and modelling in the last decades has allowed to constraint atmospheric conditions [Holo6; CKo7]. Atmospheric oxygen levels were below $10^{-4}$ of the present atmospheric level (PAL) until 2.4 billion years ago. Early Earth's oxygen concentration can be constrained by found mass independent fractionation of sulfur (MIF-S) and sedimentary rock called banded iron formation (BIF).

Between 2.4 and 2.0 billion years ago the great oxidation event (GOE) lead to a increase of oxygen to roughly one tenth of present levels. Oxygen producing cyanobacteria have existed before that, but the oxygen was insignificant in comparison to the reducing atmo-

sphere. The atmospheric oxygen and the oxidation state of the oceans reached modern levels around 600 million years ago. The exact trajectory and mechanisms of this transition is still not conclusively resolved [GLW06; Geb+17; LRP14; HTS15]. It is assumed that the modern atmospheric oxygen concentration plays an important role for the development of complex terrestrial life [Spe+13].

### 1.3.2 *Gaia and Daisyworld*

Until the middle of the last century the common scientific understanding was that environmental conditions, especially atmospheric composition, was something that was determined by geological processes. Biology was understood as a lucky beneficiary of climatic conditions and atmospheric oxygen concentration. This was famously questioned by Lovelock's *Gaia hypothesis*, suggesting that Earth's biota influence their environmental conditions and as a part of a complex planetary system self-regulates their environment into planetary homeostasis [LM74].[6]

[6] *This work built on earlier work proposing weaker forms of atmospheric control [LL72; ML74].*

For supporting this hypothesis, Daisyworld was created, a simple model that describes daisies of different colours and physiological parameters. Together they stabilize the planetary temperature against changing solar irradiation [WL83; Wood+08]. One main criticism of Daisyworld (and Gaia) concerns its relation to the theory of evolution [Sau94; Len98; LL00]. If biological species can actively modify their environment, what stops rebel organism from evolving that either do not contribute to environmental regulation or even regulate the environment away from the preferred conditions of their competition? Some artificial life models were created to tackle this, and it seems that the constraints between metabolism (growth) and environmental regulation (by-product) is decisive for the outcome of the simulations [DZ99; Dow02; WL07b].

Today it is widely accepted that the biospheres influences feedback processes controlling its environment, but the extend of this control and the question of a physical determinism towards such feedback loops (and towards life) is still not really resolved [Kno03].

Considering biotic regulation of environmental conditions also leads to interesting questions in relation to habitability and planetary evolution. If persistence of life on a planet depends on it managing to regulate environmental conditions, the sparsity of inhabited planets might not be due to the difficulty for emergence of life, but because a low probably of life managing to regulate its environment (Gaian Bottleneck) [CL16].

*Increasing radiation of stars during their lifetime means that conditions allowing life to emerge and persist (without biotic regulation) are vanishing after some time.*

### 1.3.3 *Habitability and Remote Life Detection*

While there are various measurement methods from spacecrafts that would have informed us on existing life on other planets in our solar system [Sag+93], conceiving concepts for remote life detection on exoplanets is way more difficult. There is not only a limit on the

observations that can be made, but also a big diversity in stars and planets [Sea13]. It is common to assume that possible life has to be carbon based, similar to life on Earth, and thus requires a planet to have liquid water to be habitable [Kas97]. With this assumption, habitable zones around stars are defined, with their boundaries depending on the radiative flux generated from the star.

Besides the distance of an exoplanet from its star, there are other factors to consider. Depending on bulk composition and geology, greenhouse gasses in the atmosphere can increase the atmospheric temperature. Atmospheric pressure influences the temperature at which liquid water vaporizes. The size of the planet determines for which time span of the planet's evolution plate tectonics is maintained [GRP13; Rau+13]. Plate tectonics is important as it cycles volatiles like $CO_2$, helping to regulate atmospheric temperature. It is also connected to the formation of the magnetic field of planets which protects the atmosphere from stellar plasma particles [Lam+09].

In this conundrum of uncertainties, it is pragmatic and common that astrobiological research focuses on exoplanets and scenarios for which the existence of life could be confirmed with high certainty, should it exist. One example for this is the modelling of super-Earth atmospheres and their spectra around M-dwarf stars [Rau+11; Gre+13].[7]

[7] *Resolving planetary spectra indirectly is easier for smaller stars.*

## 1.4 GLOBAL PERSPECTIVE

### 1.4.1 *Thermodynamics of Planets and Life*

*The books of Douce [Dou11] and Kleidon [Kle16] give a good introduction in this interesting area.*

Roughly 50 years ago James Lovelock proposed to use atmospheric thermodynamic disequilibrium as a general way of remote life detection [Lov65; HL67]. This idea is based on the common thermodynamic standpoint that Earth, like any dissipative structure [KP98], self organizes in a way that maximizes entropy production by reducing its entropy at the expense of its environment. By this argument, the thermodynamic disequilibrium in the atmosphere manifested through the coexistence of oxygen and methane should thus be seen as an indicator for the biosphere.

As striking as this idea is in its simplicity and its generality[8], nowadays it is widely accepted that atmospheric thermodynamic disequilibrium is necessary, but not sufficient for finding life on a planet [Sag+93]. The power maintaining thermodynamic disequilibrium has to be distinguished from the magnitude of the disequilibrium and possible inorganic sources have to be considered [SVK13]. Recent work also suggests that ocean chemistry has to be considered to quantify atmospheric chemical disequilibrium [KTBC16].

[8] *The method, if successful, would also work for non-carbon-based life.*

In spite of such difficulties, thermodynamics gives interesting insights into Earth's energetic structure [Kle10a] and the role of life therein [Kle10b]. It allows to characterize the evolution of free energy production throughout our planets history [LPW16] and to put life into a cosmological perspective [LE08].

Many recent applications of thermodynamics in Earth system science are related to the *maximum entropy production principle* (MEPP) [Oza+03; DK10]. This proposed extremal principle assumes that high dimensional systems out of thermodynamic disequilibrium self organise to maximise their entropy production [MS06; VSGCU11]. This principle has been explored for mixing processes in the atmosphere, but also for other parts of the physical climate system [Kle09].

Using the MEPP, ecosystems can be treated as self organizing molecular machines that adapt to maximise entropy production [Val10]. Even though thermodynamic methods have a long history in ecology [Odu68; MM93], it is hard to assign chemical potentials to organisms [MB07]. Thus, much of the work done on thermodynamics in this area is related to theoretical models [Mic05; Mic12]. For example, in case of the daisyworld model, controlling heat flux by MEPP leads to ideal homeostatic control. [Dyk08].[9]

[9] *See Sect. 1.3.2 for an explanation of daisyworld.*

An important reason to have thermodynamic theory for living systems is related to the hypothesis of life evolving towards more complexity. If this is true it is reasonable to expect a relationship with the *arrow of time* that the second law of thermodynamics provides.

### 1.4.2 *Ecology and Biogeochemical Cycles*

Contrary to thermodynamics of planetary systems, many classical disciplines investigating the biosphere often start with observations on a small scale. In *ecology* these are for example observations on biological individuals, counting them and measuring specific properties. From these, spatial and temporal patterns can be analysed on different scales, and explanations for their occurrence can be tested through modelling [Lev92; Lev98].

*Originally a part of biology,* ecology *has evolved into a field of its own [Bar01].*

While ecology focuses on biological organisms, their connection with each other and their environment, the focus of *biogeochemistry* lies on the matter cycles that are related to the biosphere.[10] Examples for basic observations are measurements of atmospheric fluxes and concentration for various chemical species. These observations are then combined into an integrated view [Gor91; Hed92].

[10] *There is a considerable overlap between those two fields.*

*Biogeochemical cycles* are collections of processes that lead to chemical elements being recycled on various spatial and temporal scales. Most important elements for biospheric modelling are carbon, nitrogen, oxygen, hydrogen, sulfur and phosphorus. Another important cycle is the water cycle. But other than with the others, a big fraction of water is cycling without being modified chemically, it only is transported and undergoes physical phase changes. The cycling of the different components is connected. For example, the fixation of inorganic carbon into organic matter is limited by nitrogen and phosphorus [WLP10]. The water cycle influences the transport of other elements through rivers [Auf+11]. Matter cycles are connected dynamically to each other and the climate system through various feedback loops. In light of the anthropogenic climate change, models

that describe natural fluctuations and sensitivity to disturbance are of special interest [Fal+00; CI+08; Cia+14].

For the biosphere, recycling of matter is important from an energetic standpoint. Creating organic matter, or also just precursors for it, is energetically expensive and might need complex machinery that the biological organisms can not provide. Theoretical considerations show that cycles necessarily exist in (biological) systems which are driven into thermodynamic disequilibrium [Mor66]. The amount of cycles has also been show to have an influence on thermodynamic efficiency of cyclic processes [Lay+12]. Even if ecology and biogeochemistry puts much effort into small scale observations, these observations and their interpretation is connected to thermodynamics by energy and mass balance.

## 1.5 THESIS STRUCTURE

*To guide the reader, each chapter also starts with a short summary, keywords and a side note on recommended chapters for prior reading.*

In Chapter 2 (METHODS) the formal methods used in this thesis are presented. It communicates knowledge of nonequilibrium thermodynamics and reaction networks. A new method for calculating the pathway decomposition of a reaction network and its steady state is introduced. Big random reaction networks are generated in Chapter 3 (THERMODYNAMICS OF REACTION NETWORKS) and simulated with thermodynamic constraints. Their topological structure (cycles) is investigated as a function of network flow and the internal distribution of dissipation in the networks is analysed.

In Chapter 4 (ATMOSPHERIC REACTION PATHWAYS) two chemical models of Earth's atmosphere in different states of its planetary evolution are compared by using reaction pathways. A tendency towards simpler pathways in the modern atmosphere is found. Thermodynamic characterisation reveals that the matter flow across the boundary of the modern atmosphere is stronger driven by chemical energy (of the matter) while the Archean atmosphere is more driven by photochemical reactions (radiation). To get an integrated view on pathways in a thermodynamical setting, in Chapter 5 (PATHWAYS IN ARTIFICIAL ECOSYSTEMS) artificial ecosystems are generated and evolved by a simple selection process. It is shown that evolution leads to less complex pathways (less reactions, less cycles) explaining the pronounced biogeochemical cycles found on Earth. After giving a short conclusion on the main part of this thesis in Chapter 6, a short outlook with ideas for reaction networks in planetary science and artificial life is given in Chapter 7.

*Additional information is contained in the appendix. The attached disc also contains scripts and simulation data in digital form.*

# 2 METHODS

Methods of non-equilibrium thermodynamics and formal methods of reaction networks are introduced. A new approach for calculating important reaction pathways for a given steady state of a reaction network is presented. Additionally, a method for calculating a uniquely defined (for a specific set of pathways) pathway decomposition is introduced. This method is especially well suited numerically if steady state rates span many orders of magnitude.
**keywords:** reaction network, thermodynamics, pathway analysis, elementary modes, graphs, thermodynamics

## 2.1 BASIC THERMODYNAMICS

Thermodynamics is a field of physics that is concerned with the study of energy and its transformations, especially in the context of macroscopic systems. Its development is connected to the rise of the steam engines in the 18th century. Today, concepts of thermodynamics are used in a wide range of fields, from biology to planetary science. In the following section, a very short introduction into the subject is given roughly following the book of Kondepudi and Prigogine [KP98].

*Douce [Dou11] also gives an introduction in the field with focus on planets and life.*

### 2.1.1 *Equilibrium Thermodynamics*

In thermodynamics there are different types of systems, depending on their interaction with their environment. *Isolated systems* do not exchange heat or matter with their environment. *Closed systems* do not exchange matter with their environment, but they can exchange heat and work with it. *Open systems* exchange work, matter and heat with their environment.

A system in thermodynamic equilibrium is fully specified by a set of *state variables*. All closed systems evolve irreversible towards such a state. The relation between the state variables is given by *state functions* or *thermodynamic potentials*. For closed systems the state function entropy $\mathbf{S} = \mathbf{S}(\mathbf{U}, \mathbf{V}, \mathbf{N_j})$ describes the state of the system as a function of the natural variables inner energy $U$, volume $V$, and particle numbers $N_j$. In differential form, the change of entropy is written

$$\begin{aligned} dS &= \frac{\partial S}{\partial U} dU + \frac{\partial S}{\partial V} dV + \sum_i \frac{\partial S}{\partial N_i} dN_i \\ &= \frac{1}{T} dU + \frac{p}{T} dV - \sum_i \frac{\mu_i}{T} dN_i. \end{aligned} \tag{2.1}$$

*Relationship between conjugate variables: $\frac{1}{T} = \frac{\partial S}{\partial U}$, $\frac{p}{T} = \frac{\partial S}{\partial V}$, $\mu_i = -\frac{\partial S}{\partial N_i}$*

The partial derivatives of the state function in relate conjugate pairs of potential state variables. Volume $V$ is related to pressure $p$, as well as chemical potential $\mu_i$ is related to particle number $N_i$. Conjugate pairs always consist of one intensive and one extensive variable. Extensive variables like volume and inner energy scale with the size of the system, while intensive variables like pressure and temperature do not.

*Transformation between thermodynamic potentials is done using the Legendre transformation [KP98, p. 132].*

*Relationship between conjugate variables:* $S = -\frac{\partial G}{\partial T}$, $V = \frac{\partial G}{\partial p}$, $\mu_i = -\frac{\partial S}{\partial N_i}$.

For different situations different thermodynamic potentials can be used. The Gibbs free energy $\mathbf{G} = \mathbf{G}(\mathbf{T}, \mathbf{p}, \mathbf{N_j})$ describes the work that can be extracted from a system at constant pressure $p$ and temperature $T$.

$$\begin{aligned} dG &= \frac{\partial G}{\partial T} dT + \frac{\partial G}{\partial p} dp + \sum_i \frac{\partial G}{\partial N_i} dN_i \\ &= -S dT + V dp - \sum_i \mu_i dN_i \end{aligned} \tag{2.2}$$

Using the concept of state function, the *first law of thermodynamics*, stating energy conservation in closed systems, can be formulated by saying that the inner energy $U$ is a state function with the entropy $S$ and volume $V$ as its natural variables. For its differential $dU$ this means

$$\oint dU = 0. \tag{2.3}$$

All integrals over $dU$ on a closed path are zero. This is equivalent with the statement that the path integral between two points is independent of the path chosen.

The *second law of thermodynamics* states that entropy in closed system never decreases over time. In differential form:

$$dS \geqslant 0. \tag{2.4}$$

Isolated systems in equilibrium are in their state of maximum entropy, which means that Eq. 2.4 reduces to $dS = 0$.

### 2.1.2 *Non-Equilibrium Thermodynamics*

In non-equilibrium thermodynamics fluctuations become important and interesting phenomena, such as long-range correlated structures and dissipative structures, become possible. There are no global state functions that describe the state of the system. The systems that we are focusing are locally in equilibrium, state variable like temperature and pressure can only be defined as densities. State functions can be integrated to get the entropy or inner energy of the entire system, but these integrated quantities do not fully quantify the state of the system.

A good starting point for explaining the difference to equilibrium thermodynamics (where no irreversible processes are considered) is the splitting of the entropy change $dS$ of a system into a part due to internal processes $d_i S$ and a part related to exchange processes $d_e S$.

$$dS = d_i S + d_e S \tag{2.5}$$

The part describing exchange processes $d_e S$ depends on heat transport through the boundary, while the internal part can be attributed to irreversible processes in the system. By removing heat from the system one can reduce the entropy, implying a negative $d_e S$ or even $dS$. But through irreversible processes entropy can only be generated. This leads to the inequality

$$\frac{d_i S}{dt} = \frac{1}{T}\frac{dU}{dt} + \frac{p}{T}\frac{dV}{dt} - \sum_i \frac{\mu_i}{T}\frac{dN_i}{dt} \geqslant 0, \tag{2.6}$$

that corresponds the second law of thermodynamics, but for non-equilibrium systems.

To give an example, the entropy production rate from heat conduction of an isolated system with two heat baths with distinct temperatures is calculated [KP98, p. 94]. The two heat baths have well defined temperatures $T_1$ and $T_2$. They are isolated to the outside ($d_e S = 0$), but are in contact with each other. The heat flow is $\frac{dQ}{dt} = \alpha(T_1 - T_2)$. This leads to an entropy production rate $\sigma$ of:

$$\sigma = \frac{dS}{dt} = \frac{d_i S}{d_t} = \left(\frac{1}{T_2} - \frac{1}{T_1}\right)\frac{dQ}{dt} = \ldots = \alpha\frac{(T_1 - T_2)^2}{T_1 T_2} \geqslant 0. \tag{2.7}$$

As this describes an isolated system, it will naturally approach equilibrium at some time. The entropy production ceases when the entropy reached its maximum according to the second law.

## 2.2 REACTION NETWORKS

The growing importance of investigating complex systems has lead to different approaches for handling new phenomena and growing amount of data. One are the different aspects of network theory. It is applied in social science, for investigation of the spreading of diseases, for gene regulatory networks in systems biology, for data analysis and in many other areas. The basic theory of reaction networks describes how a set of chemical species interact through a set of reactions. Formally, reaction networks [Cla88] consist of a set of species $\mathcal{M}$ combined with a set of reactions $\mathcal{R}$. They contain information on the connection of chemical species through reactions and include the stoichiometric constraints given by the reactions. Reaction networks can be represented in different ways (Fig. 2.1). Either by a list of chemical reaction equations or by different graphical representations.

Mathematically, a reaction network can also be described by two stoichiometric matrices $\mathbf{N}^{\text{in}}$ and $\mathbf{N}^{\text{out}}$. $N_{ij}^{\text{in}}$ is the coefficient of the i-th species on the left side of the j-th reaction and $N_{ij}^{\text{out}}$ is the coefficient of the i-th species on the right side of the j-th reaction. Combining both matrices gives the stoichiometric matrix $\mathbf{N} = \mathbf{N}^{\text{out}} - \mathbf{N}^{\text{in}}$, for which the element $N_{ij}$ in i-th row and j-th column gives the effective change of species i by reaction j. Given a relation $\mathbf{v} = \mathbf{v}(\mathbf{x})$ between reaction rates $\mathbf{v}$ and species concentrations $\mathbf{x}$, one can associate the reaction network with the dynamics of an ordinary differential equation (ODE):

$$\frac{dx}{dt} = \mathbf{N} \cdot \mathbf{v}(\mathbf{x}). \tag{2.8}$$

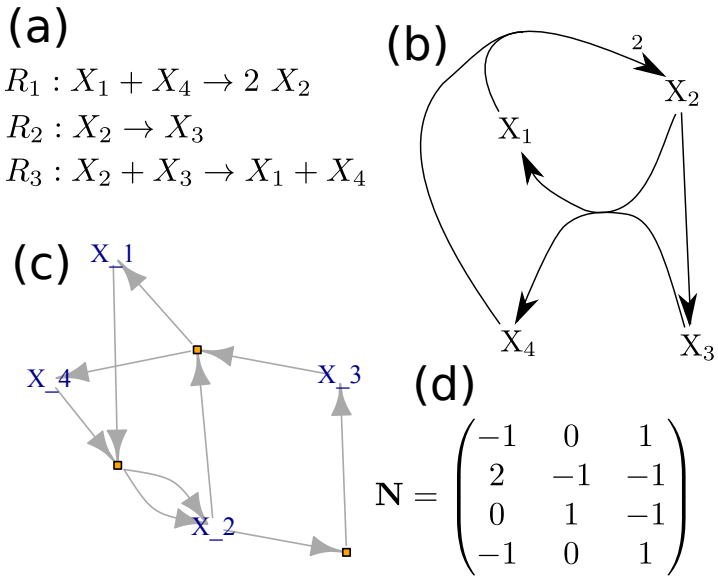



Figure 2.1: **Different representations of the same reaction network.** The same network can be represented through its reaction equations **(a)** or different graphical forms in which reactions are either represented by connected arrows **(b)** or by yellow squares **(c)**. **(d)** The stoichiometric matrix $\mathbf{N}$ does not uniquely encode the network if the network includes autocatalytic reactions (same species occurring in reactants and products of a reaction).

If the functional form of the kinetic law $\mathbf{v} = \mathbf{v}(\mathbf{x})$ is known, the differential equation can be simulated as well as analysed as a dynamical system. The fixed points can be found by solving $0 = \mathbf{N} \cdot \mathbf{v}(\mathbf{x})$ for $\mathbf{x}$. Linearization allows to determine the stability of these fixed points [Cla88].

### 2.2.1 *Mass Action Kinetics*

In principle one can imagine different kinetic laws to simulate the dynamics of a chemical system. As our aim is to simulate and analyse reaction networks in a thermodynamically consistent framework we use the kinetic law for elementary reactions: *mass action kinetics* [HJ72]. A reaction being elementary means that it happens in one step and there are no intermediates that can be detected. Because it is necessary for all reactants to collide (meet in space) for the reaction to become possible, the reaction rate $v_i$ scales with the product of all their concentrations:

$$v_i(\mathbf{x}) = k_i \prod_j x_j{}^{N^{in}_{ji}}. \tag{2.9}$$

For a general, thermodynamically realistic description, chemical reactions have to be simulated reversible. The net reaction rate is de-

fined as the difference of mass action kinetics rates (Eq. 2.9) applied for the forward and the reverse reaction:

$$\begin{aligned} \nu_i(\mathbf{x}) &= \nu_i^f - \nu_i^b \\ &= k_i^f \prod_j x_j{}^{N_{ji}^{in}} - k_i^b \prod_j x_j{}^{N_{ji}^{out}}. \end{aligned} \tag{2.10}$$

An important formula for constraining the reaction rate constant $k$ and describing its scaling with temperature is the *Arrhenius equation*,

$$k = k_0 e^{-E_a/RT}. \tag{2.11}$$

In this equation $k_0$ is a reaction dependent constant, $E_a$ the activation energy of the reaction, $R$ the ideal gas constant, and $T$ the absolute temperature. The equation can be derived using transition state theory [KP98, p. 231].

### 2.2.2 *Substrate Graphs*

To analyse the topology (connectivity structure) of reaction networks, complex network theory is used. The area of complex networks extends the older concept of a graph from mathematics and computer science. Formally a *graph* is defined as a pair $G = (V, E)$ of nodes $V$ and edges $E$. Edges are 2-element subsets of $V$ Sometimes nodes are also called vertices and edges links. If the edges have a direction they are defined as ordered pairs $E \subset V \times V$, then $G = (V, E)$ is called a *ordered graph*. An example of an ordered graph with $V = \{A, B, C, D, E\}$ and $E = \{(D, D), (A, D), (A, B), (A, B), (B, C), (C, B), (C, A)\}$ is shown in Fig. 2.2 **(a)**.

*A* complex network *can be simply understood as a graph with nontrivial statistical properties.*

*Often* $E$ *is a multiset. For clarity* $G$ *is then called multigraph.*

For comparison of large-scale properties of reaction networks they are often transformed into a graph representation. This representation is called substrate graph. The nodes in the substrate graph correspond to the chemical species of the reaction network and an edge between two species is added each time they are found on different sides of a reaction equation (Fig. 2.2 **(b)**) [SM04b; WF01]. For the sample network in Fig. 2.2 **(c)** this results in the substrate (multi)graphs shown in Fig. 2.2 **(d)&(e)**. In the substrate graph representation basic methods from computer science can be used to calculate simple topological properties like cycles [Gle+01] and shortest path.

For example, such a network theoretic analysis of chemistry found that the interstellar medium (ISM) [JD10] shows a different scaling behaviour than a biological network (metabolism of E. coli) [JD12]. Earth's atmosphere was found to have a modular reaction network comparable to other networks of living systems, while inorganic networks have a simpler structure [SM04b].

### 2.2.3 *Elementary Modes & Reaction Pathways*

In bioinformatics for looking at the steady state of $\mathbf{N}$ without knowing the function $\mathbf{v} = \mathbf{v(x)}$, often elementary (flux) mode analysis [SH94] is used. An elementary (flux) mode of a reaction network

is a collection of reactions that can operate together in steady state and can not be decomposed further.[1] Be $|(\mathcal{R})|$ the number of reactions in our reaction network, then an elementary mode $\mathbf{E} \in \mathbb{N}_0^{|(\mathcal{R})|}$ has to meet the following conditions:

[1] *See Fig. 2.3* ***(b)*** *for an example.*

**a)** Not change the concentration of any intermediary species $S_i \in \mathcal{M}$:

$$\forall S_i : \sum_j E_j N_{ij} = 0. \tag{2.12}$$

**b)** Not be a proper superset of any other (potential) elementary mode **F** (that fulfils condition a)):

$$\nexists \mathbf{F} : \left(\forall j, E_j = 0 : F_j = 0\right) \wedge \left(\exists k, F_k = 0 : E_k \neq 0\right). \tag{2.13}$$

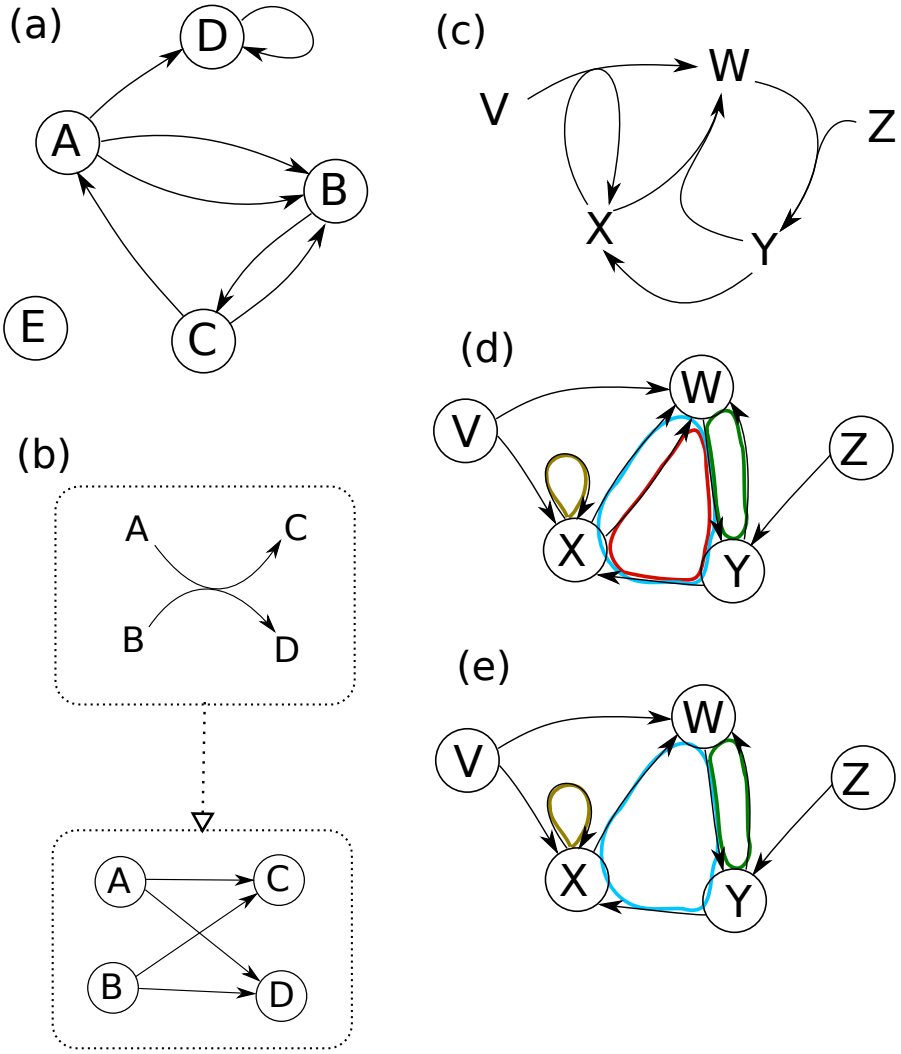



Figure 2.2: **Substrate graph and cycles. (a)** Sample of a *directed graph*. **(b)** Substrate graph of a reaction network is obtained by taking all species as nodes and adding an edge between each pair of reactants and products of every reaction. **(c)** Sample reaction network. **(d)** Substrate graph of reaction network (with multiple edges). Cycles with different size are highlighted in colour. **(e)** Substrate graph without multiple edges.

Every steady state rate vector $\mathbf{v} \in \ker(\mathbf{N})$ can be represented as a linear combination of elementary modes $\mathbf{E}^{(i)}$ with positive, real valued coefficients $\alpha_i$:

$$\mathbf{v} = \sum_i \alpha_i \mathbf{E}^{(i)}. \tag{2.14}$$

For a given steady state rate vector $\mathbf{v}$ the coefficients $\alpha_i$ are not unique. In Sect. 2.3.4 a method is introduced, that calculates coefficients that are uniquely defined for a given set of elementary modes. We will use the term *reaction pathway* to refer to elementary modes in the context of a specific steady state.[2] We will also call this expansion (Eq. 2.22) *pathway decomposition*, as it decomposes the steady state into elementary modes. An example of the pathway decomposition and calculated coefficients is shown in Fig. 2.3 **(b)**.

[2] Elementary mode *analysis is commonly used for metabolic networks in bioinformatics, where steady state rates are unknown.*

*The notion of a* reversible reaction *is not to be confused with the idea of a* reversible process *in thermodynamics!*

For application in this thesis, reaction rates $v_i$ are understood as effective rates of forward and backward reaction rates like defined in Eq. 2.10. Reaction pathways are calculated with one reaction representing the dominant reaction direction of each reaction. This is different for the analysis of metabolic networks in bioinformatics, where some reactions are treated as reversible and some as irreversible depending on their chemical equilibrium. For elementary mode analysis, reversible reactions are then represented by two identical reactions with opposing directions. This leads to pathways that only contain reactions of one direction and make it difficult to associate thermodynamic properties of reactions to pathways (cf. Eq. 2.22).

### 2.2.4 *Other Approaches*

Besides elementary flux modes [SH94; SDF99] there exists the related notion of extreme pathways [SLP00]. Those two theories are closely related and mainly differ on the way they treat reversible reactions [Pap+04].

Also related is metabolic flux analysis which tries to find the steady state using constraints and maximising the production of selected species [OTP10]. Chemical organisation theory [DF07; CD07] focuses on enumerating subsets of the reaction network that are closed and self maintaining. Similarly, the theory of autocatalytic sets [HS04; HHS10; HS12; HSK13; HSS15] specifies subsets of reaction networks that can catalyse their own production and maintain themselves.

### 2.2.5 *Thermodynamics of Reaction Networks*

Some work has been done for integration of thermodynamic constraints in reaction network formalism. With few exceptions [QB05], most of it focuses on steady state conditions. A short introduction into the formalism of thermodynamics of reaction networks is given in Qian and Beard [QB05].

Much of the research in this area focuses on connecting flux balance analysis with thermodynamics as an additional constraint [BLQ02;

Bea+04; HBHk07; PE14]. But there are also attempts to use thermodynamics to indicate regulatory control in reaction networks. [BQ05; KPH06] and to derive reaction directions [YQB05] and chemical potentials [De +12; SMH12] through thermodynamics.

Thermodynamic properties of reaction networks can be described by non-equilibrium thermodynamics as shown in Sect. 2.1.2. Similar to the change of entropy before, we are splitting the change of Gibbs free energy $dG$ into the exchange of the system with the environment $d_e G$ and change through processes in the system $d_i G$:[3]

[3] *Gibbs free energy is used because through its natural variables it is well suited for describing systems with constant pressure and temperature.*

$$dG = d_e G + d_i G. \tag{2.15}$$

Using Eq. 2.2 we get[4]:

[4] *Pressure and temperature are constant, thus* $dT = dp = 0$.

$$\frac{dG}{dt} = \frac{d_e G}{dt} + \frac{d_i G}{dt} = -\sum_i \mu_i \frac{dN_i}{dt}. \tag{2.16}$$

In steady state the concentrations do not change and $G$ is constant. But we can split the change of concentration $\frac{dN_i}{dt}$ into one part through fluxes $f_i$ (external) and one part through chemical reactions $\sum_j N_{ij}\nu_j$:

*Positive values of* $f_i$ *correspond to inflow, negative values to outflow.*

$$\begin{aligned}
\frac{dG}{dt} &= -\sum_i \mu_i \left( f_i + \sum_j N_{ij}\nu_j \right) \\
&= -\sum_i \mu_i f_i + \sum_j \left( \sum_i \mu_i N_{ij} \right) \nu_j \\
&= -\sum_i \mu_i f_i + \sum_j \Delta\mu_j \nu_j \\
&= \frac{d_e G}{dt} + \frac{d_i G}{dt}.
\end{aligned} \tag{2.17}$$

For the last simplification the definition of the Gibbs free energy of reaction $\Delta\mu_j := \sum_i \mu_i N_{ij}$ is used. Because $G$ is constant in steady state the two terms

$$\begin{aligned}
\frac{d_e G}{dt} &= -\sum_i \mu_i f_i \\
\frac{d_i G}{dt} &= \sum_j \Delta\mu_j \nu_j
\end{aligned} \tag{2.18}$$

have to cancel out. With constant pressure and constant temperature, the free energy dissipated inside the system describe the amount of heat that is generated. Thus, the entropy production $\sigma$ can be calculated by

*In steady state,* $\sigma$ *can also be calculated through boundary flow by using* $\frac{d_e G}{dt}$.

$$\sigma = \frac{d_i S}{dt} = -\frac{1}{T}\frac{d_i G}{dt} = -\sum_j \Delta\mu_j \nu_j. \tag{2.19}$$

To reformulate this, the entropy production can be defined per reaction by

$$\sigma_i = -\frac{1}{T}\Delta\mu_i \nu_i = -\frac{1}{T}\sum_j \mu_j N_{ji} \tag{2.20}$$

to get the total entropy production of the network as the sum of the entropy production of all of its reactions:

$$\sigma = \sum_i \sigma_i. \tag{2.21}$$

There is an alternative way of calculating entropy production of individual reactions in steady state by writing it in terms of forward and backward reaction rates $v_i^f$ and $v_i^b$ [KP98; BQ07]:

$$\sigma_i = (v_i^f - v_i^b) \log \frac{v_i^f}{v_i^b}. \tag{2.22}$$

For simulations done in this thesis both formulations are equivalent, though Eq. 2.22 is more general and allows to determine entropy production for complex kinetic laws like, for example, Michaelis-Menten kinetics [BQ07].

## 2.3 CALCULATING SIGNIFICANT REACTION PATHWAYS

For calculating a decomposition of the steady state of a reaction network into reaction pathways we use an algorithm inspired by Lehmann's work [Leh04]. Lehmann's algorithm was already applied to models for atmospheric chemistry [Gre+06; Sto+12a; Sto+12b] but with a slightly different aim. In these publications reaction pathways were used as a tool to explain the production or consumption of specific species. Our work has a different approach which leads to some changes to Lehmann's original algorithm:

- We treat all species as intermediate species.
- We do not use the concentration of the species as input for the algorithm. Thus we cannot calculate the residence time of the species to order them. The species are all considered as intermediate species in order of their increasing production rate.
- The system is assumed to be in steady state. Effective production and consumption of species is balanced by introducing in- and outflow pseudoreactions.
- A different criterion to decide which pathways are discarded is used. To find a decomposition of the steady state that explains a big fraction of all reactions a criterion that is independent of the rate of the reactions is used. This is especially relevant as typical steady states have reaction rates ranging over more than 10 orders of magnitude.

In the following a self contained description of the algorithm used in this thesis is given. The implementation of the algorithm in R [IG96] is freely available [Fis16a].[5]

[5] *Cf. Appendix C.*

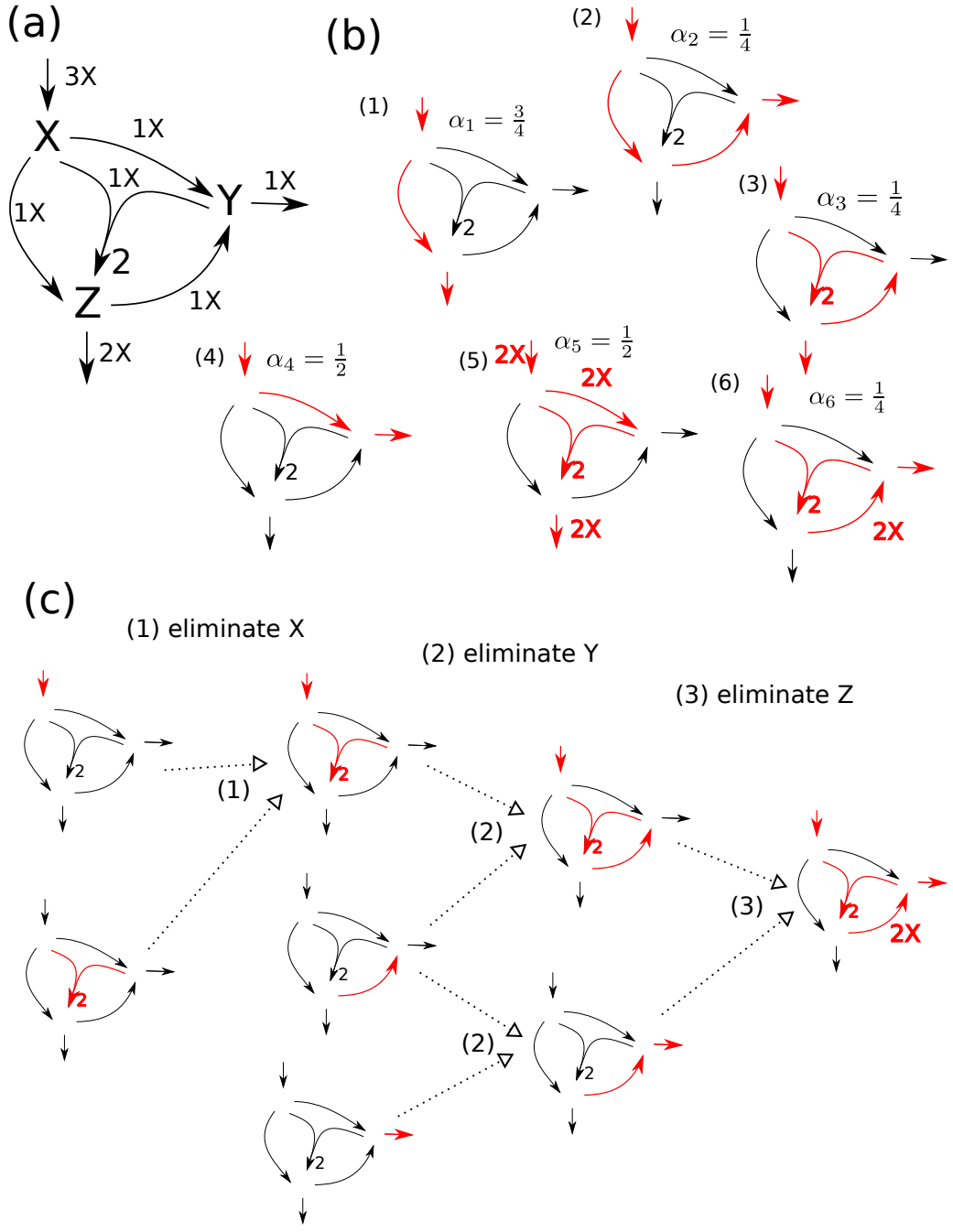



Figure 2.3: **Reaction pathways: Decomposition example and algorithm.** **(a)** Sample reaction network with steady state rates (indicated by nX). **(b)** Complete set of pathways for the reaction network. Coefficients $\alpha_i$ are calculated with the algorithm in Sect. 2.3.4 and reconstruct the given steady state. The nonuniqueness of the pathway decomposition is exemplarily shown by the sets of coefficients $\boldsymbol{\alpha} = (1,0,1,1,0,0)$ and $\boldsymbol{\alpha} = (0,1,0,0,1,0)$, which represent the same steady state. **(c)** Calculation of Pathway (6) with elementary mode algorithm (Sect. 2.3.1). Calculation starts with all reactions being represented by one individual pathways. Then, pathways producing and consuming specific species are combined, and these species "eliminated". In our example this happens in the order **(1)** X, **(2)** Y, **(3)** Z.

### 2.3.1 *General Structure*

*An example for pathway calculation (general method) is given in Fig. 2.3 **(c)***

The input of the algorithm is the structure of the reaction network given by the stoichiometric matrix $\mathbf{N}$ and a steady state rate vector $\mathbf{v}$. Before the pathways are calculated pseudo in- or outflow reactions are added for all species, so that their concentrations are balanced ($\mathbf{N} \cdot \mathbf{v} = 0$). If the system is not (yet) in steady state this can be done in the same way, but the added pseudoreactions then do not only describe the matter exchange of the reaction network in steady state with the environment, but also the change of concentrations at the corresponding point in time. For the following description of the algorithm we will also assume (without loss of generality), that all elements in the rate vector $\mathbf{v}$ are non-negative.

In the beginning, the list of pathways is initialised with the (intermediate) pathways each representing exactly one reaction.

*$\mathbf{M}^{(i)}$ is used here instead of $\mathbf{E}^{(i)}$ to distinguish intermediate from final pathways.*

$$M_j^{(i)} = \delta_{ij} \tag{2.23}$$

$$\alpha_i = v_i \tag{2.24}$$

These values fulfil $\mathbf{v} = \sum_i \alpha_i \mathbf{M}^{(i)}$. Throughout iterating all species as branching species this identity is approximately maintained and the pathways are updated and combined in a way that results in all pathways leaving the concentration of all already branched species unchanged. For ordering the species for branching, their turnover rate is used. The turnover rate $\mathbf{w}$ is defined by

$$w_i(\mathbf{v}) = \max \left\{ \sum_{j, N_{ij} \geqslant 0} N_{ij} \cdot v_j \ , \ - \sum_{j, N_{ij} \leqslant 0} N_{ij} \cdot v_j \right\} \tag{2.25}$$

and is in our case equivalent to the quantities commonly called production and consumption rate. As this shows the best runtime behaviour, we schedule the species for branching with increasing turnover rate. This means that species with high turnover rate, and probably also a high number of associated reactions, are used for branching last.

When branching at a species $X_i$, the pathways are separated in three disjunct sets, depending on how they affect the concentration of $X_i$:

$$\Delta X_i(\mathbf{M}^{(k)}) = \sum_j N_{ij} M_j^{(k)}. \tag{2.26}$$

Apart from possible dropping of unimportant pathways as described in Sect. 2.3.2, pathways that do not change the concentration of $X_i$ are left unaffected by branching and all pathways that net produce $X_i$ are combined with all pathways that net consume it. The coefficients are chosen so that the combined pathway does not change the concentration of $X_i$. The combination of the net producing pathway $\mathbf{M}^{(k)}$ with the net consuming pathway $\mathbf{M}^{(j)}$ is given by

$$\mathbf{M}^{(k,j)} = \|\Delta X_i(\mathbf{M}^{(j)})\| \cdot \mathbf{M}^{(k)} + \|\Delta X_i(\mathbf{M}^{(k)})\| \cdot \mathbf{M}^{(j)}. \tag{2.27}$$

Finally, for every newly combined pathway the integer coefficients defining the pathway are divided by their greatest common divisor and the pathway is then replaced by its elementary pathways as described in Sect. 2.3.3. When the new set of pathways has been calculated, new coefficients $\alpha_i$ are determined (Sect. 2.3.4) and the criterion for dropping pathways with minor importance is applied (Sect. 2.3.2).

After iterating all species as branching species $\mathbf{M}^{(i)}$ contains the final set of pathways and $(\alpha_i)$ the corresponding coefficients. If the quality of the expansion is insufficient, important pathways may have been dropped (2.3.2) and the coefficients $r$ and $f_{exp}$ should be adapted accordingly.

### 2.3.2 *Discarded Pathways*

There are two points in the algorithm at which pathways can be discarded. One is at branching, before combining a pathway that produces and one that consumes the branching species. For this, we split the pathways in two classes. "Important" pathways generating the branching species are combined with all pathways consuming it. "Unimportant" producing pathways however are only combined with consuming pathways that are themselves in the class of important pathways. Ultimately, every pathway producing and consuming the branching species is combined with some other pathway, but pathways contributing little to the production are not combined with pathways that have a small share in consumption.

To split the pathways in two classes, producing pathways as well as consuming pathways are ordered with decreasing fraction they have at the production or consumption of the branching species. In both cases, pathways are assigned to the class of important pathways until their cumulative fraction of production or consumption exceeds the free parameter $r \in [0, 1[$.

The second point for discarding pathways is after the two pathways are combined and the new (intermediate) decomposition of the steady state is calculated. For each pathway the maximum fraction that it explains of a reaction rate, $f_R$, and the maximum fraction that if explains of the turnover of a species, $f_T$, are calculated.

$$f_R(\mathbf{M}^{(i)}, \alpha_i) = \max_{j, M_j^{(i)} \neq 0} \frac{M_j^{(i)} \cdot \alpha_i}{\nu_j} \tag{2.28}$$

$$f_T(\mathbf{M}^{(i)}, \alpha_i) = \max_{j, M_j^{(i)} \neq 0} \frac{w_j(M_j^{(i)} \cdot \alpha_i)}{w_j(\mathbf{v})} \tag{2.29}$$

In the last equation $w_j = w_j(\mathbf{v})$ is used as defined in Eq. (2.25). After each branching step the pathways for which both measures, $f_R$ as well as $f_T$, are below the parameter $f_{exp}$ are discarded and removed from the list of pathways that will be used for branching in the following steps.

### 2.3.3 *Non-Elementary Pathways*

At each branching step the newly generated pathways have to be checked for the elementary property. In the original elementary modes algorithm this is simply done by checking for the non-existence of a proper sub-pathway (other pathway which set of reactions is a proper subset of its reactions). In our case this is not possible because some of the sub-pathways may not be included in the list of pathways as one of their precursors might have been discarded. Thus, for every combined pathway the original elementary mode algorithm is applied to the subnetwork defined by this pathway. If the pathway is not elementary (because it has proper sub-pathways) it is removed from the list of newly generated pathways and instead all its elementary sub-pathways are added.

The structure of the original elementary mode algorithm [SS93] used for this step is identical to the general algorithm as described above (Sect. 2.3.1), except that no coefficients are calculated and no pathways are discarded. With no rates known in this case, the species for branching are ordered randomly.

### 2.3.4 *Calculation of Coefficients*

For the expansion of a steady state rate vector into elementary pathways the coefficients $\alpha_i$ are not uniquely defined. To counter this arbitrariness we are using the following method to calculate a unique set of coefficients for a given steady state $\mathbf{v}$ and a given set of reaction pathways $\mathbf{M}^{(i)}$. This method iteratively distributes the steady state rate to as many reaction pathways as possible.

It starts from a set of pathways $\mathbf{M}^{(i)}$ and a steady state rate vector $\mathbf{v}$. On the course of the iterative algorithm, a list of current coefficients that is initialised with zeros ($\alpha_i = 0$) and a vector containing the yet unexpanded part of the steady state (initially $\mathbf{v}^* = \mathbf{v}$) are kept and updated. After initialising, the following steps are iterated:

- Determine the set of pathways I that are compatible with the currently unexpanded part of the steady state $\mathbf{v}^*$:

$$I = \{i | \{k | M_k^{(i)} \neq 0\} \subset \{j | v_j^* \neq 0\}\}$$

- If I is empty, the algorithm is finished. The final coefficients are in $\alpha_i$ and the remaining unexpanded rates in $\mathbf{v}^*$.

- If I is not empty, the coefficients of all pathways in I are increased by the same amount $\Delta\alpha$ until at least one additional reaction in $\mathbf{v}^*$ is depleted.

$$\Delta\alpha = \min_{j, v_j^* \neq 0} \frac{v_j^*}{\sum_{i \in I} M_j^{(i)}}$$

- Update the coefficients accordingly and afterwards continue with the next iteration.

$$\alpha_i' = \begin{cases} \alpha_i + \Delta\alpha & \text{if } i \in I \\ \alpha_i & \text{else} \end{cases} \tag{2.30}$$

The quality of the expansion can be evaluated by comparing the remainder of the rates $\mathbf{v}^*$ with the steady state rates $\mathbf{v}$.

# 3

# THERMODYNAMICS OF LARGE-SCALE RANDOM REACTION NETWORKS

Large-scale artificial reaction networks are created and their non-equilibrium steady state is investigated for various boundary fluxes. Linear and nonlinear reaction networks of different topologies are compared. The distribution of entropy production of the individual reactions inside the network follows a power law. An elevated entropy production rate is found in reactions associated with weakly connected species. Increasing the flow through nonlinear reaction networks leads to an increased number of cycles through self-organization like switching of reaction directions.
*This chapter is based on work previously published in [FKD15].*
**keywords:** reaction networks, complex networks, thermodynamics, artificial chemistry, cycles

*This chapter requires the reader to understand basics of reaction networks and thermodynamics (see Chapter 2).*

## INTRODUCTION

Connecting network theory with thermodynamics was an idea already present more than 40 years ago under the term network thermodynamics [OPK71; Per75]. Despite the fact that the terms were used in combination, the theory was merely a graphical representation of conservation equations and did not make any statements about complex networks, as they are known today. In 2006 Cantú and Nicolis [CNo6a] studied thermodynamic properties of linear networks, but limited themselves to small networks, which they were able to handle analytically. Here, we extend their study by generating large random linear and nonlinear reaction networks and simulating them to a thermodynamically constrained steady state. This might contribute to a framework that allows to test methods for reconstructing thermodynamic data of reaction networks [SH10; SMH12] and lead to a better thermodynamic understanding of reaction networks in general. Possible applications of this approach include the thermodynamic investigation of reaction models in biology [SH10; SMH12; MQN14], origin of life [HS12] and also Earth system and planetary science [KHP85; Sau+03].

*The field of* complex networks *itself was not existent 20 years ago.*

We look at reaction networks as thermodynamic systems that transform two chemical species into one another [CNo6a] (see Fig. 3.1). The environment is driving the network to thermodynamic disequilibrium by keeping the concentration of two species constant. In the following, we will call the chemical species that are kept constant *boundary species*, because they are the species to which the boundary conditions are applied to.

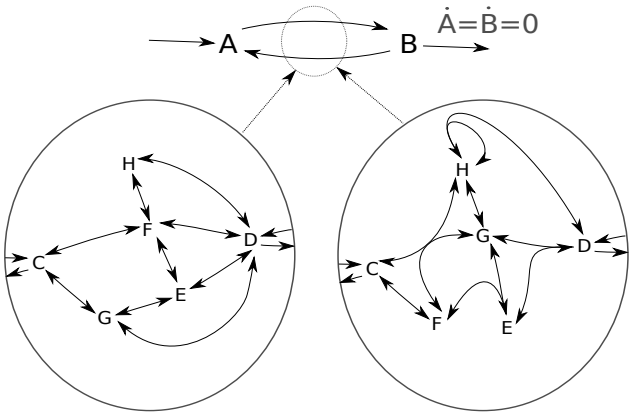



Figure 3.1: **Simple method for maintaining reaction networks in non-equilibrium steady-state.** This idea is inspired by the work of Cantú and Nicolis [CNo6a] and allows to control thermodynamic forcing through concentration of species $A$ and $B$. The implicit assumption is, that there are pathways that transform $A$ into $B$. For the external observer the network in steady state can be replaced by a single reaction. This scheme is realized with linear reaction networks **(left)** as well as nonlinear networks **(right)**.

Our basic assumption is that the network in steady state is able to transform the two boundary species into each other. This is not always possible in real reaction networks where the transformations are constrained by stoichiometry of chemical constituents. For example, any chemically sound reaction model will implicitly forbid pathways that transform $N_2O$ into $H_2O$. Even if the artificial networks we create are comparable in density to real network, they are not created with this constraint. This is due to the implications this constraint would have on the complexity of the boundary conditions. Omitting it leads to the existence of many transformation pathways between most pairs of randomly chosen boundary species, otherwise almost all pairs of boundary species would just have a steady state flow of zero between them.[1]

*The model in Chapter 5 will allow to consider stoichiometric constraints.*

We study different quantitative properties of the networks at steady state. In particular, because cycles have been reported to have important functions in networks [Mor66; Gle+01; Kre+12; PPS13], we look at the cycles that appear in the flow pattern. These cycles depend on the direction of the flow of each reaction, which in turn depends on the strength of the thermodynamic disequilibrium caused by the boundary condition.

*Cycles are investigated using the networks substrate graph as shown in Fig. 2.2.*

The next section describes a method for generating reaction networks so they resemble different complex network models and a method to simulate their reaction equations and find their non-equilibrium steady state. Results concerning the flow through the networks, the distribution of entropy production of individual reactions, and the dependency of cycle number from flow through the nonlinear networks are then presented.

[1] *The existence of pathways between boundary species is also an effect of many linear reactions in the network.*

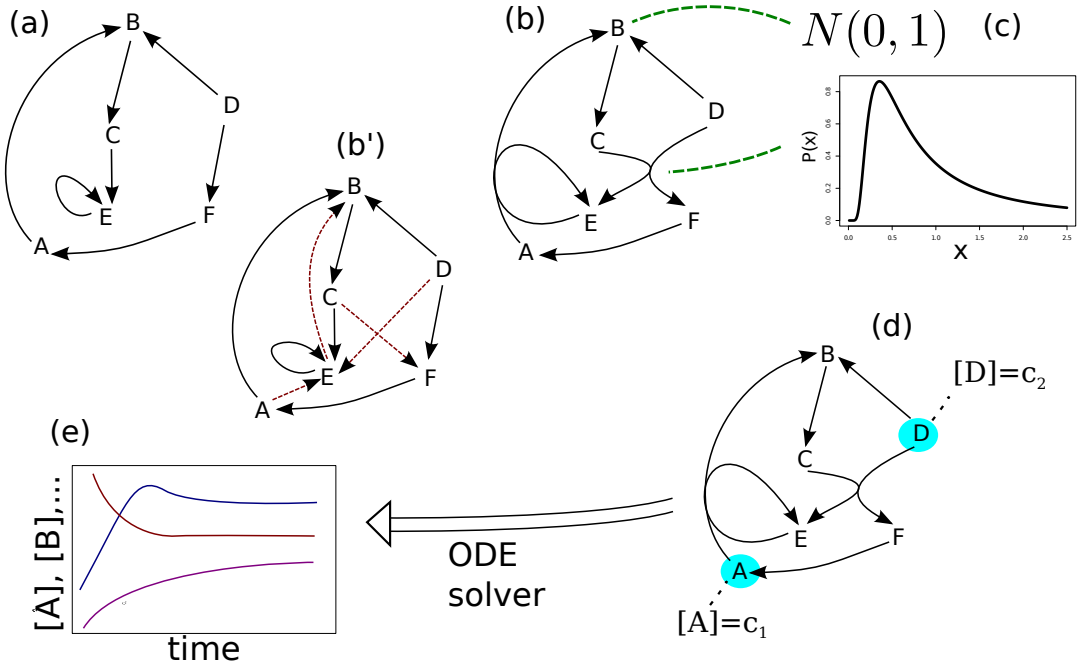



Figure 3.2: **Illustration of realized simulations: (a)** Linear reaction networks are generated from existing complex network models. (Arrows represent reactions, chemical species are indicated by uppercase letters.) **(b)** Pairs of linear reactions are coupled to form nonlinear reactions. **(b')** Substrate graph that should maintain its characteristic properties while coupling. Edges invoked by coupling are depicted with dotted arrows. **(c)** Gibbs energies of formation are assigned to species from a normal distribution, activation energies to reactions from a Planck-like distribution (Eq. 3.2). **(d)** Two boundary species whose concentrations are kept constant are selected while the other species are initialized randomly. **(e)** Reaction equation is solved numerically and final rates are taken as steady state rates.

## 3.1 METHODS

### 3.1.1 *Network Construction*

Our artificial reaction networks are generated in three steps. We first generate a simple directed network (graph) consisting out of a set $V$ of $N$ nodes and a multiset $E \subset V \times V$ of $M$ edges. These networks are generated following the models of Erdős-Rényi [ER59], Barabási-Albert [AB02] (scale-free), Watts-Strogatz [WS98] (small-world, clustering) and Pan-Sinha [PS09; PS07] (hierarchically-modular). In each case we are using variants of these network models that allow formation of self loops and multiple edges between the same nodes. Also, we generate networks with a fixed number of edges. From these complex networks the reaction network is constructed.

*Using a fixed number of edges maintains comparability in density.*

Simple reaction networks are created by translating each edge into a reaction of the form $X \rightleftharpoons Y$ with $X$ being the first and $Y$ being the second node of the edge. In the rate equation of mass action kinetics this leads to a linear dependence of the reaction rates from the

concentrations and thus we are calling these networks *linear* reaction networks.[2]

[2] *This implies a linear dependency of all concentrations from boundary concentrations.*

Nonlinear reaction networks are generated out of directed complex networks by combining pairs of edges to second order reactions of the form $X + Z \rightleftharpoons Y + W$. The selection of pairs is done with a probability distribution that maintains the characteristic properties of the substrate graph as far as possible. This is done by considering the probability of newly introduced edges in the substrate graph in the originally used network model. For example, consider the combination of the reactions $A \rightleftharpoons B$ and $C \rightleftharpoons D$ to create the reaction $A + C \rightleftharpoons B + D$. This leads to two new edges in the substrate graph between A and D as well as between C and B. The probabilities of these two edges in the original network model are then used to calculate the probability of the combined reactions.

Finally, the thermodynamic data is generated and assigned to species and reactions. In the following, the generation process of nonlinear networks specific to the different network models is explained before the generation of thermodynamic data is specified in detail.

*Erdős-Rényi (ER)*

In the Erdős-Rényi network model [ER59] all possible edges have the same probability. We create these networks by simply drawing the nodes of every edge from the set of all nodes with uniform probability. For the construction of nonlinear reaction networks, second order reaction equations are then chosen from the set of pairs of linear equations with uniform distribution. Note that linear equations that are used as part of a nonlinear equation are not replaced, meaning that the probability of all pairs of linear equations containing it is set to zero for subsequent couplings.

*Barabási-Albert (BA)*

For generating scale-free networks, the Barabási-Albert model is used [AB02]. In this model nodes are added consecutively. Newly added nodes are connected to the network by introducing edges between it and already existing nodes. The selection of nodes to attach to is done with probability scaling with their node degree (preferential attachment).

*This preference of well connected nodes leads to power law degree distribution (scale-free property).*

The coupling probability of linear reactions is calculated from the product of the node degrees of the chemical products. In principle, other functional dependencies are possible, but for simplicity we choose this one and check that it maintains the power law distribution of the node degree in the associated substrate graph (Fig. 3.3 (a)).

*Watts-Strogatz (WS)*

Networks having a comparable average path length to the Erdős-Rényi model, but with a higher clustering coefficient, are generated with the Watts-Strogatz model [WS98]. From a circular lattice like

structure, a fraction $\alpha$ of all edges is randomly reordered. Considering the size of our networks we choose a value of $\alpha = 0.1$ (see Table 3.1).

For the creation of nonlinear networks, we only form couplings between linear reactions that lead to two new close edges in the substrate graph. Here "close" means that their distance in terms of the circular lattice is not larger than the largest distance of non reordered edges in it. It would have be possible to use a more sophisticated approach and use the parameter $\alpha$ as the probability of introducing a far edge in the substrate graph while coupling. But because even our simple method does not achieve a clustering coefficient as high as equivalent linear networks (Table 3.1) we use this simple method.

*Allowing more far edges while coupling would lead to a even lower clustering coefficient.*

*Pan-Sinha (PS)*

Hierarchically-modular networks are generated starting with uniformly partitioning the nodes into $2^h$ elementary modules, with $h$ being the number of hierarchical levels of the network. On the first level, two pairs of modules on the elementary level are joined to form a new module, leading to $2^{(h-1)}$ modules on the first level. Analogous, for all other levels modules of the level below are joined pairwise, up to the h-th level where there is just one module consisting out of the entire network. When edges are added to the network, this happens with a probability proportional to the lowest level $l$ in which the two nodes to be connected share a module. Two nodes that share an elementary module are connected with the probability $p_0$ whose value is given by normalization. Nodes whose lowest common level is $l$ are connected with probability $p_0 p^l$. For our networks we choose $p = 0.5$ and $h = 8$.

When creating nonlinear reactions we assign each possible coupling a probability proportional to the product of the probability of the two newly introduced edges in the original model. Assuming a coupling leads to new edges in the substrate graph between nodes with lowest common module on levels $l_1$ and $l_2$, then the probability of choosing this coupling is scaled with $p^{l_1} p^{l_2}$.

*Parameters*

For network construction we generate linear reaction networks with $N = 1000$ species and $M = 5000$ first-order reactions. Nonlinear networks are built by generating a linear network with $M = 3000$ reactions and connecting $C = 1000$ of them to second-order reactions. To compare linear and nonlinear networks directly we also generate linear networks from the substrate graph of the nonlinear networks. This comparison is not possible with the generated linear reaction networks because their substrate graph is not as clustered. An overview of all generated networks is shown in Table 3.1.

*As nonlinear reactions have four edges in the substrate graph, this leads to the same density than the linear reference network.*

### 3.1.2 *Thermodynamics of Reaction Networks*

For generating thermodynamic data and analysis we are using the formalism introduced in Sect. 2.2.5. As we are generating artificial data and do not directly compare with chemical data, unitless equations are used in this chapter ($k_B T = 1$ or $RT = 1$).

*Generating Thermodynamic Data*

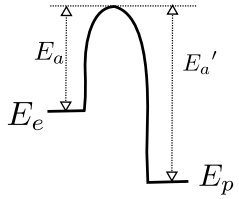
For calculating the networks dynamics, Gibbs energies of formation of the species $\mu_i^0$ are drawn from a normal distribution $N(0, 1)$. Reaction rate constants are then calculated using the Arrhenius equation (with the prefactor $A$ set to 1):

$$k = Ae^{-E_a} = e^{-E_a}. \tag{3.1}$$

Here, $E_a$ is the activation energy which is sampled from the distribution

$$P(x) = \frac{6}{\pi^2} \frac{1}{x^3(\exp(1/x) - 1)} \tag{3.2}$$

for every reaction. We have chosen this distribution, which resembles the Planck-distribution, because it has an effective non-zero lower bound while still having a large tail to the right (Fig. 3.2). We simulate all reactions reversibly. Forward and backward reaction are energetically constrained by the Gibbs energies of the species. Thus, we sample $E_a$ just once for every reaction and assign it to the reaction direction which respective educts have a higher Gibbs energy of formation, either $E_{e,i} = \sum_{j,N_{ij}<0} \mu_i^0 |N_{ij}|$ (forward direction) or $E_{p,i} = \sum_{j,N_{ij}>0} \mu_i^0 N_{ij}$ (backward direction). The activation energy of the second reaction direction is then given by the constraint $E_a' = E_a + |\sum_j \mu_i^0 N_{ij}|$. This expression is a reflection of the fact that in equilibrium forward and backward reaction rates need to balance.




*The reaction rate constant for the backward reaction is thus written* $k_b = e^{-E_a'}$.

*See Eq. 2.22.*

*Analysing Dissipation*

In steady state the entropy production of individual reaction $i$ can be as function of forward and backward reaction rates $v_i^f$ and $v_i^b$ like explained in Sect. 2.2.5:

$$\sigma_i = (v_i^f - v_i^b) \ln \frac{v_i^f}{v_i^b}. \tag{3.3}$$

This relation can be applied to calculate the total entropy production rate $\sigma_{tot} = \sum_i \sigma_i$ of a reaction network acting between two boundary species $b_1$ and $b_2$ kept at concentrations $c_1$ and $c_2$. As the entropy production rate in steady state only depends on the boundary conditions ($c_1, c_2, \nu = \nu_+ - \nu_-$) we can replace the entire network with one imaginary linear reaction $b_1 \rightleftharpoons b_2$. If we assume the Gibbs energies of formation of boundary species to be zero, the forward and backward rate coefficients are equal and we obtain the equation

*Cf. Fig. 3.1.*

$$\sigma_{tot} = \nu \ln \frac{c_1}{c_2}. \tag{3.4}$$

Alternatively one could also get this result by calculating the boundary species entropy exchange with the environment, because in steady state $0 = dS = d_eS + d_iS$.

### 3.1.3 *Network Simulation*

As we are interested in the steady state of the network under thermodynamic boundary conditions, we solve the reaction equation while keeping the concentration of two selected chemical species $b_1$, $b_2$ at fixed concentration $c_1$, $c_2$. To remove the effects of the energy difference between the boundary species on the flow, we set their Gibbs energy of formation $\mu_i^0$ to zero and recalculate reaction rates before starting the simulation. For solving the ODE, the C++-library *boost* is used. The selected algorithm is "Dormand-Prince 5". Concentrations are initialized normally distributed with $\frac{c_1+c_2}{2}$ taken as mean and $|c_1 - c_2|$ as standard deviation. Dynamics are simulated up to a time $t$ of 50000 or up to the time when the mean square change of concentration (per species and time-step size) is smaller than $10^{-20}$.

*This is an explicit method. The implicit (stiff) solver used for Chapter 5 cannot handle networks of this size.*

We assume that the greatest topological factor influencing flow through the reaction network is the shortest path distance between the boundary species. Because we cannot perform simulation and analysis for one million pairs of boundary species, we sample 50 pairs of boundary species for all values of the shortest path occurring in the network.

A simple investigation of network flow and dissipation is done with boundary species concentrations set to $c_1 = 0.1$ and $c_2 = 1$. To get an error estimate, we generate 10 independent samples of every network type.

To investigate the response of the nonlinear networks to an increase in thermodynamic disequilibrium we vary the boundary conditions accordingly. For this we keep $c_1$ at 0.1 while varying $c_2$ from 0.2 up to 60. With higher values of boundary concentrations we notice an extreme increase of computational time needed to solve the individual ODEs. Thus, we are only able to simulate one network sample of every type for this setup.

The software used is contained in the digital appendix (Appendix C) and available on GitHub[3]. The networks were generated with the *jrnf_tools* C++-program and simulated with *jrnf_int*. For analysis a set of R-tools was used (*jrnf_R_tools*).

[3] https://github.com/jakob-fischer/

Table 3.1: **Network properties:** Properties of the substrate graphs of artificially generated networks as well as of examples of real networks. Table contains the number of vertices (|V|) and edges (|E|), and the mean shortest path length (< L >). The clustering coefficient (< C >) is taken from the respective undirected network. The modularity is calculated using the walktrap community finding algorithm [PL06]. Data for real networks is taken from a database for Earth's photochemical reactions [YD98], models for the combustion of Ethanol [Mar99] and Dimethyl ether [Kai+00] and a kinetic model of Yeast's metabolism [Sta+13]. For the artificial networks and the randomizations of the real networks mean values and standard deviations are calculated from 10 samples. For real networks also reference networks obtained through randomizing reaction directions (but not association to species) are calculated.

| network | \|V\| | \|E\| | < L > | < C > | modularity | 1-cycles | 2-cycles | 4-cycles |
|---|---|---|---|---|---|---|---|---|
| ER (linear) | 1000 | 5000 | 4.5 ± 0.02 | 0.0098 ± 5e-04 | 0.234 ± 0.0037 | 4.6 ± 2.4 | 10 ± 2.3 | 151 ± 14 |
| ER (nonlinear) | 1000 | 5000 | 4.5 ± 0.03 | 0.0141 ± 0.002 | 0.296 ± 0.00921 | 4.1 ± 2 | 1634 ± 274 | 2964 ± 939 |
| BA (linear) | 1000 | 5000 | 3.9 ± 0.04 | 0.0277 ± 0.001 | 0.178 ± 0.00412 | 5.2 ± 2.4 | 47 ± 9.3 | 2138 ± 245 |
| BA (nonlinear) | 1000 | 5000 | 3.8 ± 0.04 | 0.0386 ± 0.002 | 0.266 ± 0.00664 | 9 ± 3.9 | 1254 ± 273 | 9321 ± 5306 |
| WS (linear) | 1000 | 5000 | 6.4 ± 0.1 | 0.484 ± 0.009 | 0.805 ± 0.00569 | 0.3 ± 0.48 | 2.4 ± 1.8 | 4188 ± 158 |
| WS (nonlinear) | 1000 | 5000 | 6.7 ± 0.1 | 0.255 ± 0.008 | 0.748 ± 0.00733 | 90 ± 5.1 | 1002 ± 58 | 2201 ± 161 |
| PS (linear) | 1000 | 5000 | 5 ± 0.03 | 0.0414 ± 0.003 | 0.51 ± 0.0124 | 469 ± 17 | 208 ± 24 | 316 ± 31 |
| PS (nonlinear) | 1000 | 5000 | 4.6 ± 0.06 | 0.0297 ± 0.001 | 0.314 ± 0.00817 | 285 ± 23 | 1782 ± 553 | 3758 ± 1745 |
| Earth's atm. | 280 | 1846 | 2.9 | 0.147 | 0.301 | 3 | 1337 | 48503 |
| (random dir.) | 280 | 1846 | 2.7 | 0.147 | 0.301 | 3 | 3675 (39) | 661986 (10544) |
| (randomized) | 280 | 1846 | 3.3 ± 0.03 | 0.0513 ± 0.003 | 0.202 ± 0.011 | 6.2 ± 1.6 | 21 ± 5 | 480 ± 42 |
| Ethanol | 57 | 2902 | 1.7 | 0.4977 | 0.171 | 18 | 14264 | 1.59e+07 |
| (random dir.) | 57 | 2902 | 1.7 | 0.498 | 0.171 | 18 | 13506 (93) | 1.56e7 (2e5) |
| (randomized) | 57 | 2902 | 1.4 ± 0.005 | 0.831 ± 0.006 | 0.051 ± 0.014 | 56 ± 7.4 | 1265 ± 22 | 1.49e+6 ± 19103 |
| Dimethyl ether | 79 | 2492 | 1.9 | 0.416 | 0.285 | 10 | 9995 | 3.45e+06 |
| (random dir.) | 79 | 2492 | 2 | 0.416 | 0.285 | 10 | 9329 (94) | 3374297 (70542) |
| (randomized) | 79 | 2492 | 1.7 ± 0.005 | 0.551 ± 0.007 | 0.073 ± 0.016 | 30 ± 4.9 | 492 ± 23 | 226961 ± 5134 |
| Yeast Metab. | 295 | 16954 | 2.6 | 0.1005 | 0.0175 | 27 | 355713 | 3.34e+09 |
| (random dir.) | 295 | 13742 | 2.8 | 0.101 | 0.0175 | 27 | 347764 (32317) | 2.98e+09 (4.27e+08) |
| (randomized) | 295 | 16954 | 2 ± 0.01 | 0.51 ± 0.02 | 0.025 ± 0.045 | 59 ± 6.1 | 1715 ± 208 | 2.77e+06 ± 6.5e+05 |

## 3.2 RESULTS

For this study, multiple random networks are generated for each network model (Table 3.1). The networks are simulated with various boundary conditions and the resulting steady state is analysed. In the following, we first compare the artificial networks with real networks and then go into detail of how the flow and and energy dissipation depend on network structure and boundary condition.

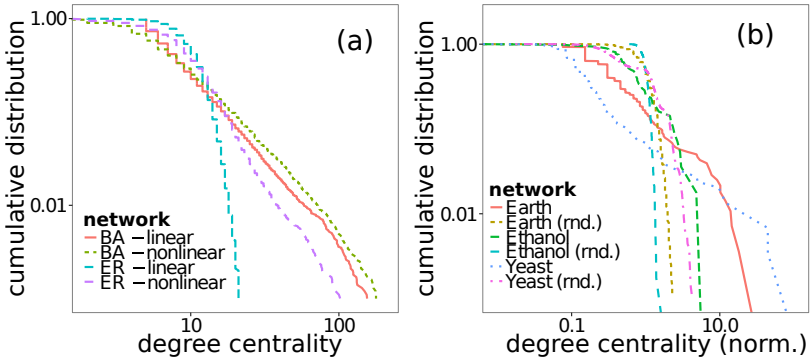



Figure 3.3: **Comparison of artificial and real networks: (a)** Nonlinear BA networks maintain scale-free degree distribution. **(b)** Cumulative degree scaling of real network's substrate graphs shows pronounced scale-free property in comparison with their null models (randomized counterparts) for Earth's photochemistry [YD98] and a kinetik model of Yeast's metabolism [Sta+13].

### 3.2.1 *Network Structure*

We compare the topological features of the artificially generated networks with real world networks (Table 3.1). For this a compilation of chemical reactions in Earth's atmosphere [YD98] and models for the combustion of Methane [Mar99] and Dimethyl ether [Kai+00] is used. Also a kinetic model of the metabolic network of Yeast [Sta+13], available through the BioModels Database [Li+10], is investigated. To avoid that the representation of networks as substrate graphs biases our results [ZN11], we compare each network with a randomized version of itself. When randomizing an artificial network one would obtain an Erdős-Rényi network with the same density and the same types of reactions. Thus, rows for randomized BA, WS, and PS networks are omitted in Table 3.1.

*For nonl. BA, WS and PS networks the reference is naturally the nonl. ER network as it has the same constraints on clustering.*

The power law scaling for Earth's atmospheric reaction network and the metabolic network of Yeast are clearly pronounced in comparison with their respective null models (Fig. 3.3 (b)). This is not true for the Ethanol combustion chemistry whose size of 57 species (nodes) does not allow to unambiguously decide on the scale-free property. The substrate graphs of the two combustion chemistries show the properties of small-world networks, they have a small mean shortest path length and a high clustering coefficient. As their null models show the same properties, this can be attributed to their high density.

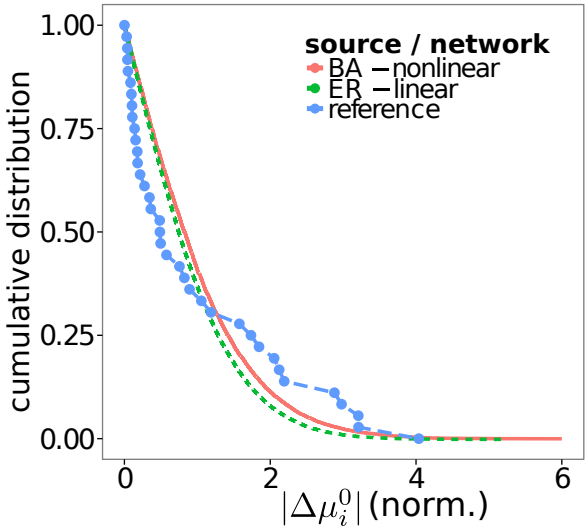



Figure 3.4: **Comparison of thermodynamic parameters of artificial and real networks:** Cumulative distribution of the standard change of Gibbs energy of formation ($\Delta\mu_j^0 = \sum_i N_{ij}\mu_i^0$) for artificial networks and respective thermodynamic reference data for glycolysis (see [Li+11], Table 4, $\Delta_r G'^0$). Distributions were (linearly) rescaled to have a mean of one.

All reaction networks have more cycles than their randomized counterparts. With the exception of the network from Yeast's metabolism all real networks also have a higher clustering coefficient or a higher value for modularity.

To see if the occurrence of cycles depends on the dynamical state of the system represented in the reaction directions, we additionally calculate the properties of the different reference networks with randomized directions. As only cycles depend on the directions of edges in the substrate graph, all other properties are equal to the originals. For the number of cycles with randomized directions we find that only Earth's atmosphere deviates significantly from its original network. Interestingly the number of cycles goes up. That means that if the actual reaction directions are result of a self organization process in the atmospheric chemistry, it self organizes towards fewer cycles.

*The atmosphere self organizing towards less cycles might be related to the finding of a less complex modern atmosphere in Chapter 4.*

For a comparison of the artificial reaction networks with real thermodynamic data we use a table of reaction free energies ($\Delta_r G^0$) of reactions in glycolysis [Li+11]. In our networks this corresponds to $\Delta\mu_j^0 = \sum_j N_{ij}\mu_i^0$. Because there is no way to assign a unique reaction direction to the reactions in the artificial networks, we are only comparing the distributions of absolute values $|\mu_j^0|$. The normalized (mean set to one) cumulative distributions show a more localized distribution with a wider tail for the data from glycolysis (Fig. 3.4). The distribution for the artificial networks is over all more regular. The bimodal distribution for the data from glycolysys might be related to the fact that it describes two distinct processes, the tricarboxylic acid cycle and the pentose phosphate pathway.

*If those two work on different energy scales, this would explain the bimodal distribution.*

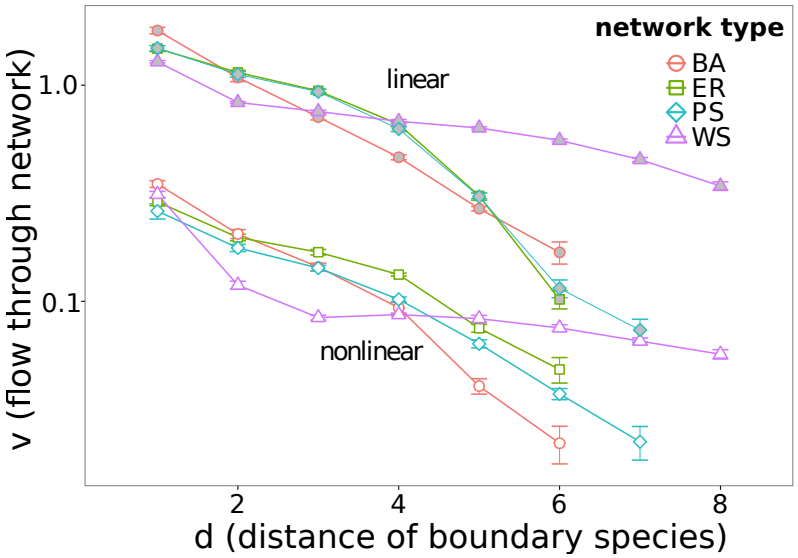



Figure 3.5: **The flow $v$ through the network depending on boundary species distance $d$.** All networks are simulated with a boundary concentration difference of $|c_1 - c_2| = 0.9$ and a base concentration of $\min(c_1, c_2) = 0.1$. Filled (grey) symbols represent linear networks, empty (white) the nonlinear ones. Error bars show the standard error of the mean.

### 3.2.2 *Distance Dependency of Flow*

To characterize the strength of the steady state flow for different network types, we start with the intuitive assumption that the main factor determining the flow is the distance between the two boundary species in the reaction network, measured by shortest path length $d$ in the substrate graph. The dependency of the mean flow on shortest path length is shown in Fig.3.5.

The flow through reaction networks created with small-world and clustering topology (Watts-Strogatz model) shows to be especially weakly dependent on boundary species distance $d$. In the linear as well as the nonlinear case these networks have a lower mean flow for small $d$ ($\leqslant 4$) while for larger values of $d$, they have generally a larger flow than the other networks. We hypothesize that the flow for boundary points whose distance is close to the diameter is limited by the sparse connection of those boundary species to the network. The high clustering of Watts-Strogatz networks (cf. Table 3.1) apparently leads to their exceptional high flow for boundary points with a large distance $d$. This also agrees with the low sensitivity to boundary species distance that the Watts-Strogatz networks show.

The linear networks generated out of the Erdős-Rényi model and those generated with the Pan-Sinha model show a strikingly similar behaviour. This may be due to their similar degree distribution (not shown).

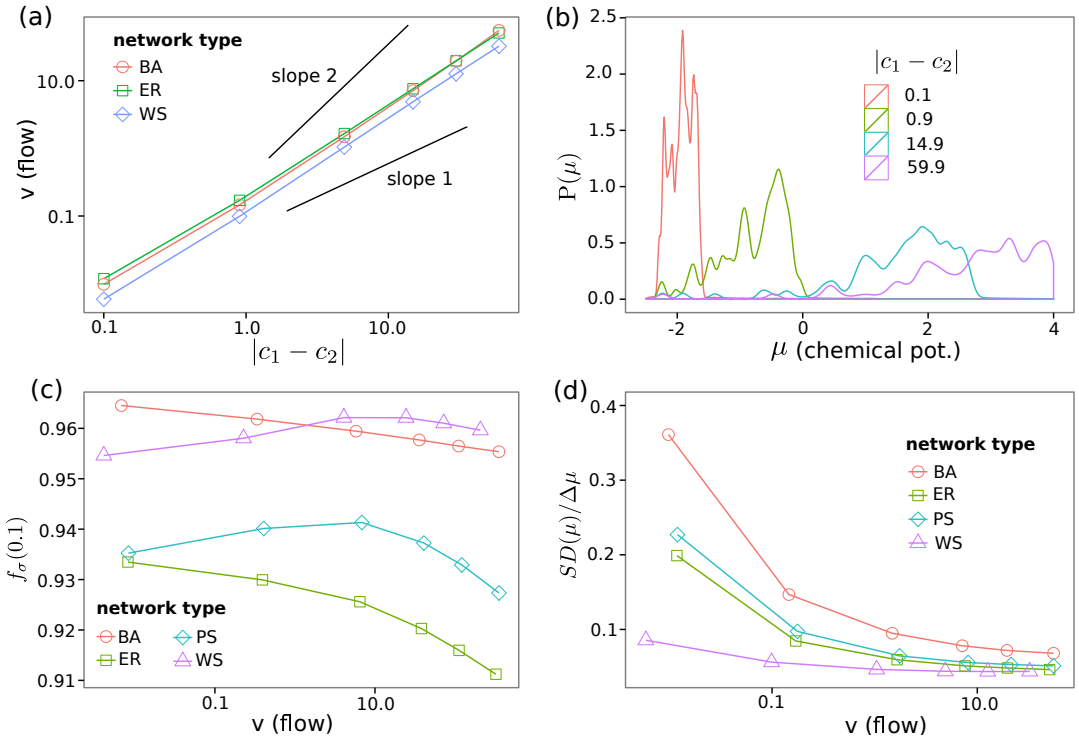



Figure 3.6: **Varying flow through large nonlinear networks.** Each data point is the average of all simulations with specific boundary species concentration ($c_1 = 0.1$ $c_2 = 0.2 \ldots 60$) and a shortest path between boundary species of 3. **(a)** Dependency of flow from concentration difference. Pan-Sinha results are not shown as they overlap with the Erdős-Rényi ones. **(b)** Distribution of species chemical potential $\mu_i$ for different boundary condition strengths of Barabási–Albert (BA) networks. **(c)** The fraction of dissipation in the network explained by the most dissipating 10 percent of reactions, $f_\sigma(0.1)$. **(d)** Standard deviation of chemical potentials $\sigma_\mu$ normalized by difference between boundary species' potentials $\Delta\mu = |\mu_{b_2} - \mu_{b_1}|$ shows a more localized distribution of chemical potentials for larger flows.

### 3.2.3 *Varying Flow for Nonlinear Networks*

Unlike in linear networks, the flow and dissipation distribution in nonlinear networks depend on the absolute concentrations of the boundary species. For the variation of boundary concentration, flow dependency of the concentration difference is in an intermediate regime (Fig. 3.6 (a)) and the slope in log-log plot takes a value between 1 and 2. This is plausible since the network consists of a mix of linear reactions and nonlinear reactions with at best quadratic behaviour. Theoretically a stronger than quadratic dependency of flow from concentration difference would be possible for a specific boundary condition and a specific concentration range, but this possibility seems not to influence the mean behaviour.

We look at the distribution of chemical potentials $\mu = \mu_0 + \ln(x_i)$ inside the reaction network for different strengths of the boundary condition. In Fig. 3.6 (b) the distributions $P(\mu)$ are shown for the sim-

ulated Barabási-Albert (BA) reaction networks with boundary species distance d of 3. The distributions in general are localized between the chemical potentials of the boundary species $\mu_{b_1} = \ln(x_{b_1})$ and $\mu_{b_2} = \ln(x_{b_2})$ While the distributions are almost uniform in this range for low flows, at higher flows the distributions are more shifted towards the upper part. Normalizing the standard deviation $\sigma_\mu$ by $\Delta\mu = |\mu_{b_2} - \mu_{b_1}|$ confirms this finding (Fig. 3.6 (d)) and shows a narrower distribution relative to the chemical potentials of the boundary species.

*Remember that $\mu_0$ for boundary species is set to zero.*

The distributions of dissipation values of the reactions are to noisy to find out if they also get narrower for higher flows. Thus, we calculate the fraction of the dissipation explained by the 10% of reactions with the highest dissipation, $f_\sigma(0.1)$. We see that with higher flows the fraction of dissipation explained by these 10 percent of the network decreases (Fig. 3.6 (c)). The networks generated from the Watts-Strogatz (WS) and the Pan-Sinha (PS) networks show an increase of $f_\sigma(0.1)$ for lower values, but above a flow of around 5 they also decrease. Put differently, for higher flows one needs a larger part of the network to explain a given fraction of its dissipation. Together with the narrower distribution of chemical potentials we interpret this as the thermodynamic disequilibrium leading to a tighter coupling of the reaction network. This coupling leads to the chemical potential of different species to be closer and to the dissipation being more evenly distributed among reactions.

*Note that this happens under the specific boundary conditions of a flow network that adapts its mass to higher flows. Networks investigated in Chapter 5 behave differently.*

### 3.2.4 *Flow Dependency of Cycle Number in Nonlinear Networks*

There are many indicators that cycles have an important function in networks [Gle+01; Kre+12; PPS13]. Cycles function as feedback mechanisms and stabilize the dynamics of the system against perturbations. Also cyclicity has been related to thermodynamic efficiency in thermodynamic power cycles [Lay+12]. To check if there is a dependency of the number of cycles from the flow through the networks, we count the number of small cycles (2- and 4-cycles) in the directed substrate graph for different values of $\nu$. Note that even if the simulated reactions do not change, a change in the effective flow of a reaction can imply a change of direction and by this a change in the directed substrate graph.

The number of small cycles is dependent on local topological properties of the network models. Thus, for evaluation we subtract the number of cycles found in networks with randomly chosen reaction directions (Table 3.1). For all network types we find a clear increase in the number of cycles with increasing flow (Fig. 3.7). This formation of additional cycles can be understood as the network self-organising in thermodynamic disequilibrium to increase its flow and dynamic stability. This supports the idea of a closer coupling of the network with higher degree of disequilibrium that was presented in the previous section.

*This analysis only considers the reaction direction but not the reaction rates (strength of cycles). Strength of cycles will be included in the following chapters through using pathway analysis.*

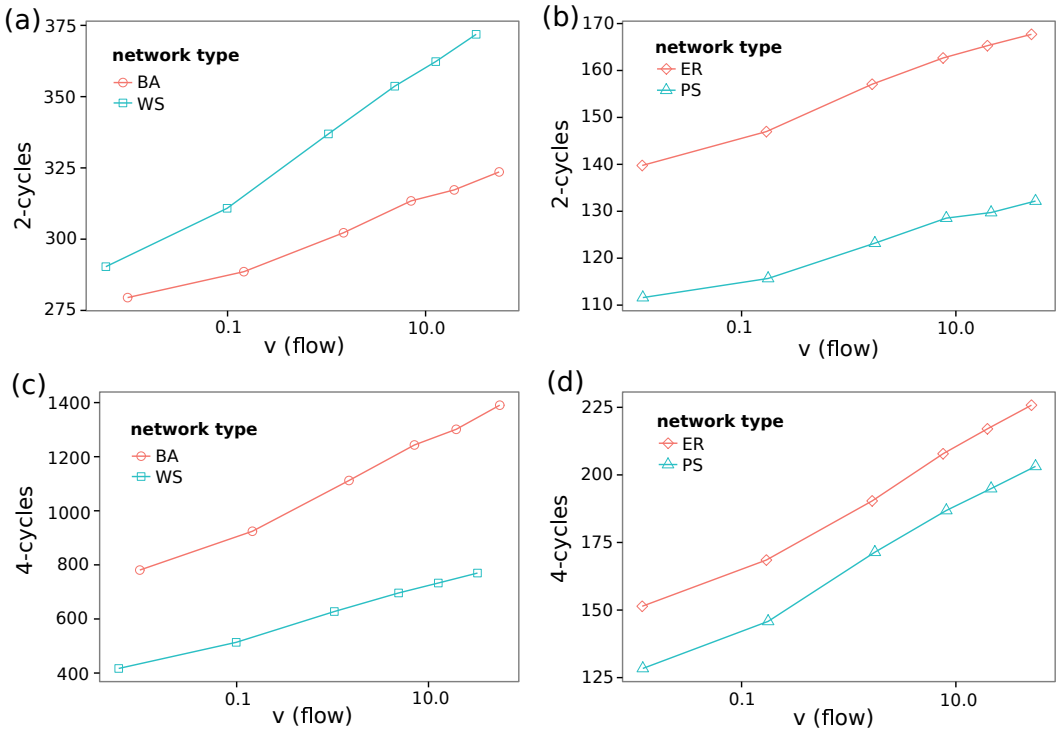



Figure 3.7: **Number of 2- and 4-cycles in the (directed) substrate graphs of the nonlinear reaction networks.** The plots show the number of additional cycles depending on the flow through the network in comparison to the same network with random reaction directions (Table 3.1). Each data point is the average of all simulations with boundary points distance of 3 and fixed boundary concentrations ($c_1 = 0.1$ $c_2 = 0.2 \dots 60$).

### 3.2.5 *Distribution of Entropy Production Rates*

To see how dissipation is distributed inside of the networks, we calculate the entropy production rate for the individual reactions $\sigma_i$ (Eq. 2.22) and look at their distribution for specific network topologies and boundary conditions. To better see the power law dependency, instead of $P(\sigma)$ we plot the cumulative distribution $1 - \int_{-\infty}^{\sigma} P(\sigma')d\sigma'$, which describes the probability of the entropy production rate being higher than $\sigma$ [New05].

The distributions show no large qualitative differences between the different network models (Fig. 3.8). The power law in the intermediate regime is differently pronounced in its extent for different network types, but the greatest difference is clearly seen between the slopes of linear and nonlinear networks. Assuming that $P(\sigma)$ follows a power law, we get an exponent of about $-1.5$ for linear networks and of $-1.66$ for nonlinear networks. The steeper slope of the nonlinear networks can be interpreted as an effect of their reactions being better coupled. This can be seen by the fact that nonlinear ($A + B \rightleftharpoons C + D$) reactions are not depleting a potential between two species directly but there is always the probability that they increase the potential between two other species. The coupling implies a stronger connec-

tion of the flow between individual reactions and by this a stronger connection with the magnitude of dissipation.

### 3.2.6 *Connectivity Dependence of Dissipation*

To evaluate how the dissipation of a reaction depends on the connectivity of the involved species, for every species we calculate the mean dissipation of all reactions connected to it. Plotting the mean dissipation depending on the degree centrality of the species (in the substrate graph) shows a relatively high dissipation for reactions adjacent to lowly connected species (Fig. 3.9). This effect is more pronounced for nonlinear networks. When looking for reactions with high dissipation we should search in the vicinity of lowly connected

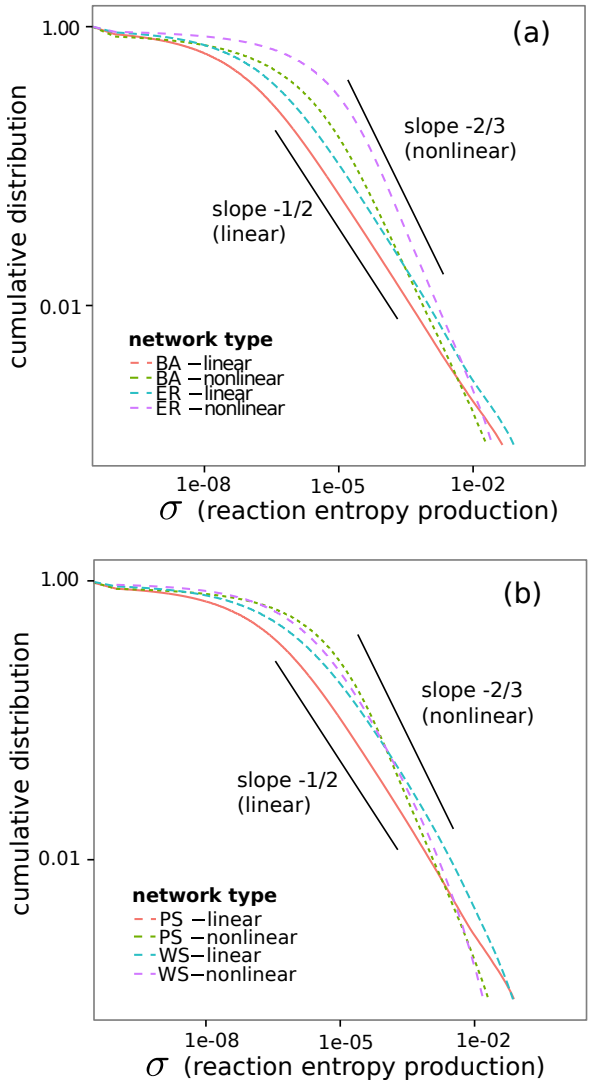



Figure 3.8: **Cumulative distribution of the entropy production of the reactions.** All simulations are performed with boundary concentration values of $c_1 = 0.1$, $c_2 = 1.0$ and a shortest path between boundary species of length 4. **(a)** Distributions for Barabási-Albert (BA) and Erdős-Rényi (ER) networks. **(b)** Distributions for Watts-Strogatz (WS) and Pan-Sinha (PS) networks.

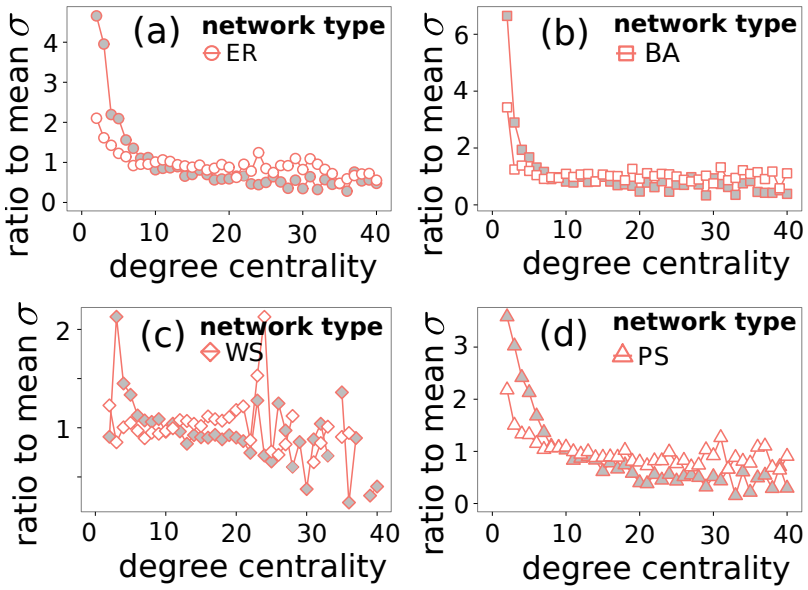



Figure 3.9: **Mean entropy production σ associated with nodes of degree** f. Values are normalized by mean entropy production in the specific network. Grey filled points show nonlinear networks, white filled points show linear networks. Data was taken from all simulation runs of the specified network type with $\min(c_1, c_2) = 0.1$, $|c_1 - c_2| = 0.9$ and shortest path $d = 4$.

species. This can be explained by the stronger connection between reactions generating and consuming the species. When the rate of a reaction that produces a species is increased, the additional flow has to be distributed to all consuming reactions. If there are many consuming reactions, there are more potential pathways to forward the flow while keeping the mean dissipation rate low.

*To illustrate this, one might think of an electric network which is going to overheat at a bottleneck.*

## 3.3 CONCLUSIONS

The simulated random reaction networks under thermodynamic constraints have provided insight into how energy is dissipated in complex reaction networks in thermodynamic disequilibrium. We observe a clear difference between linear and nonlinear networks. The power law has an exponent of $\approx -1.5$ for linear and a slightly lower exponent of $\approx -1.66$ for nonlinear networks. However, there are no qualitative differences between the distributions of entropy production rate for different complex network models like Erdős-Rényi, Barabási-Albert, etc. (Fig. 3.8). The differences between thermodynamic reaction networks of different topologies are more pronounced in the flow (Fig. 3.5) than in the other properties investigated.

We found that a greater thermodynamic disequilibrium in nonlinear reaction networks is associated with a more tightly coupled network. For a greater flow, the network self-organises and maintains a higher number of cycles (Fig. 3.7). A greater flow also leads to a narrower distribution of chemical potentials (Fig. 3.6 (b), (d)). This is associated with results that suggest that for higher flows, a larger

fraction of the network is necessary to explain a given fraction of its dissipation (Fig. 3.6 (c)). We interpret this as an increase in the complexity of the system that comes along with a higher thermodynamic disequilibrium.

Finally, we found that reactions involving lowly connected species tend to dissipate more energy, which is more pronounced in non-linear networks, but is also found in linear networks (Fig. 3.9). This might help to identify reactions that play central roles in the energy dissipation of a complex reaction network.

We also showed how our artificial networks share topological properties with real reaction networks. The artificial networks are topologically more similar to Earth's atmospheric chemistry and Yeast's metabolism than to the two investigated combustion chemistries. The main discriminating factor here is the high density of those two combustion chemistries. The distribution of thermodynamic parameters in the artificial networks only roughly matches data from reactions of glycolysis (Fig. 3.4). Obviously, the amount of thermodynamic data (37 reactions) is quite limited. Current progress in bioinformatical methods to reconstruct thermodynamic data [Can14; Cha+13; DM13] may improve the availability of such data in future and allow a better analysis.

Nevertheless, a fundamental problem of such a comparison remains. It is the way the data of reaction networks is obtained. In networks from chemical models, experimentalists and modellers have made a decision on which reactions they deem relevant. Experimentalists only find reactions that are occurring and are measurable in the systems they investigate. Also the modellers might just decide to exclude reactions with low reaction rates from their models. Hence, the reaction network taken from a model is already biased with respect to the purpose of the model. Our approach with artificial networks, however, assumes the artificial network is a set of (hypothetically) possible reactions. Which reactions become important emerges from the dynamics and can be different depending on the boundary conditions.

A solution for this mismatch between different approaches could be found in future development in automated network generation. Methods using quantum chemical and molecular dynamics simulations to construct reaction networks have been explored in recent years [Wan+14; Wan+16; Dön+15]. When these techniques can generate big, meaningful networks, these can be used in computer-aided modelling of real system as well as reference to evaluate artificial chemistries.

# 4

# THERMODYNAMICS OF ATMOSPHERIC REACTION PATHWAYS

Two models of Earth's atmosphere (modern, Archean) are compared in terms of their chemical reaction pathways. Pathways of the modern Earth atmosphere are simpler (less reactions, fewer cycles). Estimating thermodynamic properties of the pathways shows that the Archean atmosphere is stronger driven by radiation. The modern atmosphere obtains a larger part of its energy through matter exchange (chemical energy). This novel approach allows to analyse and distinguish complex reaction systems with a pathway oriented view on thermodynamics.
**keywords:** atmospheric chemistry, Earth, Archean Earth, reaction networks, reaction pathways, thermodynamics

*This chapter makes use of most of the methods from Chapter 2 (reaction networks, thermodynamics, reaction pathway analysis).*

## INTRODUCTION

Since Lovelock proposed the idea of using atmospheric thermodynamic disequilibrium as a general way of remote life detection [Lov65; HL67] non-equilibrium thermodynamics of planets has been an exciting topic. The idea has been fascinating to many, because it connects two very fundamental questions with a macroscopic physical theory [LE08]. These questions concern the reason for the existence of life and how our planet manages to be so well suited to maintain it.

Alongside many considerations regarding extraterrestrial life, like its likelihood and detectability [CL16], progress in geology and modelling in the last decades lead to a constant refinement of our view of our own planet's history [Holo6; CK07]. Atmospheric oxygen levels were below $10^{-4}$ of the present atmospheric level (PAL) until 2.4 billion years ago. From there on the atmospheric oxygen level and the oxidation state of the oceans increased until reaching modern levels around 600 million years ago. The exact trajectory of this transition is still not conclusively resolved [LRP14]. Yet the increase in oxygen in the atmosphere with a mostly reduced solid Earth suggests an evolution towards greater chemical disequilibrium which is associated with evolution towards more complex forms of life [LPW16].

*Much of recent progress made in geology is related to the analysis of isotope distributions.*

Regarding the detectability of life by atmospheric chemical disequilibrium there is an important distinction to be made between the magnitude of the disequilibrium and the strength of the force maintaining it [SVK13]. The work by Simoncini, Virgo, and Kleidon [SVK13] also showed that the biological power maintaining the chemical disequilibrium of the atmosphere is of the same order of magnitude as abiological geochemically driven processes at Earth's surface. Recent

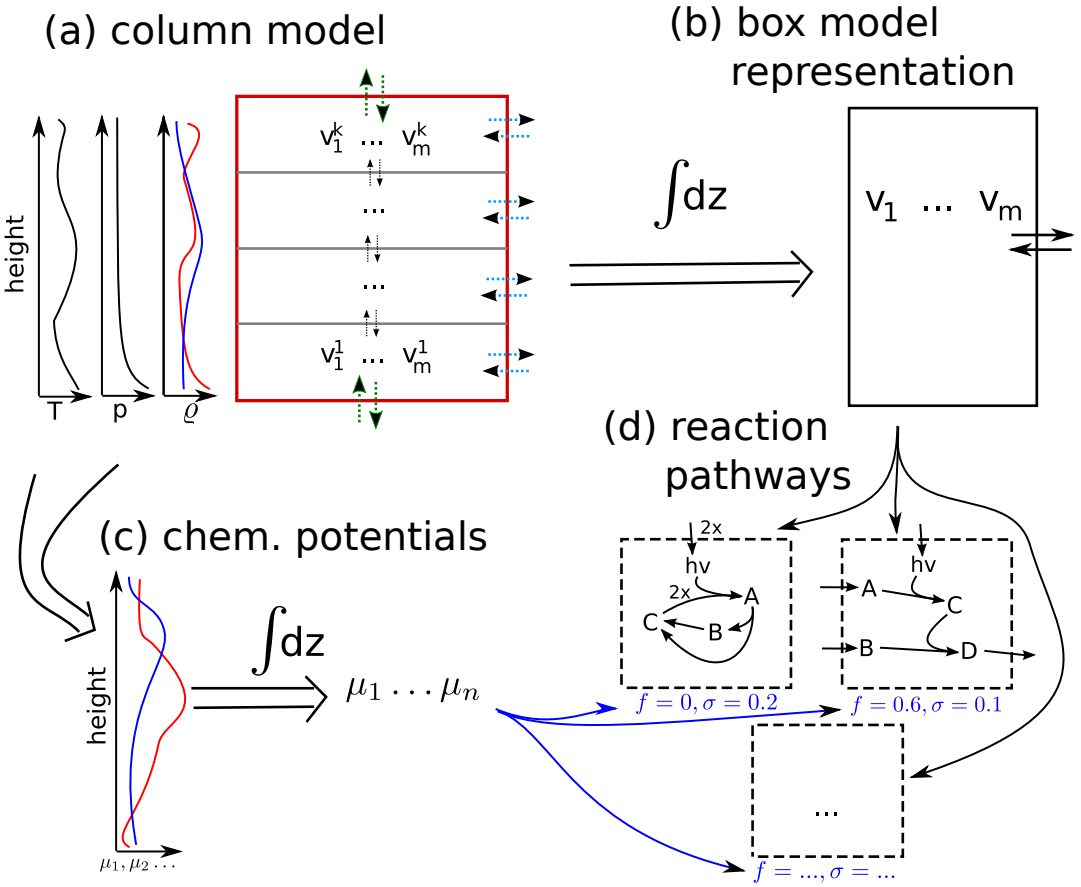



Figure 4.1: **Sketch showing how the data of the two atmospheric chemistry models is processed.** The reaction rates of the gas phase chemistry **(a)** are vertically integrated to get a box model representation **(b)**. Integrated reaction rates are assumed to represent the steady state. This steady state is decomposed into reaction pathways **(d)**. Concentration, pressure and temperature profiles **(a)** are used to calculate profiles for chemical potentials **(c)**. From these an effective chemical potential in the box model representation is derived. This allows a thermodynamic characterisation of the individual reaction pathways.

*This suggests one needs information on ocean chemistry to use thermodynamics as proxy for life on exoplanets.*

work [KTBC16] quantifies thermodynamic disequilibrium of various planets and finds that Earth has a distinctively high disequilibrium if one includes oceanic chemistry.

Another approach to discriminate different types of planets by their atmospheres is in the realm of looking at the topological structure of atmospheric chemical reaction networks. In the field of complex network theory a basic comparison of planetary atmospheric chemistries [YD98] and other reaction networks has been done by Solé and Munteanu [SM04b]. Gleiss et al. [Gle+01] found that metabolic networks and networks from atmospheric chemistry have an unusual high amount of short cycles. An important recent work was the inclusion of reaction directions in the topological analysis of atmospheric chemistries [Est12]. This was done by using a concept called network returnability, an application of statistical physics to characterise network topology. Artificial networks self organise to form additional cycles in higher thermodynamic disequilibrium [FKD15] and cycles

also have been related to dynamic stability [Kre+12]. This suggests that topology and especially cycles are important for understanding chemical disequilibrium in Earth's atmosphere.Earth's atmosphere

The goal of this chapter is to introduce a method that allows to reveal the structural features of thermodynamics in such complex reaction networks. A novel element of this approach is not only using topology and reaction direction, but also including reaction rates as input of the analysis. Such rates can be taken from models that describe the chemical composition of the atmosphere. For this work we choose two atmospheric chemistry models from the group of James Kasting. The first model describes the composition of an atmosphere of early Earth without an extensive biosphere [PBK01; KKS05; Seg+07]. The second model describes the composition of modern day Earth's atmosphere [KHP85; PK02]. Both models are column models and use overlapping sets of reactions. Additionally they include some reactions, chemical species, and parametrisations for boundary conditions that are specific to the respective planetary situation. Because of this, besides comparing the dynamical steady state of the models, we also compare their structure. This is in accordance with the implicit assumption that the modellers have chosen an appropriate model structure for the specific planetary state. From these column models we derive a box model representation and effective chemical potentials for this representation. Fig. 4.1 sketches how the data from these original models is used to obtain a better topological understanding of their respective atmospheric chemistries. For simplicity we will refer to properties of the "modern atmosphere" and the "Archean atmosphere" in this chapter, meaning the two models that we investigate.

*An overview of models and pre-procesing of the data is given in Appendix B.1.*

*An introduction in methods used here is given in Chapter 2.*

The following section provides a short recapitulation of reaction pathways and the role thermodynamics plays for their understanding. An interpretation for reaction pathways in the context of atmospheric chemistry is given. The results are structured in a general presentation of the pathways found, their thermodynamic properties, and an interconnection of their properties with the boundary fluxes of the atmosphere. We conclude by explaining how reaction pathways and thermodynamics together can be used to analyse atmospheres for signatures of life on the basis of our model comparison. Processing of the data from the atmospheric chemistry models for analysis and calculation of chemical potentials is given in the appendix of the thesis (Sect. A.2 and Sect. B.1).

## 4.1 METHODS

To compare the two states of the atmosphere we use classic reaction kinetic modelling[1] and existing thermodynamic data [MGR93] in combination with elementary mode type pathway analysis [SS93]. We use the algorithm for determining all reaction pathways relevant for explaining a certain fraction of the simulated steady state fluxes.

[1] *This has been done by the creators of the chemical models (see Sect. B.1).*

Furthermore these pathways are characterised thermodynamically by estimating the chemical potentials for the chemical species (Fig. 4.1).

### 4.1.1 *Thermodynamic Consistency*

Assuming mass action kinetics for a reaction network, with the knowledge of thermodynamic data the parameters of the kinetic law $\mathbf{v}(\mathbf{x})$ are specified and the dynamics of the system can be calculated.[2] But even with a rate vector $\mathbf{v}$ that fulfils the steady state condition $\mathbf{N} \cdot \mathbf{v} = 0$ from some other kinetic law, one can use thermodynamics to check the consistency of a model [Ksc10]. Thermodynamic consistency for our purposes means that the reaction direction matches with the chemical potentials of the species. Formally this means that the reaction rate $\nu_i$ and the change of chemical potential $\Delta\mu_i$ of a reaction must have different signs:

[2] *See Chapter 3.*

$$\nu_i \cdot \Delta\mu_i = \nu_i \cdot \sum_j N_{ji}\mu_j \leqslant 0. \tag{4.1}$$

This implies that all reactions have to be understood as in principle reversible reactions for which the rate $\nu_i$ corresponds to the effective, thermodynamically favoured rate in the specific state of the system. As no reaction can spontaneously increase the chemical energy (Gibbs free energy) of the system, Eq. 4.1 has to hold for all reactions.

### 4.1.2 *Reaction Pathway Analysis*

For analysing steady states of the reaction networks we use their decomposition into reaction pathways as shown in Section 2.2.3. Reaction pathways are collections of reactions that can operate together in steady state (without changing concentration of species) and can not be decomposed further. Steady state vectors $\mathbf{v}$ can be represented as a linear combination of reaction pathways $\mathbf{E}^{(i)} \in \mathbb{N}^{|\mathcal{R}|}$ with positive coefficients $\alpha_i$:

*Cf. Eq. 2.14.*

$$\mathbf{v} = \sum_i \alpha_i \mathbf{E}^{(i)}.$$

The coefficients $\alpha_i$ are not uniquely defined by the equation above. We use the method presented in Sect. 2.3.4 to calculate $\alpha_i$, which leads to coefficients that are unique for a given set of elementary modes. The algorithm has the additional advantage of working well if not all reaction pathways have been found and if reaction rates span many orders of magnitudes. The description of the algorithm used to calculate all significant reaction pathways can be found in Sect. 2.3.

***This interpretation of pathways will be relevant in the following chapters!***

This algorithm is a slightly modified version of Lehmann's algorithm [Leh04]. Yet the work presented here does not only differ in details of the used algorithms but also comes with a different interpretation of the meaning of an elementary mode or reaction pathway. In bioinformatics the technique is widely used for analysing metabolic networks [TWS09; Pap+03]. The reactions are enzymatic reactions that are controlled by genes. Thus a found pathway is associated with a set of genes and can be understood as a product of

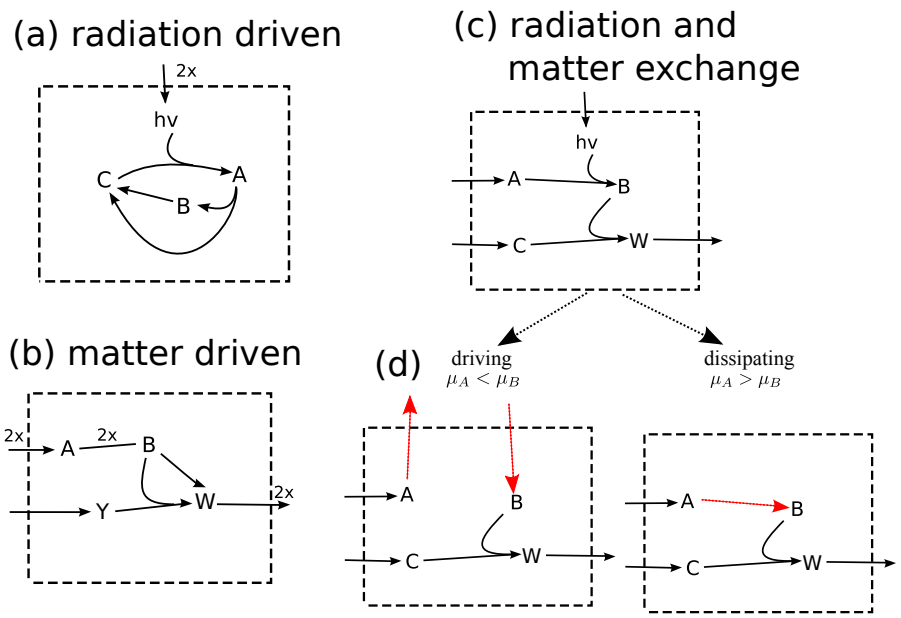



Figure 4.2: **Pathways found in atmospheric chemistry can be thermodynamically divided into three categories,** even without knowledge of exact thermodynamic potentials. **(a)** Pathways only driven by photochemical reactions that do not exchange any matter with the boundary of the atmosphere. **(b)** Pathways that do not include any photoreactions, but are instead driven by matter exchange across the boundary of the system. **(c)** Pathways that include both, photochemical reactions and matter exchange. These can in theory transform radiative energy into chemical energy. To decide if this is the case one needs to know the chemical potentials of the species. **(d)** For thermodynamic analysis of the mixed pathways, photoreactions are substituted with marked pseudoreactions that maintain their effect. If the photoreaction increases the chemical energy of the system, multiple marked exchange reactions are used. Otherwise the photoreaction is replaced by one single dissipating reaction.

evolution. It represents a mechanism with certain objectives in an organism.

In atmospheric chemistry there is no direct effect of information processing as it is present in biological systems that are formed by evolution. But the biosphere might be able to shift the concentration of certain chemicals as it happened with oxygen on Earth. This change in concentration indirectly influences the pathways present in the system. Additionally some pathways may be directly driven by the biosphere through matter exchange. The difference of the work of this chapter and the work by Lehmann [Leho4; Gre+06], whose algorithm we adapted, is that we try to pronounce this analogy to pathways usage in bioinformatics. We treat all species as intermediate species and then obtain pathways that describe how the reaction networks connects in- and outflow of the system (Fig. 4.3 (a)). Lehmann's existing algorithm treats only shorter lived species as intermediates and then obtains pathways that connect longer lived species inside the reaction network.

### *Pathway Types and Thermodynamics*

With this interpretation of pathways as structural units that connect the inflow and outflow of the atmosphere we can already characterise three types of pathways that can be found (Fig. 4.2). This distinction relates to the type of exchange, whether the pathway contains photochemical reactions or mass exchange. Every pathway has to be connected to the environment of the system by some exchange process to obtain the chemical energy that drives the chemical reactions and is dissipated.

*Finding pathways without any interactions indicates thermodynamic inconsistencies in the model.*

Pathways that are driven by photoreactions without exchanging matter (Fig. 4.2 (a)) are purely dissipative in terms of thermodynamics. This means the energy originating from photonic radiation is transformed into heat by the reactions in the pathway. Pathways without photoreactions have to be driven by chemical potentials (Fig. 4.2 (b)). Overall the chemical energy (Gibbs free energy) through mass exchange has to decrease and its amount must exactly correspond to the energy dissipated by all reactions in the pathway.

Most interesting are those pathways that use radiative energy and transform chemical species (Fig. 4.2 (c)). For these we can define a thermodynamic quantity that we will call *transformation efficiency*. We do not consider the actual chemical potential of the radiation to do this [Wur82], but replace photochemical reactions by pseudoreactions (Fig. 4.2 (d)). The chemical energy necessary for all these reactions to occur is then taken as the radiative energy which is used by the pathways. If we consider the change of chemical energy by all conventional exchange reactions as (chemical) work $\Delta\mu_{ex}$ and the change of chemical energy by photoreactions as radiative energy input $\Delta\mu_{h\nu}$, we can calculate the *transformation efficiency* in analogy to the thermal efficiency of heat engines known from basic thermodynamics:

$$\eta = \begin{cases} \frac{\Delta\mu_{ex}}{\Delta\mu_{h\nu}} & \text{if } \Delta\mu_{ex} \geqslant 0 \\ 0 & \text{else} \end{cases} . \tag{4.2}$$

The second part of the equation captures the case of no work being done by the pathway. This case is possible even if photoreactions are driving the pathway. In analogy to thermodynamic efficiency the transformation efficiency tells us how much of the energy[3] absorbed by the system is transformed into chemical energy usable by some other processes outside of the atmosphere.

[3] *Classically this is heat, here the energy is radiation.*

### *Cycles In Pathways*

The found pathways are characterised in terms of cycles of reactions that in part transform a starting compound back to its original form (Fig. 4.3 (b)). Investigation of such cycles in biological and chemical systems, their abundance and formation, has a long history [Mor66; Gle+01]. In comparison to much of the existing work in this area, the analysis through pathways allows to weight the importance of each cycle found in the network using the importance of associated pathways. Cylces are calculated similar to the work presented in Chapter 3.

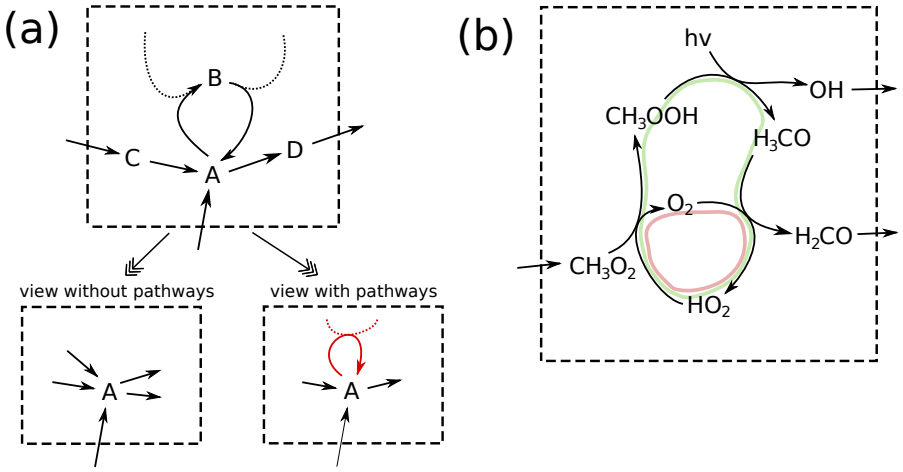



Figure 4.3: **Pathway analysis gives a more detailed view on production and consumption of species. (a)** Pathways allow to separate contribution of cycles inside the network and use this to better understand its function. **(b)** For every pathway the number of cycles can be counted individually and then interpreted weighted with the pathways' coefficient $\alpha_i$. The shown atmospheric pathway contains two cycles of length two and three. The length of a cycle corresponds to the number of chemical species in it.

Instead of doing this for the entire network, the cycles are calculated for the subnetwork defined by each pathway. Cycles are defined in the substrate graph of the subnetwork which is obtained by defining the chemical species as nodes and inserting directed edges between all reactants and products of the reactions occurring in the pathway. An example of a pathway for the modern atmosphere with two cycles is given in Fig. 4.3 (b). To limit computational complexity we calculate only cycles with four edges or less. Cycles are defined by the order of the chemical species that they contain. Even if there are multiple identical cycles from different reactions this is only counted as one cycle.[4]

*Cycles are calculated like in Fig. 2.2 (e) here.*

*[4] This differs from Chapter 3, where multiple edges are considered (Fig. 2.2 (d)).*

## 4.2 RESULTS

### 4.2.1 *Reaction Pathways*

Despite the fact that the algorithm we use reduces the complexity of finding all reaction pathways to finding those that have a large contribution to the steady state it still requires substantial computing time. This is mostly because we want to have a decomposition of the steady state into reaction pathways that explain a large fraction of most reaction rates and these span more than 10 orders of magnitude. Thus, we have to choose quite strict parameters. For the model of the modern atmosphere we choose $r = 0.1$, $f_{exp} = 5 \times 10^{-3}$. For these parameters the decomposition takes roughly a day to finish on a a single Xeon CPU at 2.90Ghz. We find 30,157 reaction pathways and corresponding coefficients that decompose the steady state perfectly (error below numerical accuracy). For the Archean atmospheric model the

*See Sect. 2.3.1 for an explanation of the parameters.*

parameters are $r = 0.3$, $f_{exp} = 10^{-3}$. Here, the analysis takes a comparable amount of time and results in 24,627 reaction pathways that also decompose the respective steady state exactly.

As described in Sect. 4.1.2 we can differentiate the pathways into three types, depending on whether they contain photoreactions and exchange matter with the outside of the gas phase chemistry. The least common ones are the pathways that do not contain any photoreactions but are driven solely by the flux of matter. For the modern (high $O_2$) atmosphere there are 39 pathways of this type that together only explain a fraction of $1.67 \times 10^{-06}$ of the steady state rates. For the Archean (high $CO_2$) atmosphere there are just two pathways that explain $3.59 \times 10^{-8}$ of the steady state. Most common in numbers are the pathways that contain photoreactions as well as exchange of matter. There are 29,778 of this type for the modern atmosphere and 24,384 for the Archean one. Nevertheless these explain only a fraction of $6.1 \times 10^{-4}$ and 0.073 of their respective steady states.

For both atmospheres the most important pathways of the different types are shown in Table 4.1 (modern atmosphere) and Table 4.2 (Archean atmosphere). They are ordered according to their importance for explaining their respective steady states (cf. Eq. 4.3). A visualisation of samples of important pathways of the three different types for the high $O_2$ atmosphere is given in Fig. 4.4.

Table 4.1: **The three most important pathways of each pathway type (Fig. 4.2) for the modern atmosphere.** $\Delta\mu_{h\nu}$ is the photochemical energy used by the pathway and $\Delta\mu_{ex}$ the chemical work (matter exchange) done by the pathway. Both are given in $W/m^2$ scaled by the pathway coefficient $\alpha_i$. ("NA" indicates values that are undetermined because of missing thermodynamic data. "*" pathways purely dissipate radiation to heat.)

| | | | | #cycles | $\Delta\mu_{h\nu}$ | $\Delta\mu_{ex}$ |
|---|---|---|---|---|---|---|
| | $O + O_2 + M$ | $\rightarrow$ | $O_3 + M$ | | | |
| | $O_3 + hv$ | $\rightarrow$ | $O_2 + O$ | (1) | 5 | 2.03 | 0 |
| net*: | $hv$ | $\rightarrow$ | | | | | |
| | $O + O_2 + M$ | $\rightarrow$ | $O_3 + M$ | | | |
| | $O(_1D) + N_2$ | $\rightarrow$ | $O + N_2$ | | | |
| | $O_3 + hv$ | $\rightarrow$ | $O_2 + O(_1D)$ | (2) | 5 | 2.23 | 0 |
| net*: | $hv$ | $\rightarrow$ | | | | | |
| | $O + O_2 + M$ | $\rightarrow$ | $O_3 + M$ | | | |
| | $O(_1D) + O_2$ | $\rightarrow$ | $O + O_2$ | | | |
| | $O_3 + hv$ | $\rightarrow$ | $O_2 + O(_1D)$ | (3) | 7 | 0.90 | 0 |
| net*: | $hv$ | $\rightarrow$ | | | | | |
| 2X | $O + O_2 + M$ | $\rightarrow$ | $O_3 + M$ | | | |
| | $O_2 + hv$ | $\rightarrow$ | $2O$ | (23) | 1 | $3.5 \times 10^{-4}$ | $2.6 \times 10^{-4}$ |
| net: | $hv + 3O_2$ | $\rightarrow$ | $2O_3$ | | | | |
| 2X | $H_2O + O(_1D)$ | $\rightarrow$ | $2OH$ | | | |
| 2X | $H + O_2 + M$ | $\rightarrow$ | $HO_2 + M$ | | | |
| 2X | $O + O_2 + M$ | $\rightarrow$ | $O_3 + M$ | | | |
| | $O_2 + hv$ | $\rightarrow$ | $2O$ | | | |
| 2X | $O_3 + hv$ | $\rightarrow$ | $O_2 + O(_1D)$ | (24) | 8 | $1.4 \times 10^{-4}$ | $-8.6 \times 10^{-5}$ |
| 2X | $CO + OH$ | $\rightarrow$ | $CO_2 + H$ | | | |
| 2X | $CH_3O_2 + HO_2$ | $\rightarrow$ | $CH_3OOH + O_2$ | | | |
| 2X | $CH_3OOH + OH$ | $\rightarrow$ | $CH_3O_2 + H_2O$ | | | |
| net: | $3hv + O_2 + 2CO$ | $\rightarrow$ | $2CO_2$ | | | | |

(continued on next page)

Table 4.1: (continued - pathways of modern atmosphere)

| | | | | #cycles | $\Delta\mu_{h\nu}$ | $\Delta\mu_{ex}$ |
|---|---|---|---|---|---|---|
| 2X | $H_2O + O(_1D)$ | $\rightarrow$ | $2OH$ | | | |
| 2X | $H + O_2 + M$ | $\rightarrow$ | $HO_2 + M$ | | | |
| 2X | $OH + HO_2$ | $\rightarrow$ | $H_2O + O_2$ | | | |
| 2X | $O + O_2 + M$ | $\rightarrow$ | $O_3 + M$ | (27) | 7 | $1.4 \times 10^{-4}$ | $-8.6 \times 10^{-5}$ |
| | $O_2 + hv$ | $\rightarrow$ | $2O$ | | | |
| 2X | $O_3 + hv$ | $\rightarrow$ | $O_2 + O(_1D)$ | | | |
| 2X | $CO + OH$ | $\rightarrow$ | $CO_2 + H$ | | | |
| net: | $3hv + O_2 + 2CO$ | $\rightarrow$ | $2CO_2$ | | | |
| | $Cl + O_2 + M$ | $\rightarrow$ | $ClO_2 + M$ | (120) | 1 | 0 | NA |
| net: | $O_2 + Cl$ | $\rightarrow$ | $ClO_2$ | | | |
| | $H + O_2 + M$ | $\rightarrow$ | $HO_2 + M$ | | | |
| | $H + HO_2$ | $\rightarrow$ | $2OH$ | (587) | 3 | 0 | $-1.5 \times 10^{-6}$ |
| 2X | $CO + OH$ | $\rightarrow$ | $CO_2 + H$ | | | |
| net: | $O_2 + 2CO$ | $\rightarrow$ | $2CO_2$ | | | |
| 2X | $H + O_2 + M$ | $\rightarrow$ | $HO_2 + M$ | | | |
| | $OH + O_3$ | $\rightarrow$ | $HO_2 + O_2$ | | | |
| | $O + O_2 + M$ | $\rightarrow$ | $O_3 + M$ | | | |
| 2X | $CO + OH$ | $\rightarrow$ | $CO_2 + H$ | | | |
| 2X | $HCO + O_2$ | $\rightarrow$ | $HO_2 + CO$ | | | |
| 2X | $H_2CO + OH$ | $\rightarrow$ | $H_2O + HCO$ | | | |
| 2X | $CH_4 + OH$ | $\rightarrow$ | $CH_3 + H_2O$ | (1472) | 11 | 0 | $-6.9 \times 10^{-7}$ |
| 2X | $CH_3 + O_2 + M$ | $\rightarrow$ | $CH_3O_2 + M$ | | | |
| | $2CH_3O_2$ | $\rightarrow$ | $2H_3CO + O_2$ | | | |
| 2X | $H_3CO + O_2$ | $\rightarrow$ | $H_2CO + HO_2$ | | | |
| 7X | $NO + HO_2$ | $\rightarrow$ | $NO_2 + OH$ | | | |
| 7X | $NO_2 + O$ | $\rightarrow$ | $NO + O_2$ | | | |
| net: | $8O + 2CH_4$ | $\rightarrow$ | $4H_2O + 2CO_2$ | | | |

Table 4.2: **Three most significant pathways of each pathway type (Fig. 4.2) for the Archean atmosphere model.** The last two pathways were the only ones found that are solely driven by flow of matter. $\Delta\mu_{h\nu}$ is the photochemical energy used by the pathway and $\Delta\mu_{ex}$ the chemical work (matter exchange) done by the pathway. Both are given in $W/m^2$ scaled by the pathway coefficient $\alpha_i$. ("NA" indicates values that are undetermined because of missing thermodynamic data. "*" pathways purely dissipate radiation to heat.)

| | | | | #cycles | $\Delta\mu_{h\nu}$ | $\Delta\mu_{ex}$ |
|---|---|---|---|---|---|---|
| | $H + O_2 + M$ | $\rightarrow$ | $HO_2 + M$ | | | |
| | $HO_2 + O$ | $\rightarrow$ | $OH + O_2$ | | | |
| | $CO_2 + hv$ | $\rightarrow$ | $CO + O$ | (1) | 6 | $8.3 \times 10^{-3}$ | 0 |
| | $CO + OH$ | $\rightarrow$ | $CO_2 + H$ | | | |
| net*: | $hv$ | $\rightarrow$ | | | | |
| | $SO_2 + hv$ | $\rightarrow$ | $_1SO_2$ | | | |
| | $_1SO_2 + M$ | $\rightarrow$ | $SO_2 + M$ | (2) | 3 | NA | 0 |
| net*: | $hv$ | $\rightarrow$ | | | | |
| | $H + CO + M$ | $\rightarrow$ | $HCO + M$ | | | |
| | $HCO + hv$ | $\rightarrow$ | $H + CO$ | (3) | 5 | 0 | 0 |
| net*: | $hv$ | $\rightarrow$ | | | | |
| | $H + O_2 + M$ | $\rightarrow$ | $HO_2 + M$ | | | |
| | $H + HO_2$ | $\rightarrow$ | $2OH$ | | | |
| | $OH + O$ | $\rightarrow$ | $H + O_2$ | | | |
| | $O(_1D) + M$ | $\rightarrow$ | $O + M$ | (4) | 7 | $1.9 \times 10^{-3}$ | 0 |
| | $CO + OH$ | $\rightarrow$ | $CO_2 + H$ | | | |
| | $CO_2 + hv$ | $\rightarrow$ | $CO + O(_1D)$ | | | |
| net: | $hv$ | $\rightarrow$ | | | | |

(continued on next page)

Table 4.2: (continued - pathways of Archean atmosphere)

| | | | | | #cycles | $\Delta\mu_{h\nu}$ | $\Delta\mu_{ex}$ |
|---|---|---|---|---|---|---|---|
| | $CO_2 + hv$ | → | $CO + O$ | | | | |
| 2X | $H + CO + M$ | → | $HCO + M$ | | | | |
| | $2HCO$ | → | $H_2CO + CO$ | | | | |
| 2X | $H_2CO + hv$ | → | $H_2 + CO$ | (31) | 4 | $2.9 \times 10^{-5}$ | $1.2 \times 10^{-5}$ |
| | $CH_3 + O$ | → | $H_2CO + H$ | | | | |
| | $CH_4 + hv$ | → | $CH_3 + H$ | | | | |
| net: | $4hv + CO_2 + CH_4$ | → | $2H_2 + 2CO$ | | | | |
| | $H_2O + hv$ | → | $H + OH$ | | | | |
| | $CO_2 + hv$ | → | $CO + O$ | | | | |
| | $CO + OH$ | → | $CO_2 + H$ | | | | |
| 4X | $H + CO + M$ | → | $HCO + M$ | | | | |
| 2X | $2HCO$ | → | $H_2CO + CO$ | (32) | 6 | $2.6 \times 10^{-5}$ | $5.6 \times 10^{-6}$ |
| 3X | $H_2CO + hv$ | → | $H_2 + CO$ | | | | |
| | $CH_3 + O$ | → | $H_2CO + H$ | | | | |
| | $CH_4 + hv$ | → | $CH_3 + H$ | | | | |
| net: | $6hv + H_2O + CH_4$ | → | $3H_2 + CO$ | | | | |
| | $H_2O + hv$ | → | $H + OH$ | | | | |
| 2X | $CO_2 + hv$ | → | $CO + O$ | | | | |
| 2X | $CO + OH$ | → | $CO_2 + H$ | | | | |
| 4X | $H + CO + M$ | → | $HCO + M$ | | | | |
| 2X | $2HCO$ | → | $H_2CO + CO$ | (34) | 7 | $2.7 \times 10^{-5}$ | $4.7 \times 10^{-6}$ |
| 3X | $H_2CO + hv$ | → | $H_2 + CO$ | | | | |
| | $CH_3 + O$ | → | $H_2CO + H$ | | | | |
| | $CH_4 + O$ | → | $CH_3 + OH$ | | | | |
| net: | $6hv + H_2O + CH_4$ | → | $3H_2 + CO$ | | | | |
| | ${}_1SO_2 + M$ | → | ${}_3SO_2 + M$ | (5897) | 1 | 0 | NA |
| net: | ${}_1SO_2$ | → | ${}_3SO_2$ | | | | |
| | $SO_3 + H_2O$ | → | $H_2SO_4$ | (16883) | 0 | 0 | $-1.3 \times 10^{-13}$ |
| net: | $H_2O + SO_3$ | → | $H_2SO_4$ | | | | |

### 4.2.2 *Topology of Reaction Pathways*

With both models having roughly 30,000 reaction pathways whose coefficients $\alpha_i$ are ranging from around $10^{-28}$ to $10^{15}$, it is important to weight the pathways when comparing their topological properties. Instead of just taking the pathway coefficient $\alpha_i$, we are using the fraction of the steady state explained by it. This leads to the pathways being weighted by the so called *rate fraction*:

*$f_i$ is the fraction of the rate vector explained by pathway $i$, measured by the $L^1$-norm.*

$$f_i = \frac{\alpha_i \sum_j |E_j^{(i)}|}{\sum_k \alpha_k \sum_j |E_j^{(k)}|}. \tag{4.3}$$

By weighting pathway properties with this factor instead of their coefficient $\alpha_i$, a higher weight is put onto pathways with more reactions. For each pathway we calculate the number of unique cycles with a length smaller than five as described above (Sect. 4.1.2, Fig. 4.3). Weighted with $f_i$ this allows to calculate a cycle number for each of the models. This number indicates the importance of cycles in the pathway representation of the network's steady state. A high number indicates complex pathways (many cycles) with high coefficients while a low number indicates that most pathways with higher coefficients have few or no cycles. We get a cycle number of 5.22 for the modern atmosphere and 5.51 for the Archean one. For the weighted

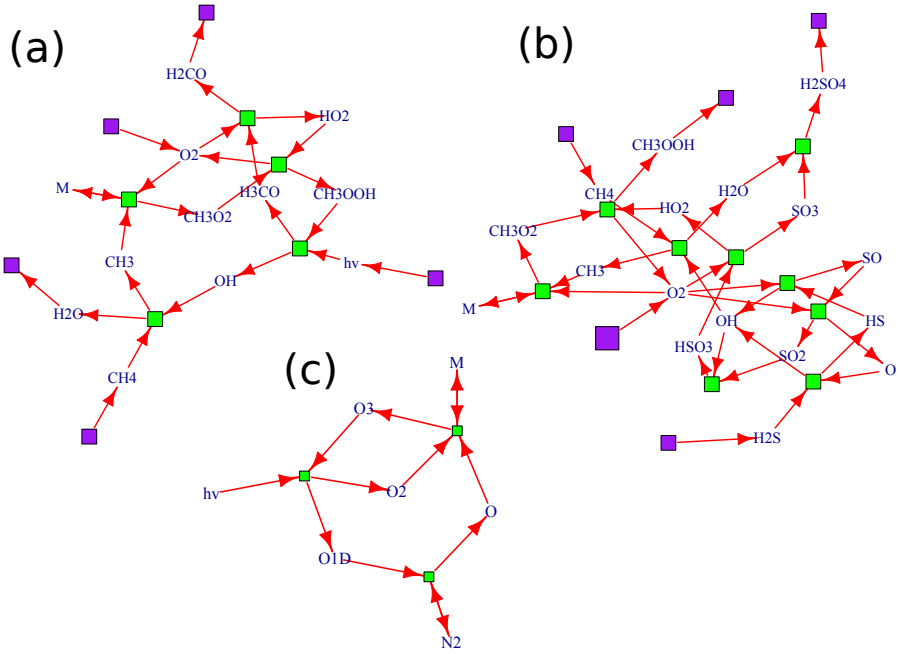



Figure 4.4: **Example of pathways with different thermodynamic characteristics from the model of the modern atmosphere.** Bright green squares represent conventional reactions and dark purple squares in- or outflow reactions. Size corresponds to the coefficient of a reaction in the pathway. **(a)** Pathway number 20,549 is one of a large number of pathways that oxidise methane using light. **(b)** Pathway 680 oxidises methane without using any photoreactions. **(c)** Pathway number 2 (see Table 4.1) dissipates radiation without any exchange of matter through the boundary of the gas phase chemistry of the atmosphere.

number of reactions contained in the reaction pathways the distinction is more pronounced. This value is 3.38 for the modern, and 5.42 for the Archean atmosphere. This means that with respect to this measure the modern atmosphere is slightly less complex.

The cumulative distribution (Fig. 4.5) shows this difference in greater detail. The graph shows the probability of finding a pathway with more cycles / reactions than the value given on the x-axis. All pathways are weighted with their respective $f_i$. In this representation a steep slope corresponds to a high probability in the original distribution. While the difference between distributions for cycles is evenly distributed over a wide range, for the number of reactions the difference is more localised in the shape of the distribution. We see a notable drop of the probability of longer pathways around pathways with five reactions. This drop might be explained by the high importance of ozone related pathways in the modern atmosphere that have a relatively simple structure.

*The first three (ozone related) pathways in Table 4.1 together explain 97% of the steady state.*

### 4.2.3 *Chemical Potentials*

Chemical potentials $\mu$ are calculated from the values of concentrations, temperature and pressure in each simulated atmospheric layer (see Appendix A.2). From these vertical profiles of chemical

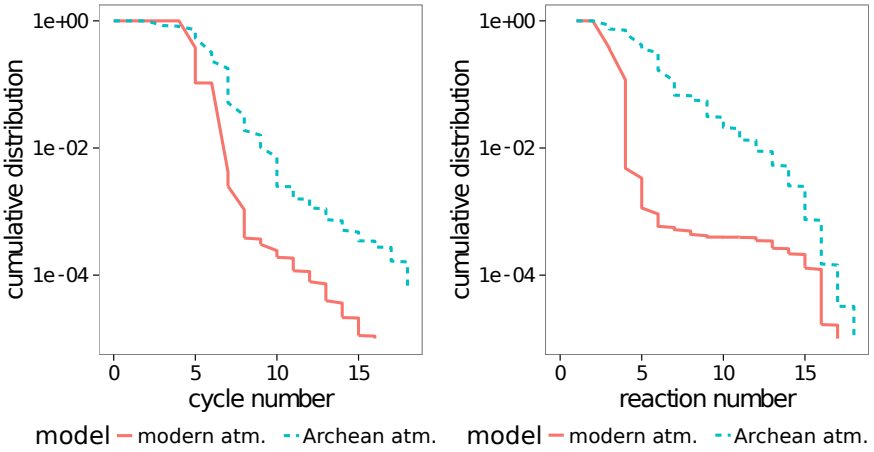



Figure 4.5: **Comparison of atmospheric reaction networks by pathway complexity.** Weighted cumulative distribution of cycles in pathways **(left)** and number of reactions in pathways **(right)**. Value on the y-axis indicates the (weighted) fraction of pathways that has more cycles / reactions than shown on the corresponding value on the x-axis. It can be clearly seen that the Archean atmosphere has a greater amount of complex reactions in the intermediate region (more cycles, more reactions).

potentials an effective potential is calculated by averaging with the concentration of the respective chemical species as weight. Note that these are not chemical potentials in the strict physical sense.[5] Though the simplification made through the box model representation should not matter for reaction processes that happen on timescales much faster or much slower than the timescale of vertical mixing, it is possible that inconsistent pathways in intermediate timescales are found, even if one assumes a column model that is perfectly thermodynamically consistent.

[5] *There is no kinetic law that generates the same dynamics in the integrated box model representation.*

For the model of the modern atmosphere we can calculate the effective chemical potentials for all species except the following: $CH_3OOH$, $HO_2NO_2$, $S_2$, $CH_3O_2$, HS, $ClONO_2$, ClONO, $_1SO_2$, $_3SO_2$. For the Archean atmospheric model the chemical potentials for the following species can not be calculated because of missing data: $CH_3O_2$, HS, $_1SO_2$, $_3SO_2$, $S_3$, $S_4$, $S_8$AER, $C_2H_4OH$, $C_2H_2OH$, HCAER, $C_3H_2$, $C_3H_6$, $C_3H_7$, $C_3H_5$,$NH_2$s, $CH_3C_2H$. Pathways exchanging these species can not be quantified thermodynamically. These unquantifiable pathways will be called "undecidable" in the following paragraph and their contribution to the steady state will be calculated to evaluate the impact of missing thermodynamic data.

*For some complex species even the model itself is unclear about what isomers the species refer to exactly.*

For the calculated chemical potentials of the modern atmosphere model 164 reactions are correct in terms of the reaction directions, 11 reactions are incorrect, and 39 reactions can not be decided (Eq. (4.1)). For the chemical potentials calculated for the second model 276 reactions have correct direction, 11 reactions are incorrect, and 71 reactions can not be decided because of missing values for chemical potentials. To weaken the effect of unknown or erroneous chemical potentials reaction pathways are quantified directly, without calcu-

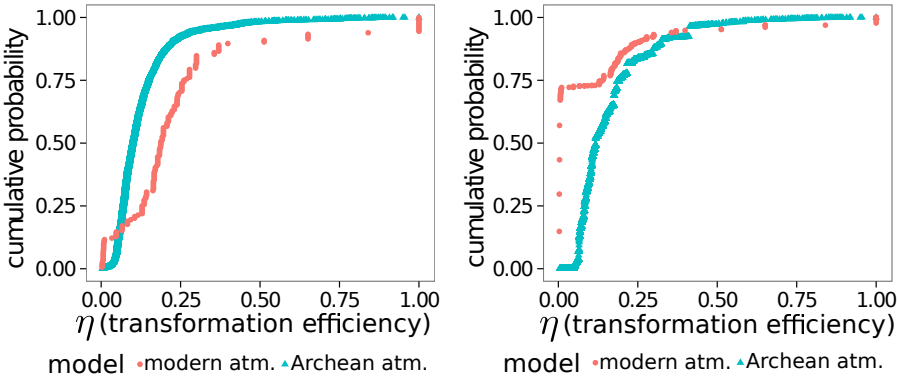



Figure 4.6: **Comparing both models by looking at their thermodynamic transformation efficiency. left:** Cumulative distribution of thermodynamic transformation efficiency. Only pathways with positive efficiency are considered. **right:** Cumulative distribution of thermodynamic efficiency weighted by rate fraction of pathways. Two pathways that transform $O_2$ in $O_3$ and $O_1$ dominate the distribution (contribute 90%of the steady state rate) and are excluded.

lating thermodynamic properties of contained reactions first. Only species that take part in photochemical reactions or matter exchange are relevant for this characterisation. On the pathway level, from 30,157 pathways of the modern atmosphere 19,983 are compatible with the chemical potential, 13 are incompatible and 10,161 can not be decided. Weighted (Eq. (4.3)), the fraction of incompatible pathways is 0.01 and the fraction of undecidable is $7.3 \times 10^{-5}$. From the 24,627 pathways of the Archean model 11,917 are compatible with the chemical potential, 3 are incompatible and 12,707 can not be decided. Weighted (Eq. (4.3)), the fraction of incompatible pathways is 0.09 and the fraction of undecidable is 0.22. This big fraction is mainly due to Pathway 2 (undecided) and Pathway 3 (incompatible) which are shown in Table 4.2. A list of all pathways incompatible with the estimated chemical potentials is given in the supplementary material (Appendix C).

### 4.2.4 *Thermodynamic Characterisation*

From those pathways that can be fully quantified and that match the chemical potentials, that is 99% (rate weighted) for the modern atmosphere and 70% for the Archean atmosphere, thermodynamic properties of individual pathways and the atmosphere as a whole[6] can be calculated. In total, the accountable part of the modern atmosphere model consumes 5.33 $W/m^2$ of radiative energy and the matter flux across its boundary has a balance of -2.6 $mW/m^2$. For the Archean atmosphere, the photochemical reactions only consume 26 $mW/m^2$ and the flux of matter has a balance of 0.29 $mW/m^2$. The comparable large amount of consumed radiative energy of the modern atmosphere is due to pathways related to ozone formation in the stratosphere. Just the three most significant pathways (Table 4.1) explain 5.16 $W/m^2$ of radiative energy consumed and all three involve

[6] *Excluding the part of non-quantifiable pathways.*

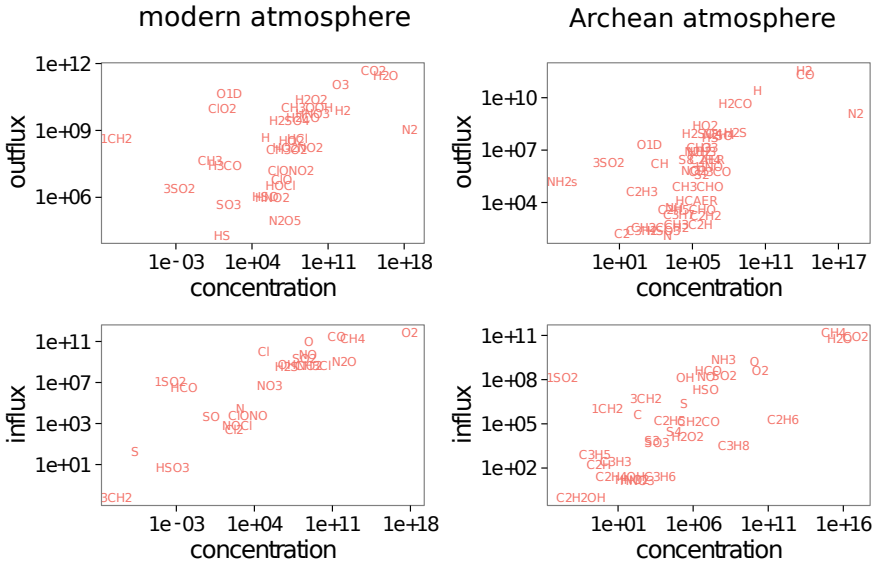



Figure 4.7: **Scatterplot showing the relationship between inflow and outflow of chemical species and their concentration** for the modern (high $O_2$) and the Archean (high $CO_2$) model. Both quantities are integrated vertically. Units are $1/cm^2$.

ozone. Its overall positive balance of chemical energy fluxes implies that the Archean atmosphere transforms light into chemical energy as a whole with an overall transformation efficiency of 1.1%. In contrast, matter passing the modern atmosphere is effectively driven by its own chemical energy that is dissipated inside the atmosphere.

This clear distinction changes if we focus our analysis only on those pathways of the third type (photochemical reactions and matter exchange) that actually have a positive balance of chemical energy through their matter exchange. Weighting all those pathways with the radiative energy that they consume we get a transformation efficiency of 79% for the modern and 15% for the Archean atmosphere. Further analysis of the transformation efficiency values reassures this by showing that the distribution tends more toward the extremes for the modern atmosphere (Fig. 4.6). A distinct feature of the model of the modern atmosphere shows is a significant number of pathways whose transformation efficiency is almost zero or almost one. The latter ones are pathways that consist of single photochemical reactions. As these reactions are substituted by exchange reactions for analysis (Fig. 4.1 (d)) there is no dissipation happening in these pathways and thus they have 100% transformation efficiency. This indicates an overall more efficient Archean atmosphere compared with the modern atmosphere. However, a few pathways of the modern atmosphere much more efficient as their Archean counterparts.

### 4.2.5 *Boundary Fluxes and Methane Pathways*

With the coexistence of molecular oxygen and methane in the modern atmosphere being the main contributions to its thermodynamic disequilibrium [HL67; SVK13] pathways depleting this disequilibrium are

of special interest. We find 7,666 pathways that consume methane and oxygen to create water and carbon dioxide. All of those use photoreactions, and in total they dissipate 1.6 $mW/m^2$ of chemical energy and 1.4 $mW/m^2$ of radiative energy. Additionally there are 9,861 pathways that also consume methane but miss at least one of the other components. They dissipate additional 0.3 $mW/m^2$ of chemical and 0.2 $mW/m^2$ of radiative energy. The value calculated in existing work [SVK13][7], normalised by Earths surface area, corresponds to 1.3 $mW/m^2$. This roughly matches our estimation of 1.6 $mW/m^2$. Getting a little higher results can be explained by complex pathways including additional species whose consideration increases the change of chemical energy.

[7] *The estimated power necessary to maintain Earth's methane-oxygen equilibrium is 0.67 TW.*

For the Archean atmosphere there is a higher methane influx and higher methane concentration (Fig. 4.7). Because the atmosphere is reducing, there are no pathways that oxidise methane with $O_2$ to $H_2O$ and $CO_2$. The most important methane pathways perform reaction with $H_2O$ to $H_2$ and CO (see Table 4.2). The chemical energy balance for the 10,404 pathways consuming $CH_4$ is an increase by 0.29 $mW/m^2$ and the radiation consumed by them is 2.0 $mW/m^2$. This is a difference to the present atmosphere where chemical energy is decreased by the pathways. In both cases methane pathways do not only make up for a significant number of the pathways but also explain a big fraction of the chemical energy balance through mass exchange (75% for modern atmosphere; 99.7% for Archean atmosphere).

To get more details on the structure of the pathways we look at the species that are part of their outflow, inflow and appear in them as intermediates. Being an intermediate means that the species is produced as well as consumed by some reaction in the pathway. If it is net produced or consumed an intermediate species can be also an inflow or outflow species. The contribution of the most significant species in these categories is visualised in Fig. 4.8. The values are normalised by the amount of $CH_4$ consumed by the specific pathway.[8] That means they correspond to the mean number of times a species appears as inflow, outflow or intermediate for one molecule of methane being consumed. Besides verifying the big impact of the significant pathways on these figures we notice an especially interesting phenomena looking at species that function as inflow or outflow as well as intermediates. These species are of interest because the coefficients of a pathway are naturally limited by the concentrations of all the intermediates. High concentrations of intermediates thus promotes its respective pathways.

[8] *Naturally, these values are also weighted by the pathway coefficients $\alpha_i$.*

The methane consuming pathways for the modern atmosphere are mainly promoted by inflow species ($O_2$, OH, O) while those for the Archean atmosphere are primarily promoted by outflow species (H, CO, $H_2CO$). This might be understood as an autocatalytic structure (pathways creating their own catalysts) that goes along with the pathways transforming radiation in chemical energy. The modern atmosphere does not show this kind of self organisation because the fluxes through its boundary are not as much driven by photochemical reac-

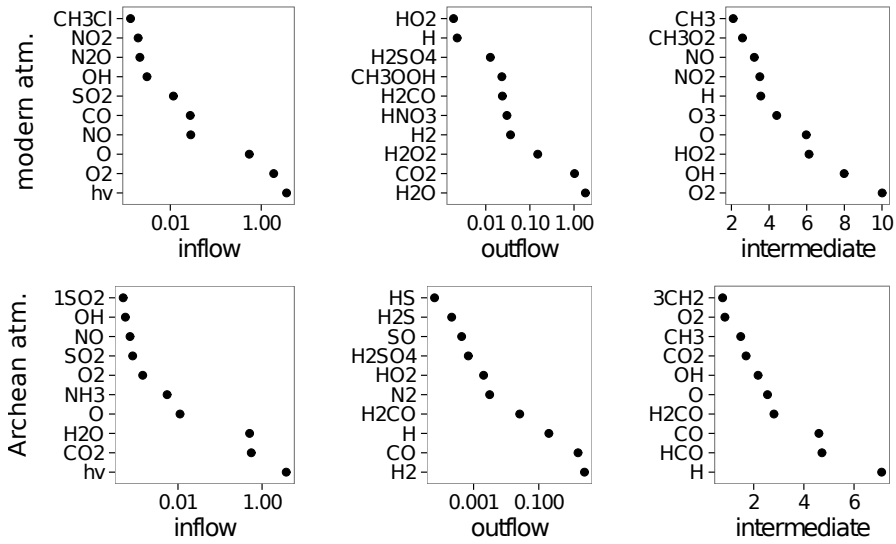



Figure 4.8: **Common species in methane consuming pathways.** Occurence of species for modern atmosphere **(top)** and Archean atmosphere **(bottom)**. Numbers are normalised by pathway coefficient $\alpha_i$ and number of methane species consumed by pathway.

tions but through mass exchange of species with different chemical potential.

## 4.3 DISCUSSION

In this chapter it has been shown that pathway analysis can be a useful computational tool for atmospheric modelling by providing additional information in comparison to a simpler view on reaction networks (Fig. 4.7). We found a tendency of the modern atmosphere to pathways with fewer cycles and fewer reactions. This decreased complexity indicates a flow of matter that is more direct and can be seen as an effect of a stronger connection of the atmosphere with other planetary processes. The decrease of complexity might seem paradoxical at first as the chemical development towards life is commonly described as an increase in chemical complexity[Pea93]. But in this context, complexity especially refers to molecular structure and spatial organisation, it would hardly be contained in our type of analysis even if it would extend from the atmosphere into biochemical details of living organisms. Generally the much greater chemical disequilibrium is between the organic carbon of the biosphere (with all of its complexity) and the oxygen in the atmosphere, and not within the atmosphere itself [Kle16]. But also, for biochemistry of metabolism the point of common and comparable simple universal pathways is made [SM16; MS07].

One major assumption that the results of pathway analysis depend upon is that the models accurately describe the respective atmosphere and its planetary system. The model is constrained by reaction rate constants that are derived from laboratory measurements. Additionally there is a constraint through proxy data and geological insight

that is used for choosing the right model structure and deposition fluxes. However, there is no explicit consideration of thermodynamic constraints in modelling itself. This is especially noteworthy as a considerable mismatch of pathway directions with thermodynamic estimates was found for the second model (high $CO_2$, Archean atmosphere). Even if this mismatch only affects 3 pathways, these are explaining 9% of the system's steady state. For obtaining thermodynamically consistent models, calculating these pathways and modifying the model to eliminate them should be straightforward.

*The list of inconsistent pathways is available in the digital appendix. See Appendix C.*

The found consistency of the quantification of methane oxidation i the modern atmosphere with existing work is encouraging regarding the conducted estimation of chemical potentials. The change of methane consuming pathways from being solely driven by radiation in the Archean atmosphere to being driven together by radiation and matter flow shows to be the important indicator for widespread life and the oxidising atmosphere. The increased transformation efficiency in few pathways transforming radiative into chemical energy (79% vs. 15%) in the modern atmosphere due to simpler pathway structure does not influence the overall picture of a more chemical energy driven modern atmosphere.[9] For remote life detection [Sag+93; Kas97; Pac01] our technique may have some potential on the long term. The requirement of an atmospheric chemistry model is challenging nowadays[10] in comparison to existing ideas of looking at specific spectra of absorption. Nevertheless, in the future an increased number of observations, better understanding of planet formation, and improvement in modelling might allow to estimate atmospheric models from sparse observations.

[9] *The high efficiency for few selected pathways might even be an artefact of the choice of boundary conditions.*

[10] *It is even more challenging if one wants characterise contribution of different sources to exchange fluxes.*

## 4.4 CONCLUSION

The approach of this chapter allows to analyse and distinguish complex reaction systems with a pathway oriented view on thermodynamics. In the model of the modern atmosphere, reaction pathways are shorter than in the Archean atmospheric model. Furthermore the Archean atmosphere has more cycles and could be regarded as more complex than the modern atmosphere. This difference can be explained by a big contribution of simple pathways (especially ozone related) in the modern atmosphere. A clear difference between the two models and its respective atmospheres can be found through thermodynamically characterising the found pathways and focusing especially on pathways that contain photochemical reactions as well as matter exchange through the boundary of the atmosphere. The energy driving these pathways mainly comes from photochemistry (radiation) for the Archean atmosphere. In the modern atmosphere these pathways are generally driven by the chemical energy of the exchanged matter and photochemistry only acts as an additional driver catalysing the pathways. Our methodology thus allows to separate between the influence of the purely abiotic photochemistry within

the atmosphere and of the biotically influenced mass exchange on maintenance of atmospheric chemical disequilibrium.

*For more detailed characterisation an annotation of exchange would be necessary.*

This pathway analysis method characterises the structure of an atmospheric disequilibrium by the set of thermodynamically most important pathways, which goes beyond simply calculating and comparing the strength of the chemical disequilibria. The application of this allowed us to distinguish the Archean from the modern atmosphere, revealing differences in pathway complexity, cycle number, and mass exchange. This suggests that for understanding the distinction between biotic and abiotic systems the associated pathways should be studied in detail. To link these insights to the possibility of life detection, this work implies that there may be characteristic sets of pathways and associated concentrations of species that could serve as fingerprints of biotic metabolisms in planetary environments.

# 5

# EMERGING PATHWAYS IN ARTIFICIAL ECOSYSTEMS

Artificial ecosystems consisting of abiotic parts and evolving organisms are simulated. The ecosystems are created from artificial chemistries with thermodynamic constraints and a stoichiometric substructure. Networks are driven by reactions that resemble photochemistry to obtain cycling patterns as they are found in real ecosystems. Pathway analysis shows that the artificial evolution leads to the formation of simpler, more pronounced pathways.
**keywords:** artificial chemistry, artificial ecosystem, ecology, evolution, emerging pathways

*Reading the previous chapter (4) helps to understand the motivation of the work presented here. Most methods from Chapter 2 are used. Thermodynamic data was generated analogous to Sect. 3.1.2.*

## INTRODUCTION

An important topic in current climate research is the quantification of global matter cycles. Especially the carbon cycle [Fal+00] and nitrogen cycle [GG08] are of interest as they connect biosphere, geosphere and atmosphere. More precisely, each of these cycles consists of many processes that cycle elements on various spatial and temporal scales. The fastest part of the carbon cycle includes atmospheric $CO_2$ which is removed by photosynthesis of plants and returned by respiration. Much slower, in comparison, is the geological carbon cycle, which includes sedimentation, subduction in the Earth's mantel, and release of $CO_2$ through volcanoes on the timescale of millions of years. It is assumed to have an important function in stabilizing Earth's temperature through feedback loops [Ber99; Ber03].

All these matter cycles control the abundance of chemicals in the atmosphere. After looking at pathways of atmospheric chemistry in Chapter 4, it seems to be the obvious next step to get reaction networks of biosphere and geosphere, connect them to our atmospheric network and see if we can find the reaction pathways that correspond to those global matter cycles. Unfortunately, these networks are not as well-modelled as those in the atmosphere.[1] This is why, for this work, we are building an artificial life model which is inspired by the Flask model [WL07b; WL07a] and the GUILD/METAMIC models [DZ99; Dow02]. These, in turn, were inspired by Daisyworld [Woo+08] and have been developed to see how biological species might gain control of parameters of their environment.

[1] *Possible approaches are discussed in the outlook (Chapter 7).*

To model the emergence of life on a molecular scale, often artificial chemistries are used [DZB01; BF92; MOI03]. The main idea of their application is to investigate the emergence of life-like structures from abiotic components. In this chapter, we want to roughly follow an

*Another interesting artificial ecosystem has been build by Rönkkö [Rön07] using particle-based simulation.*

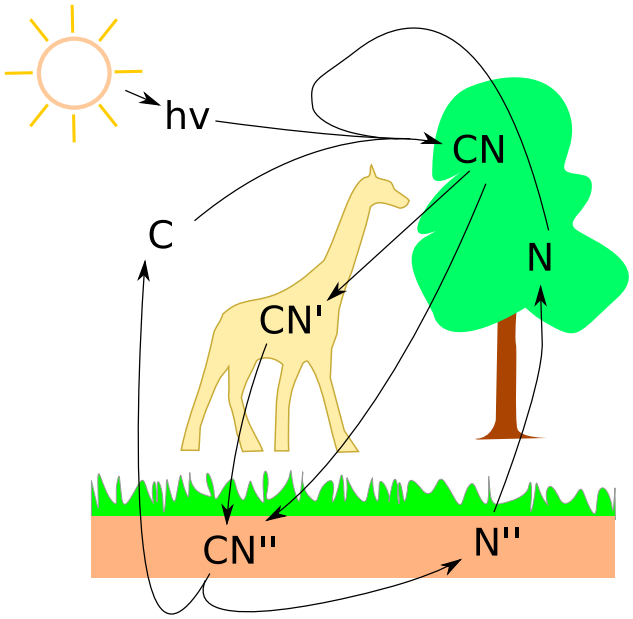



Figure 5.1: **Flow of matter and energy in ecosystems** in quasi-steady state can be described by a reaction network. The same chemical complexes in different places are represented by formally distinct species of the reaction network. *C* and *N* do not directly correspond to carbon and nitrogen here but are abstract species with their names chosen to highlight the analogy. The system is energetically driven by photosynthesis (in plants) which drives cyclic flow though other parts of the system.

idea by Dorin and Korb [DK07] and see if artificial chemistries can be used to explain properties of ecosystems rather than single organisms (see Fig. 5.1).

*A short summary of the use of the term* ecosystem *[Rei86] in artificial life is given in Dorin, Korb, and Grimm [DKG08].*

In general, the term *artificial ecosystem* can mean two things. First it refers to the specific artificial chemistry that is constructed to have properties of real ecosystems like modularity and evolution. The second meaning is that of an artificial chemistry in which ecosystem properties emerge without being explicitly designed. This chapter focuses on the exploration of artificial ecosystems in the first sense. The techniques and methods discussed here, however, should be general enough to make it possible to use them in artificial life simulations. With few exceptions [VI13; Per14; YN13], artificial chemistry models are most often built ignoring thermodynamics (energetic constraints). Here, the work done in Chapter 3 is extended to create thermodynamically consistent artificial ecosystems that can be analysed using pathway analysis methods as presented in Chapter 2 and already applied in Chapter 4.

## 5.1 METHODS

Instead of a flow system that is driven by the inflow and outflow of chemicals, we are building a closed system. The system does not exchange matter with its environment but is driven to thermodynamic disequilibrium by the inflow of energy in form of a chemical pseu-

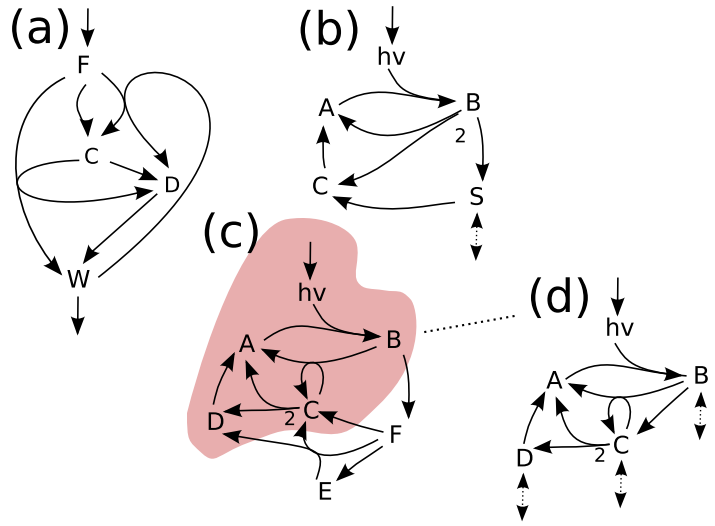



Figure 5.2: **Different ways of driving a simulated reaction network** to disequilibrium by ways of thermodynamics and supplying it with matter. **(a)** Network entirely driven by inflow of high energy food species. To be able to reach a steady state, some waste species have to be continuously removed (outflow). Either food and waste species are kept constant through contact with the environment, or a constant influx of food is induced and the outflux of waste is allowed to adapt proportional to the concentration of waste. **(b)** Network is energetically driven by photoreactions but allowed to adapt its mass through one buffered (low energy) species. **(c)** Total mass in network is kept constant (fixed through initial condition) and driven to disequilibrium by photoreactions. **(d)** Networks simulated without mass exchange can be analysed as open systems by selecting a *core* subnetwork (red) and treating the effect of the other reactions as environment.

dospecies hν which is kept constant while being used in photochemical reactions. We are using thermodynamically derived dynamics (Chapters 2 & 3) to simulate the systems and pathway analysis (Chapters 2 & 4) to analyse their steady state.

### 5.1.1 *Artificial Chemistry Model*

To drive an artificial chemistry energetically there are different ways that keep a consistent thermodynamic formalism. Common ideas are a constant inflow or a constant concentration of some high energy food species. Many common artificial chemistries (no consideration of thermodynamics) work with a fixed set of food species and outflow of all species through using a reactor with a fixed number of molecules[DZB01].

Having fixed inflow species and outflow species (Fig. 5.2 **(a)**), as used for the model in Chapter 3, comes with some drawbacks. If one wants to include substructure of species that constrains reactions, it has to be ensured that there are pathways from food species to waste species at all. The stoichiometric constraint would also undo one of the biggest advantages of the flow system, which is that its dynamics mostly does not depend on its initial condition.[2] To stress

[2] *One would have to at least guarantee pathways to generate all species from food species exist.*

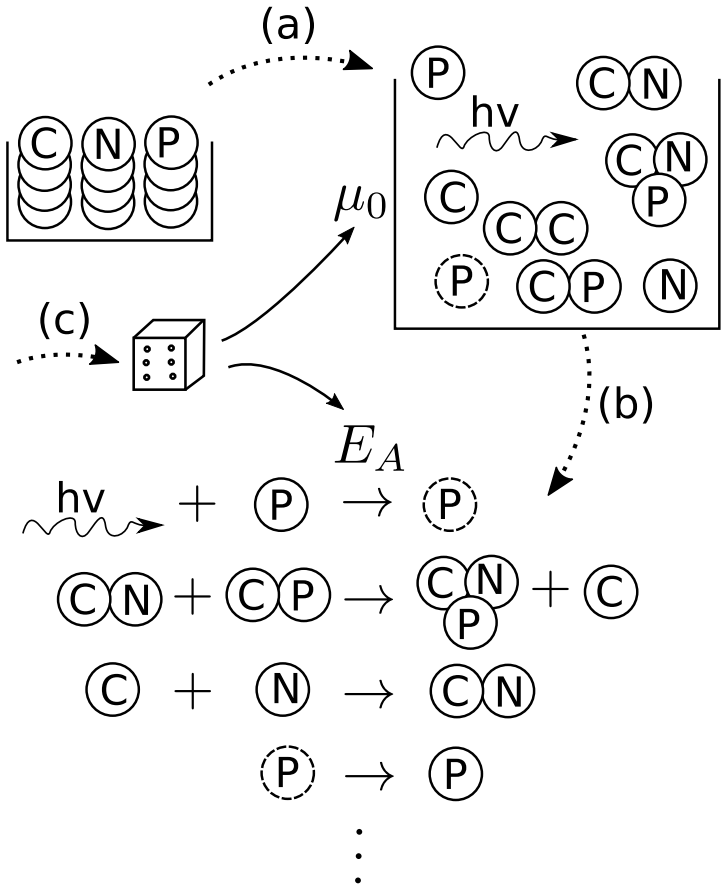



Figure 5.3: **Sketch showing the generation of the artificial chemistries. (a)** Elementary composition of the chemical species is chosen. **(b)** Fixed number of reactions (of different types) are chosen. The set of possible reactions is constrained by chemical species elementary composition. **(c)** Chemical potentials of species and activation energies of reactions are drawn from probability distributions.

the analogy to planetary ecosystems we use a model which is driven energetically by photoreactions of the form $A + h\nu \rightarrow B$ and does not exchange matter with its environment (Fig. 5.2 **(c)**). By setting the formation enthalp $\mu_0$ of $h\nu$ to a high value of 50 and controlling the concentration of $h\nu$, the strength of the force driving photochemistry can be controlled implicitly.

The artificial chemistry is generated in multiple steps that are visualised in Fig. 5.3. First, the stoichiometric mass composition of individual species is generated. Then reactions are selected from the set of all reactions compatible with the stoichiometric mass composition of the species. In the last step, activation energies $E_A$ and formation enthalpies $\mu_0$ are sampled with the method described previously (Sect. 3.1.2).

*Generating Species Stoichiometry*

Before random reactions can be generated we have to create the elementary structure of the chemical species. The model includes a minimal substructure of the species, there is no spatial conformation

*If one wants an even richer substructure a reaction network could be created through rule based expansion. [BFS03b; BFS03a; Fac+05]*

or even spatial symmetry that constrains the reactions, but just the stoichiometric composition. Chemical species are made up of $e_c$ elementary components. We use an upper bound for the stoichiometric coefficient of an elementary component $\hat{s}$. This means the elementary composition of each species is chosen from the set $\{0 \dots \hat{s}\}^{e_c}$. For all simulations presented here, we use a maximal stoichiometric coefficient $\hat{s} = 10$.

We use the poison distribution

$$P(x) = \frac{\lambda^x e^{-\lambda}}{x!} \tag{5.1}$$

*For generating the networks presented here, a value of $\lambda = 0.1$ is used.*

to define a distribution $P_c$ over all possible elementary compositions.

$$P_c : [0, \hat{s}]^{e_c} \to [0, 1], \qquad (s_1 \cdots s_{e_c}) \mapsto P(s_1) \cdot \ldots \cdot P(s_{e_c}) \tag{5.2}$$

This distribution is then used to sample the compositions of the N species while making sure the number of duplicates (same composition of different species) is limited). Each composition value can have one duplicate[3] and in total there can not be more than $M/3$ duplicates.[4] The probability of values whose sampling would violate the duplicate condition are set to zero.

[3] *That means one composition can occur in at most two species.*

[4] *For $e_c = 1$ duplicates are allowed without limitations.*

### *Sampling Reactions*

For selecting a set of M reactions between the N species (Fig. 5.3 **(b)**), we explicitly construct the probability distribution of all reactions of the form $A + B \to C + D$ in memory. Values in this form can represent one of the N species, the $h\nu$-pseudospecies or be empty. This leads to a distribution of size $(N + 2)^4$. In this distribution probabilities of reactions breaking mass conservation of elementary components are disabled. Additionally, duplicates by symmetry (e.g. $B + A \to C + D$, $C + D \to A + B$, ...) are disabled.

*Having a distribution over all $\sim N^4$ reactions in memory has the advantage that one can easily disable subsets of reactions, e.g. $A + X \to B + X$, for all X.*

After separating the reactions in linear[5] reactions ($A \to C$ or $A + E \to C + E$ type), photoreactions (contains $h\nu$) and nonlinear reactions (rest) we can sample a given number of reactions for each type. Inside the different types duplicates are prohibited. Each time after a reaction has been sampled, the probability of the duplicates of the same type is set to zero. Duplicates are defined in terms of effective concentration change. E.g. $A \to C$ and $A + E \to C + E$ are duplicates of each other. Only one of those reactions can be included in a reaction network. The reaction $A + h\nu \to C$ can be combined with one of the reactions mentioned before, as it is of a different type.

[5] *Linear reactions are identified with respect to the stoichiometric matrix.*

### *Algorithmic Complexity*

The networks generated in this chapter are all relatively small, so we can simulate a large number and analyse their reaction pathways rapidly. Nevertheless, an estimate for the complexity of network generation should be given. The sampling of the elementary composition as well as the sampling of the reactions is done by explicitly writing

the distribution to memory and sampling from it. After each new sample the distribution is updated by setting probabilities to zero to avoid duplicates if necessary. In exchange for having to update the distribution we get a deterministic run time.

For species composition the number of possible values (distribution size) is $n \sim \hat{s}^{e_c}$. With a primitive algorithm, sampling a value and updating the distribution both are $\mathcal{O}(n)$ in time. For sampling the composition of all $N$ species this gives a time complexity of $\mathcal{O}(N \cdot n) = \mathcal{O}(N \cdot \hat{s}^{e_c})$.

*For $\hat{s}$ and $N$ there is an offset of one or two that we ignore in the estimate here.*

Reactions are sampled in the same way, except that the size of the distribution is $m \sim N^4$.[6] Sampling $M$ reactions in this case has $\mathcal{O}(M \cdot m) = \mathcal{O}(M \cdot N^4)$ time complexity. Put together, for sampling the entire artificial chemistry the time complexity is

[6] *Each reaction has a maximum of four participating species.*

$$T_{RN} = \mathcal{O}(N \cdot \hat{s}^{e_c} + M \cdot N^4). \tag{5.3}$$

For future use, the algorithm can be improved by using a binary tree like structure for sampling from the discrete probability distributions. In this tree, each node contains the accumulative probability of its subtree. The data structure takes linear time to create, but samples can be drawn and probabilities can be updated in logarithmic time. For network generation time this means[7]

$$\begin{aligned} T_{RN} &= \mathcal{O}(\hat{s}^{e_c} + N \log(\hat{s}^{e_c}) + N^4 + MN \log(N^4)) \\ &= \mathcal{O}(\hat{s}^{e_c} + N \cdot e_c \log \hat{s} + N^4 + 4 \cdot MN \log N). \end{aligned} \tag{5.4}$$

[7] *The $N$ in the fourth term originates from the $N$ equivalent reactions for which the distribution might have to be updated.*

For reasonable small values of $\hat{s}$ and $e_c$ and a number of reactions $M$ comparable to the number of species $N$, the dominating term is the $N^4$ term for generating the distribution of all possible reactions for the calculated set of species. As the possible reactions are marked in an array of that size, this also describes the memory consumption of the program.[8] If we assume each (theoretically possible) reaction takes 8 byte, generating a network with 150 species would need 8.1GB of memory.

[8] *Just saving the valid reaction does not allow for time efficient deletion of duplicates of linear reactions.*

### *Simulating Chemistry*

The reaction equations of the networks are defined in the same manner as in Chapter 3. Thermodynamic parameters are generated as described there (Sect. 3.1.2) and reactions are simulated reversibly with mass action kinetics (Sect. 2.2.1). Because the systems here do not exchange matter with the environment a solver with a low error is important. Thus we are using an implicit solver for integrating the differential equation. The necessary equations are derived in Appendix A.1, the source code is contained in the package *jrnf_int* (see Appendix C).

*An* implicit solver *calculates the next time step by solving an equation system involving partial derivatives (Jacobi-matrix).*

To simulate a chemistry with different initial conditions, but still having the same mass distribution of elementary components in the system, we use the following scheme. All chemical species (except $h\nu$) are initialised with a given initial concentration $c$. Then, before starting the integration of the ODE, all reactions are applied in random

*The concentration of $h\nu$ is not affected by this randomisation.*

order and direction. To ensure that no species has a concentration of zero when the integration starts, each reaction is only applied until one of its educts has a concentration of $c/100$.

### 5.1.2 *Evolvable Artificial Ecosystems*

The artificial ecosystems we simulate are constructed from one anorganic part and multiple organisms (Fig. 5.4). All parts are constructed using the artificial chemistry model presented previously. There is no a priori difference between the chemistry of organisms and the anorganic part. The anorganic part of the network contains the $h\nu$-pseudospecies that supplies the system with energy. Organisms are created by first choosing a given number of interaction species from the anorganic chemistry and then creating the stoichiometry of a number of internal species that are only visible to the organism. Afterwards the "internal" reactions of the organism[9] are sampled as described in Sect. 5.1.1. For species composition and reactions inside the anorganic network and inside each organism the same constraints regarding duplicates apply as described before. When selecting interacting species for an organism $h\nu$ is treated like any chemical species. If it is not chosen as interacting species, no photoreactions are contained in the respective organism.

[9] *This are reactions that can contain interaction species and internal species.*

Once assembled, an anorganic chemistry and connected organisms form an artificial chemistry (Sect. 5.1.1) and its reaction equations can be simulated to steady state in the same way. We want to simulate the same anorganic chemistry with different organisms and have a comparable mass in the system. To achieve that, we set the initial concentration of organic species (before randomizing) to be one thousandth of that of anorganic species. That means the mass of the system is determined by the stoichiometry of the elementary components of the anorganic species, which do not change when changing organisms.

To evolve the network, an additional layer of dynamics is added on top of the chemical differential equation.[10] In each step the system is simulated for a fixed amount of time that should be enough to reach steady state for most cases of networks and initial conditions. Then for each organism the fitness $L_i$ is determined by the product of its fraction of the total mass $w_i$ and its fraction of the absolute steady state rate $r_i$.[11] Formally those are defined by

$$w_i := \frac{\sum_{S_j \in O_i} m_j}{\sum_j m_j} \qquad r_i := \frac{\sum_{R_j \in O_i} |\nu_j|}{\sum_j |\nu_j|} \qquad L_i := w_i \cdot r_i, \tag{5.5}$$

with $S_j \in O_i$ indicating that species $j$ is part of organism $j$ and $R_j \in O_i$ meaning the same for reaction $j$. If multiple evaluation runs (ODE integrations) are performed for the same network, each species fitness is determined by the maximum of the fitness values of the individual runs. After the organism with the lowest fitness is determined it is removed from the network and a new organism is sampled and added to the ecosystem.

[10] *This technique resembles a simple version of evolutionary optimisation [BS02].*

[11] $r_i$ *has a similar meaning than* rate fraction $f_i$ *defined in Eq. 4.3. But it refers to an organism instead of a pathway.*

Table 5.1: **Parameters for generation (and evolution) of artificial chemistries (*A, B, C*) and artificial ecosystems (*D,E*).** Parameter values are given in the order they are needed for simulation, starting with the number of elementary components and cutoff for activation energy $\widehat{A_E}$. "*" X+1 indicates X species plus $h\nu$-pseudospecies. "†" Linear reactions are identified in respect to the stoichiometric matrix (includes $A + E \rightarrow C + E$ type).

| | Network / Ecosystem | A | B | C | D | E |
|---|---|---|---|---|---|---|
| | el. components $e_c$ | 1 | 2 | 2 | 1 | 2 |
| | $\widehat{A_E}$ | 5 | 5 | 5 | 2 | 3 |
| anorganic network | species* | 4+1 | 4+1 | 12+1 | 12+1 | 18+1 |
| | total reactions | 6 | 6 | 18 | 25 | 25 |
| | linear reactions† | 2 | 2 | 5 | 5 | 10 |
| | photoreactions | 2 | 2 | 3 | 5 | 5 |
| organisms | interacting sp. | | | | 4 | 4 |
| | internal species | | | | 3 | 4 |
| | total reactions | | | | 8 | 10 |
| | linear reactions† | | | | 2 | 2 |
| | photoreactions | | | | 2 | 2 |
| | organism number | | | | 4 | 4 |
| | evaluation runs | | | | 5 | 5 |
| | generations | | | | 500 | 500 |
| ODE solver | initial concentration | 0.1; 1; 10 | 0.1; 1; 10 | 0.01...100 | 10 | 10 |
| | simulated time units | $10^7$ | $10^7$ | $10^7$ | $10^9$ | $10^9$ |
| | $h\nu$ concentration | $10^{-4} \ldots 5 \cdot 10^9$ | $10^{-2} \ldots 5 \cdot 10^9$ | $10^{-3} \ldots 5 \cdot 10^4$ | 10 | 10 |

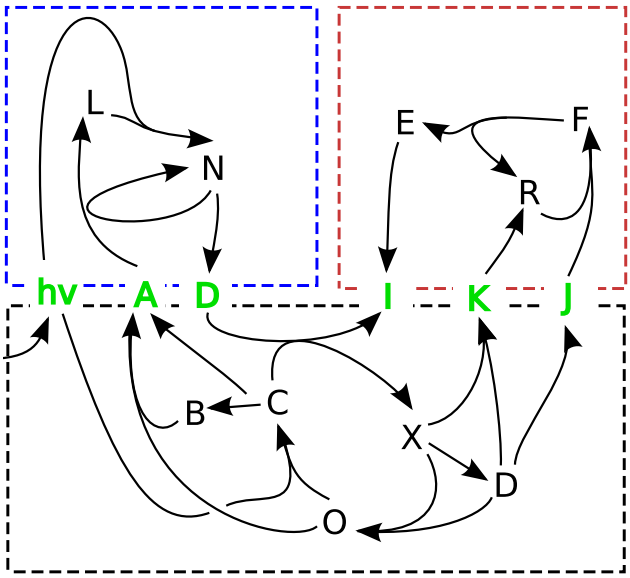



Figure 5.4: **Evolvable reaction network made of organisms *(top / colour)* and their abiotic environment *(bottom).*** Organisms are added by first selecting interacting species *(green)* and then adding internal species that are only visible inside the organism. Then random reactions can be added between internal species and interacting species. By successively replacing organisms with low fitness the network structure is evolved.

### 5.1.3 *Reaction Pathway Analysis*

Steady state rate vectors $\mathbf{v}$ of the simulated reaction networks are analysed by decomposing them into a linear combination of reaction pathways $\mathbf{E}^{(i)}$ with positive coefficients $\alpha_i$ as described in Sect. 2.2.3, *Cf. Eq. 2.14*

$$\mathbf{v} = \sum_i \alpha_i \mathbf{E}^{(i)}, \tag{5.6}$$

using the algorithm for pathway decomposition (Sect. 2.3).

For weighting pathways we use the *rate fraction* as introduced in Sect. 4.2.2: *Cf. Eq. 4.3.*

$$f_i = \frac{\alpha_i \sum_j |E_j^{(i)}|}{\sum_i \alpha_i \sum_j |E_j^{(i)}|}. \tag{5.7}$$

It describes which fraction of the steady state is explained by a certain pathway in this specific expansion. This quantity has the advantage of being normalized (if sufficiently many pathways have been calculated). Thus, $f_i$ can be interpreted as a probability and be used to weight quantities associated with individual pathways, like the number of cycles or the number of reactions.

Besides for analysis of complete networks, we are also using the pathway analysis for subnetworks, specifically those of the static (anorganic) part of our artificial ecosystems (Fig. 5.4). These subnetworks are treated like any other network, except that in- and outflow reactions are added to represent the effect of the outside part of the original network (see Fig. 5.2 **(d)**).

### 5.1.4 *Simulations*

To present how reaction pathways and artificial chemistries work together, some simple artificial chemistries are simulated. Afterwards ecosystems constructed from more complex networks are evolved. A summary of the different simulations and their parameters is given in Table 5.1.

*Sample Networks (A, B, C)*

Two simple network with 4+1 chemical species are created and simulated. This means 4 conventional species and the "+1" is indicating the $h\nu$-pseudospecies for controlling photochemical reactions. Both networks have 6 chemical reactions. Network A (Fig. 5.5 **(a)**) is created with one elementary component, Network B (Fig. 5.6 **(a)**) with two elementary components. Both are simulated for $10^7$ time steps for various combinations of initial concentration and energy input. Network C (Fig. 5.7) is created with two elementary components, from which 12+1 chemical species and 18 reactions are sampled. It is simulated with similar parameters as the other two networks (Table 5.1).

*Evolving Ecosystems (D, E)*

The size of the ecosystems is limited by the performance of the pathway analysis that is performed after network evolution. For the ecosystem with one elementary component (Ecosystem D) we create anorganic core networks with 12+1 species and 25 reactions. The chosen parameters for creating the organisms lead to the final network (anorganic core and four organisms) having 24+1 species and 49 reactions.

The ecosystem with two elementary components (Ecosystem E) seems to be handled better by pathway analysis, which means that we can generate larger networks. The anorganic core is made of 18+1 species and 25 reactions. With 4 organisms the network size is 34+1 species and 65 reactions. For the choice of all parameters (Table 5.1), the aim was to allow for some complexity inside of each organism as well as some interaction between organisms.

## 5.2 RESULTS

To show the basic properties of model and analysis, the simple reaction networks are simulated to steady state ($1 \cdot 10^7$ time units) and the final rate vector is decomposed into reaction pathways.

### 5.2.1 *Pathway Structure*

Thermodynamically closed systems *can exchange heat and work with their environment, but do not exchange matter.*

Cycling in the thermodynamically closed system is driven by photochemical reactions, which in turn are controlled by the concentration of the chemical species $h\nu$ in the network. We vary concentration of $h\nu$ to see how photochemical activity and reaction pathways change

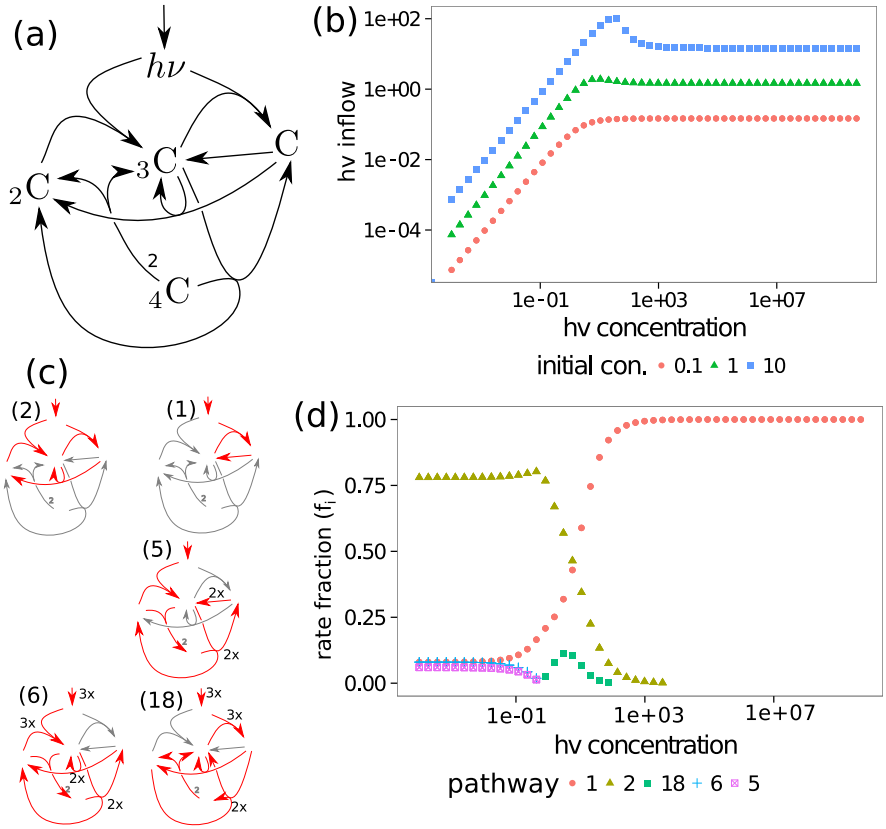



Figure 5.5: **Simulation of Reaction Network A** (1 elementary component; 4+1 species; equations shown in Table 5.3). **(a)** Plot of network structure. **(b)** Relationship between inflow of hν and its concentration for different initial concentrations (mass of the system). **(c)** The five reaction pathways of the reaction network for the states (reaction directions) occurring in the simulations. **(d)** Change of pathway decomposition with driving force (hν) for simulations with initial concentration of 1 for all species.

with thermodynamic forcing.[12] This reveals two distinct regimes, a power law with an exponent between one and two and a saturated phase with a constant photochemical activity above a certain concentration of hν (Fig. 5.5 **(b)** & Fig. 5.6 **(b)**). The plateau is an effect of the constant mass of the system. All pathways are constrained through their non-photochemical reactions, which are proportional to concentrations of their reactants.[13]

[12] Photochemical activity *means consumption of* hν *by photoreactions.*

[13] *This works as long as no pathways with only photochemical reactions exist.*

Interestingly, between the two regimes also a qualitative change in pathway decomposition occurs. The tendency of the sample systems (Fig. 5.5 & 5.6) is to have fewer, more important pathways in the plateau regime. These pathways also tend to have fewer reactions. For Network A we can see how a reaction changes in the transition between pathways (5),(6) and (18) (cf. Fig. 5.5). The coexistence of pathways (1) and (3) in the higher regime of Network B (Fig. 5.6) can be explained by them sharing species $N$, which occurs as a product of photochemical reactions in both pathways.

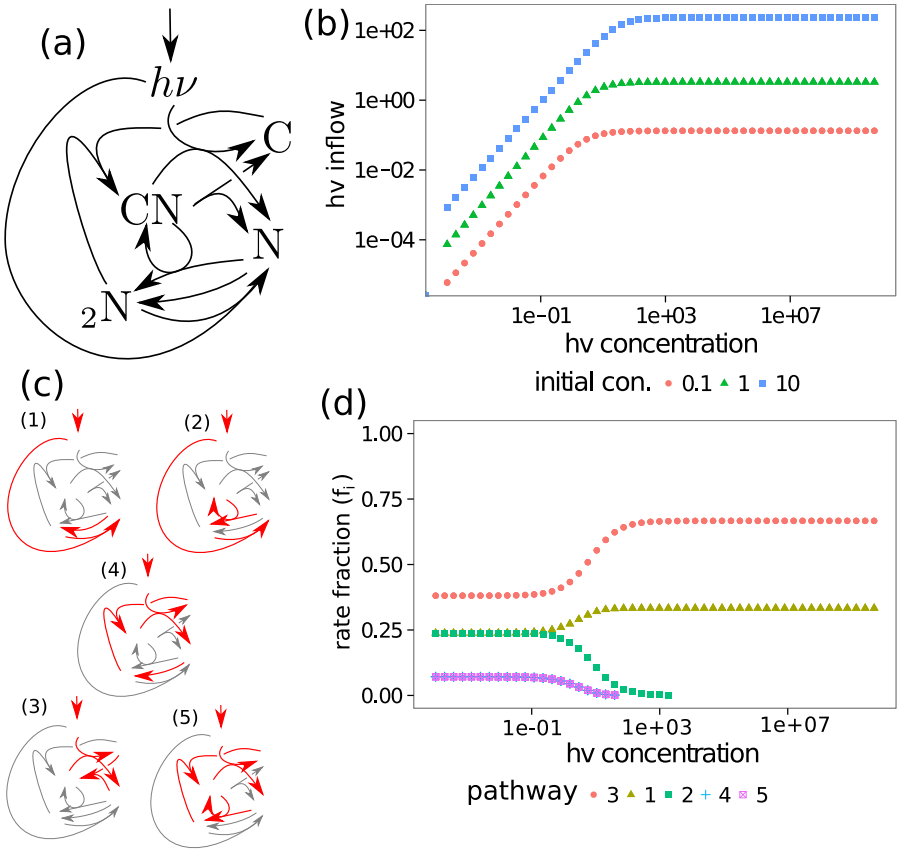



Figure 5.6: **Simulation of Reaction Network B** (2 elementary components; 4+1 species; equations shown in Table 5.4). **(a)** Plot of network structure. **(b)** Relationship between inflow of hν and its concentration for different initial concentrations (mass of the system). **(c)** The five reaction pathways of the reaction network for the states (reaction directions) occurring in the simulations. **(d)** Change of pathway decomposition with driving force (hν) for simulations with initial concentration of 1 for all species.

### 5.2.2 *Network Self Organisation*

The larger sample chemistry (Network C) is shown in Fig. 5.7 and its simulation results are depicted in Fig. 5.8. We notice the same transition between different regimes and change in pathway decomposition between them like in the smaller sample chemistries. The pathway distribution also seems more skewed for the upper regime, where the most important pathway have a larger contribution to the steady state than their lower flow counterparts.

To see how mass inside the network redistributes, we are looking at the part of the network that accumulates mass. To determine this, the chemical potential $\mu$[14] in non-equilibrium steady state is compared with its value in a reference simulation with disabled photoreactions. We call those species that have an increased chemical potential *core network*. Looking at the size of this core network for different concentrations of hν shows how the mass of the system is concentrating in a smaller section of the network if hν, and thus the thermodynamic disequilibrium, is increased (Fig. 5.8 **(b)**).

[14] $\mu = \mu_0 + \log c$

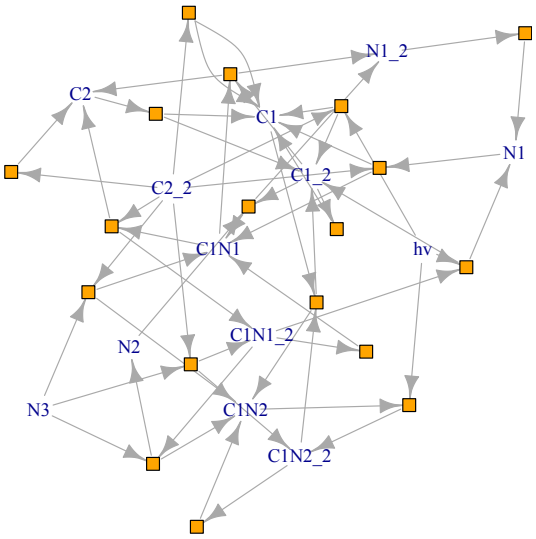



Figure 5.7: **Plot of Reaction Network C.** Larger sample network with 2 elementary components and 12+1 species. Reaction equations are given in Table 5.6.

### 5.2.3 *Pathway Evolution*

When evolving artificial ecosystems, not only pathways or pathway coefficients change, but in each generation the network is partially rewritten. Throughout each evolution run, the steady state rates of the anorganic network and of the entire ecosystem are decomposed into reaction pathways.[15] Using the rate fraction of the pathway (Eq. 5.7), we are calculating the weighted number of cycles (of length shorter 5) as an indicator for the change of pathway complexity (Fig. 5.9 **(e)** & **(f)**). From the initial, random network of the ecosystems, the weighted number of cycles decreases by around 50% in roughly 50 generations for both ecosystems.

*Ecosystem generation and evolution is introduced in Sect. 5.1.2.*

*[15] Because of its computational complexity, pathway analysis on the full ecosystem is only performed for 10 out of the 500 generations.*

To look at this change in greater detail, the most important pathways for explaining their respective steady state (Eq. 5.7) are visualized for selected generations in Fig. 5.10 and Fig. 5.11. The structure of the pathways in the evolved ecosystem are much simpler than the most important pathways in the initial (random generated) ecosystem. Also the contribution of the most important pathway increases from 30% to 65% for Ecosystem D and from 10% to 50% for Ecosystem E.

To better characterise this tendency towards fewer, more important pathways, we interpret the values for the rate fraction $f_i$ (Eq. 5.7) as probability distribution from which the informational entropy H can be calculated:[16]

$$H = -\sum_i f_i \log f_i. \tag{5.8}$$

*[16] Throughout this work* log *refers to the natural logarithm. The choice of base only changes a constant factor of* H.

*Lower values of* H *imply a more uneven distribution.*

Plotting the changing informational entropy for the anorganic part of the artificial ecosystem confirms the hypothesis of a simpler pathway distribution (Fig. 5.9 **(c)** & **(d)**). Interestingly, most of the effect of evolution on pathways seems to happen in the first 50 generations.

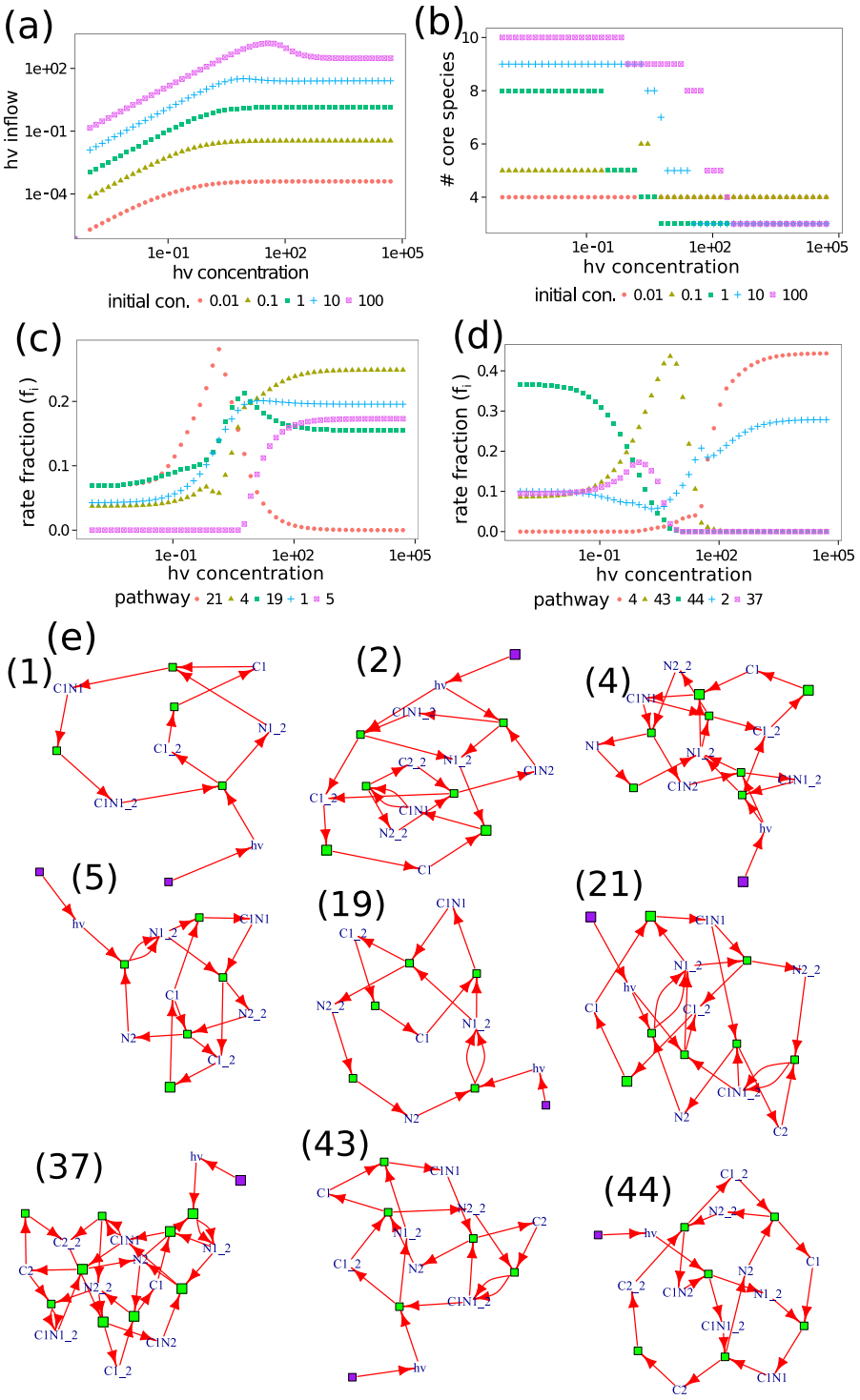



Figure 5.8: **Results for Reaction Network C.** (See Fig. 5.7 and Table 5.6 for network structure.) **(a)** Dependency of photochemical activity ($h\nu$ inflow) from driving force for different initial concentrations. **(b)** Number of core species as function of driving force ($h\nu$ concentration). Core species are defined as species that have a higher chemical energy $\mu$ than in thermodynamic equilibrium (same initial concentration; no photoreactions). **(c)** Decomposition of steady state into pathways for an initial concentration of 1. **(d)** Decomposition of steady state into pathways for species initial concentration equal to 10. Plots show the rate fraction for the five most important pathways as a function of concentration of pseudospecies $h\nu$. **(e)** Pathways that are occurring in decompositions shown in Panels (c) and (d). (Pathways are also described in Table 5.5.)

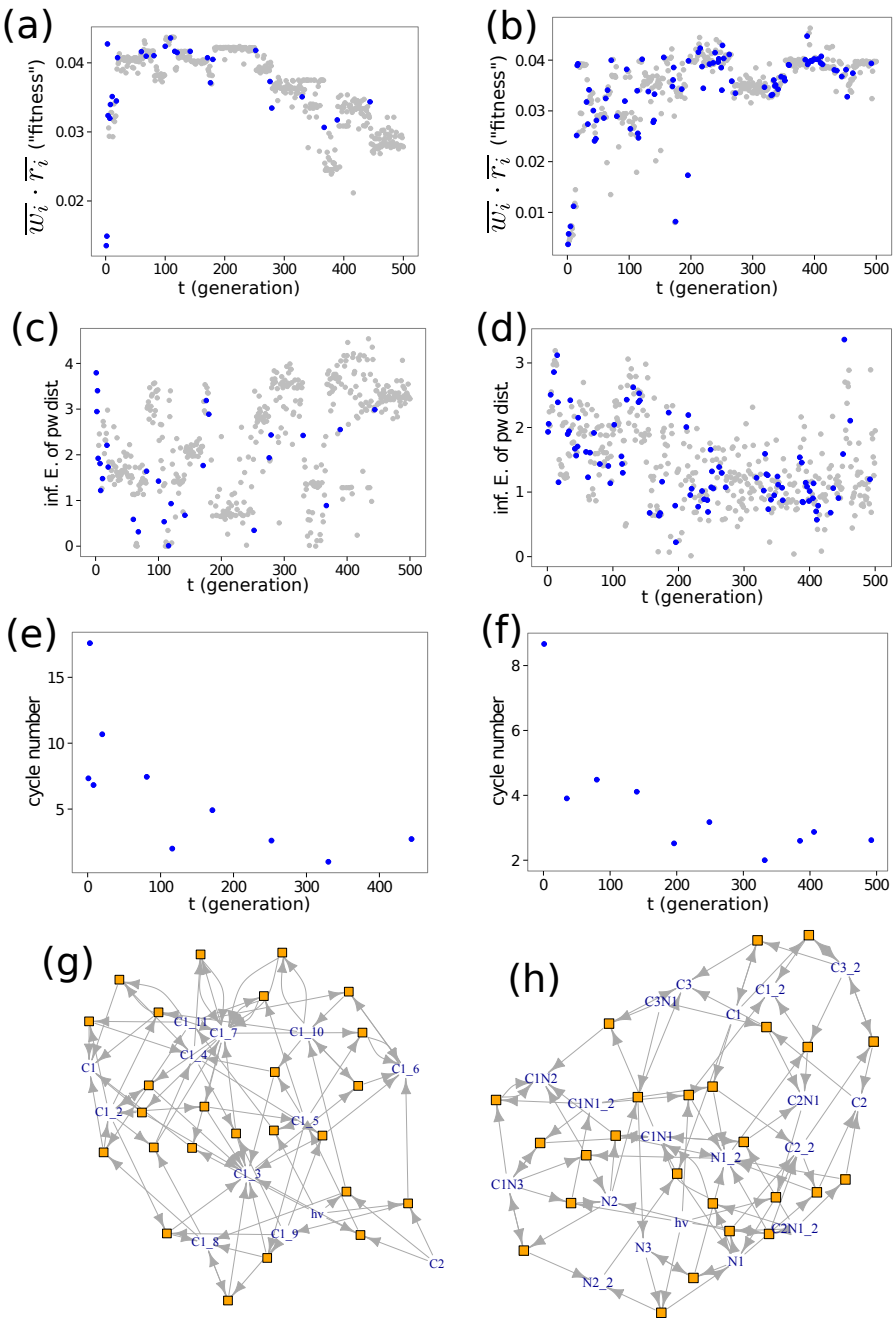



Figure 5.9: **Evolution of Ecosystem D *(left)* and E *(right)*.** Bright blue data points show generations in which a newly generated organism survived longer than one generation (indicates innovation). **(a)**,**(b)** Evolution of the ecosystems ($w_i \cdot r_i$ is used as fitness function). **(c)**,**(d)** Informational entropy (Eq. 5.8) of pathway decomposition of the (anorganic) core network. **(e)**,**(f)** Cycle number of entire ecosystem (obtained through pathway decomposition). **(g)**,**(h)** Anorganic parts of the evolved ecosystems.

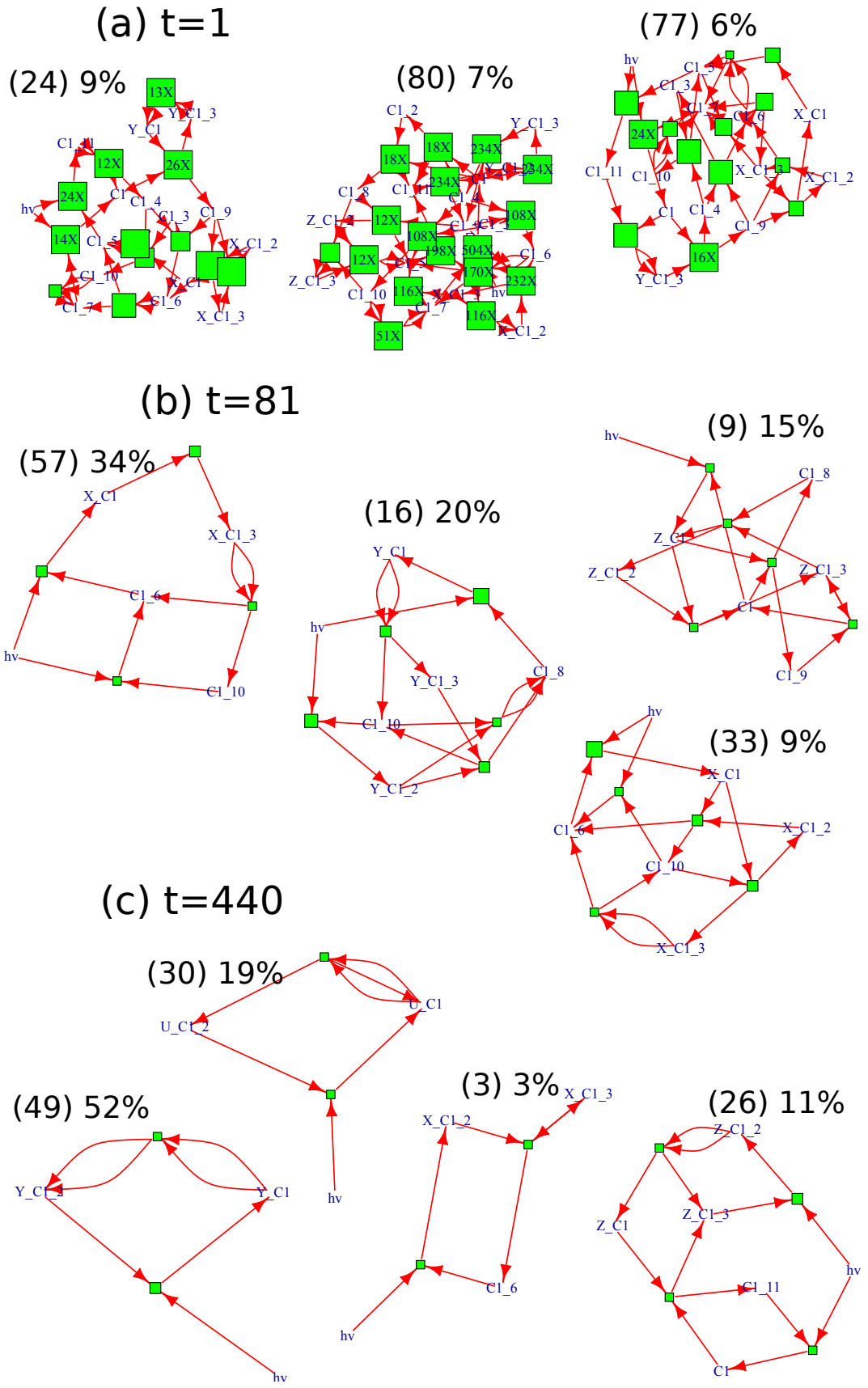



Figure 5.10: **Pathways of Ecosystem D (1 elementary component).** Most relevant pathways for explaining the steady state of the ecosystem and their rate fractions (Eq. 5.7) in percent. **(a)** Before evolution, **(b)** after 80 steps, and **(c)** after 439 steps of evolution. (Size of squares symbolizes coefficient of respective reaction. Leading letters *X*, *Y*, *Z*, *U* separated by underscore indicate internal species of organism. Final number separated by underscore indicates isotopes.)

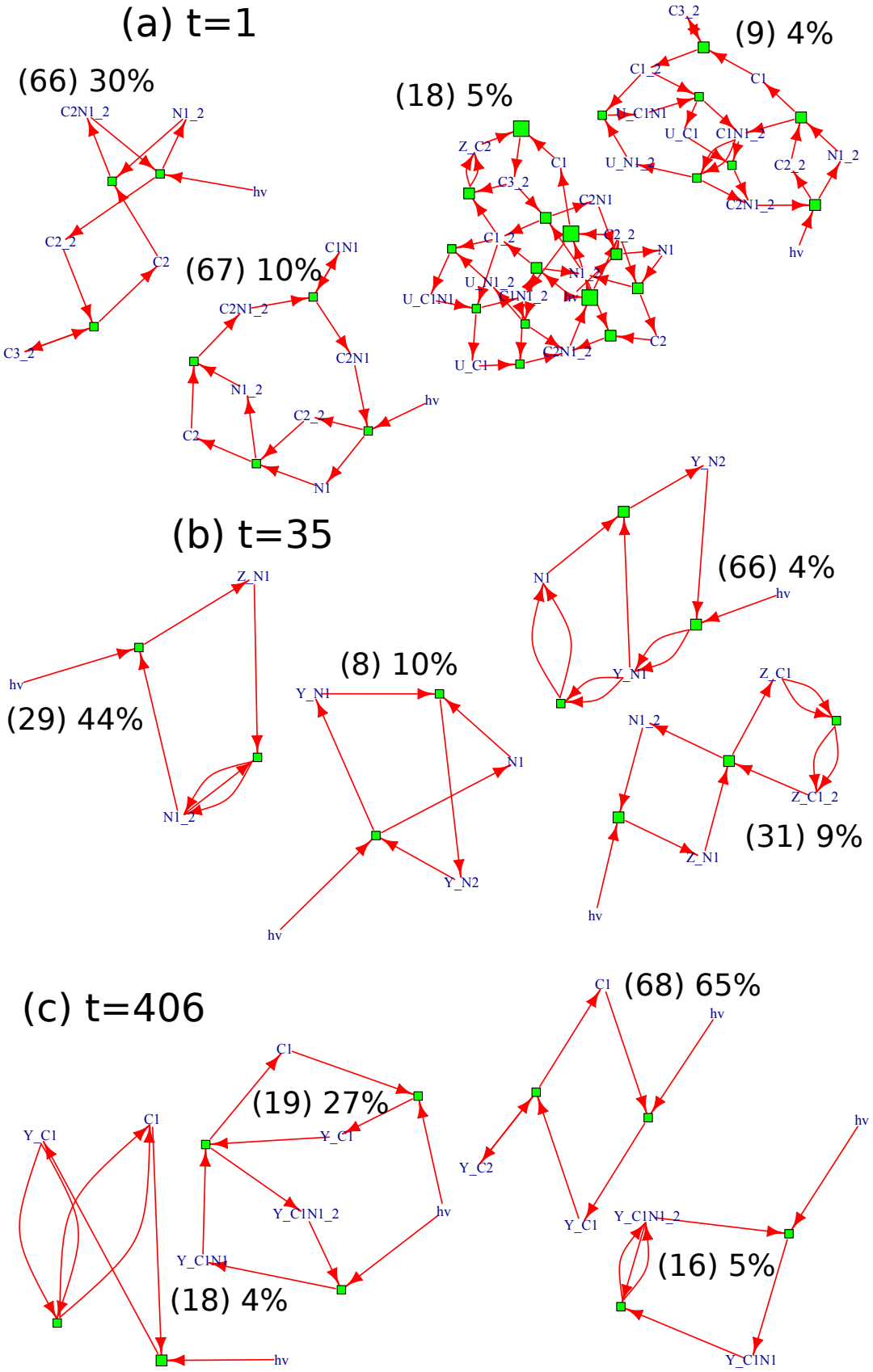



Figure 5.11: **Pathways of Ecosystem E (2 elementary component).** Most relevant pathways for explaining the steady state of the ecosystem and their rate fractions (Eq. 5.7) in percent. **(a)** Before evolution, **(b)** after 34 steps, and **(c)** after 405 steps of evolution. (Size of squares symbolizes coefficient of respective reaction. Leading letters $X$, $Y$, $Z$, $U$ separated by underscore indicate internal species of organism. Final number separated by underscore indicates isotopes.)

After that time there are no pathways that are spanning multiple organisms, indicating that evolution did not lead to formation of a form of cooperation, but rather independent organisms.

### 5.2.4 *Statistics of Evolution Runs*

The evolution of both of the two different ecosystems shows that evolved ecosystems have a simpler pathway structure. To get an estimate on the statistical significance of this, we repeat network construction and evolution 20 times for each of the two network types. Reaction pathways are evaluated like described in Sect. 5.2.3 and the change of all parameters between the initial, random generated, ecosystem and all states of the ecosystems after generation 50 of the evolution is calculated. The statistics shows a significant decrease of all of the parameters between initial and evolved ecosystems (Table 5.2) showing that the effect of evolution towards simpler pathways that was found for two specific ecosystems (Sect. 5.2.3) is statistically significant.

We find the efficacy of the evolution being limited to the beginning of the simulation (Fig. 5.9). In later stages of the evolution, mutation of ecosystems leads to large fluctuations of fitness. To better understand the nature of this fluctuations, simulations with a different number of organisms and with a different number of evolution runs are concluded. If changing the number of evolution runs per network influences the fitness this would indicate that the fluctuations are (partially) due too random errors in determining the fitness. Some initial concentrations would lead to the ODE-simulations not converging fast enough and an incorrect fitness of the organisms being determined. Change of fluctuations with number of organisms in an ecosystem would indicate the process of evolution by mutation through simple replacement of organisms as the origin of the fluctuations.[17]

[17] *Replacing the least fit organism of a bigger number should naturally have less of an impact.*

Five anorganic core networks are created with the parameters of Ecosystem D and the same number with the parameters of Ecosystem E . To compare the effect of organism number and fitness evaluation, we use the same anorganic networks with different parameters for number of organisms $o$ and number of evaluation runs $r$. The number of evaluation runs per generation $r$ does not have a large influence on the evolution process (Fig. 5.12 & 5.13). In contrast to this, increasing the number of organisms from 4 to 8 leads to reduced fluctuations in most cases. In the case of network type D (one elementary component) there are still jumps in the fitness, but those seem more like discrete jumps at few time points instead of the fluctuations visible in the reference simpulations (Fig. 5.12).

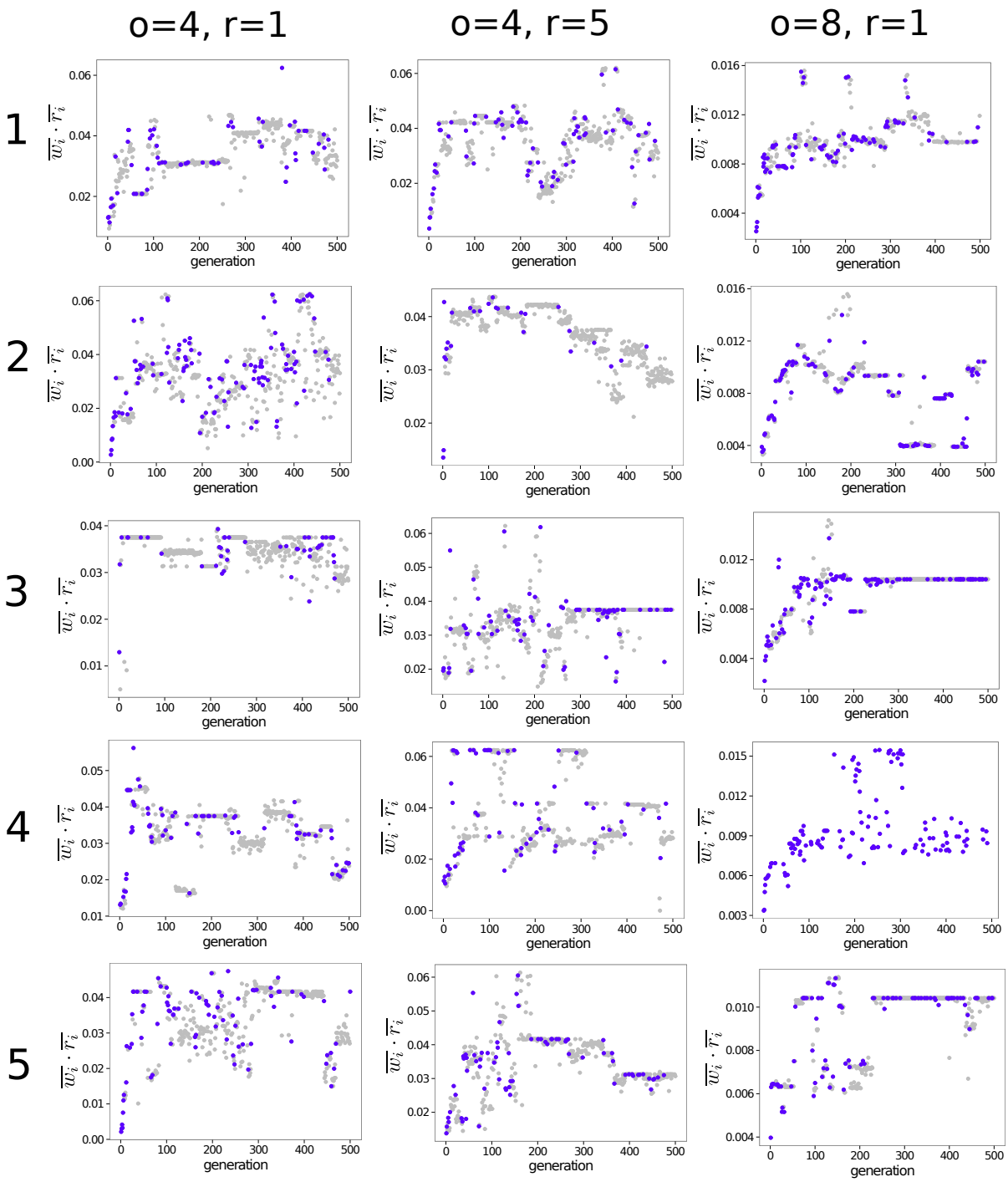



Figure 5.12: **Multiple evolution runs of artificial ecosystem (Network D-type).** Figure shows evolution of quantity mirroring fitness of the organisms for multiple runs *(vertical)* and for different parameters (horizontal) for number of simulation runs for evaluation r and number of organisms per ecosystem o. Each row of simulation runs was started with the same anorganic network but initialized with different organisms. (See Table 5.1 under *Network D* for additional parameters.)

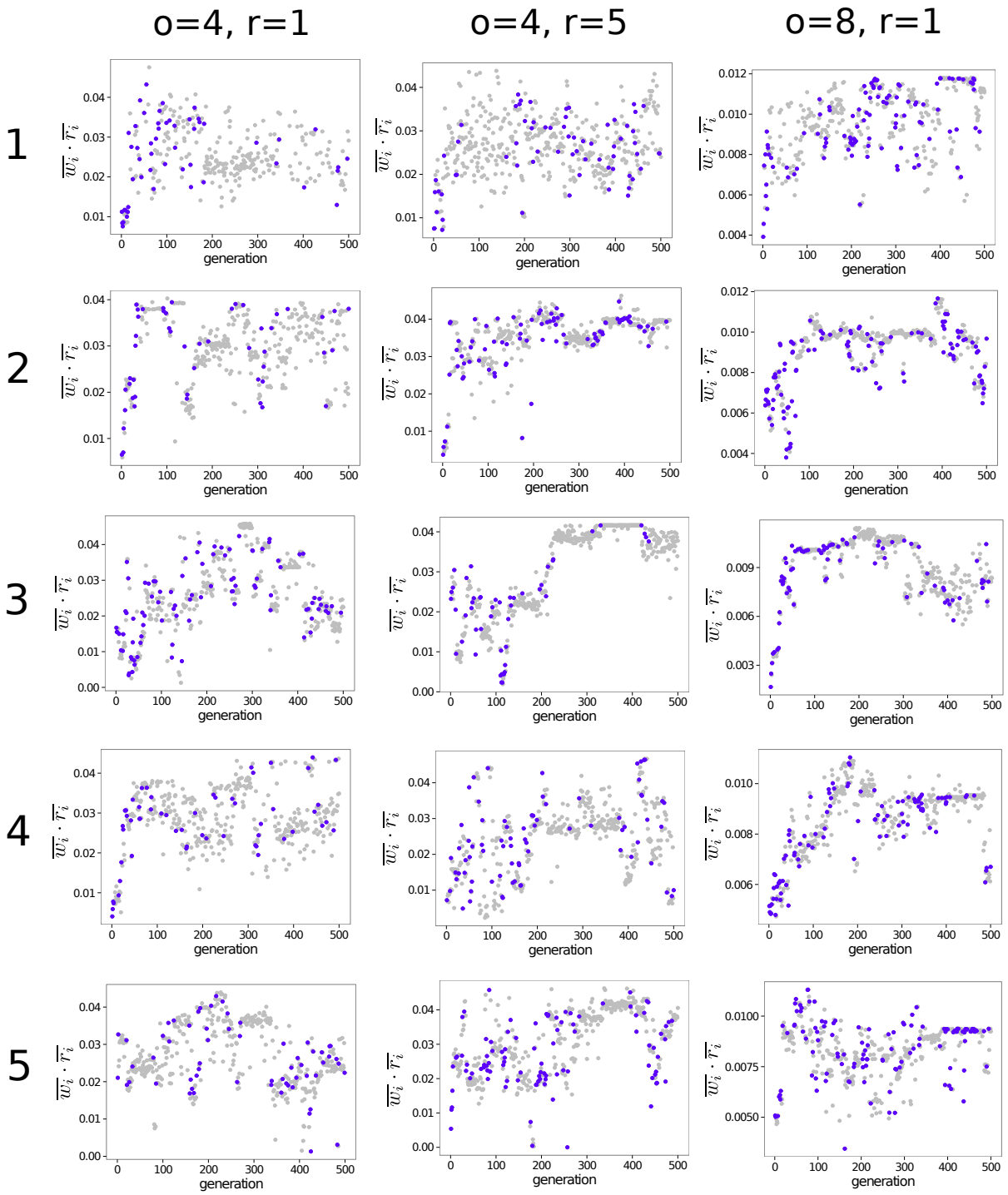



Figure 5.13: **Multiple evolution runs of artificial ecosystem (Network E-type).** Figure shows evolution of quantity mirroring fitness of the organisms for multiple runs *(vertical)* and for different parameters (horizontal) for number of simulation runs for evaluation r and number of organisms per ecosystem o. Each row of simulation runs was started with the same anorganic network but initialized with different organisms. (See Table 5.1 under *Network E* for additional parameters.)

Table 5.2: **Statistics of changing pathway properties for multiple evolution runs.** For each ecosystem type 20 ecosystems are generated and evolved. For each weighted pathway property the mean ratio between evolved ecosystems (later than generation 50) and the non evolved ecosystem is calculated. The error is estimated calculating the standard error of the mean.

| | | Ecosystem D | Ecosystem E |
|---|---|---|---|
| | Fitness | $19.4 \pm 13.4$ | $3.66 \pm 0.46$ |
| core NW | cycle number | $0.85 \pm 0.14$ | $0.52 \pm 0.18$ |
| | reaction number | $0.85 \pm 0.08$ | $0.73 \pm 0.07$ |
| | species number | $0.85 \pm 0.06$ | $0.82 \pm 0.05$ |
| | PW inf. Entropy | $0.70 \pm 0.07$ | $0.68 \pm 0.13$ |
| full NW | cycle number | $0.60 \pm 0.12$ | $0.61 \pm 0.09$ |
| | reaction number | $0.57 \pm 0.06$ | $0.58 \pm 0.05$ |
| | species number | $0.59 \pm 0.06$ | $0.66 \pm 0.04$ |
| | PW inf. Entropy | $0.78 \pm 0.11$ | $0.48 \pm 0.09$ |

## 5.3 DISCUSSION

### 5.3.1 *Evolution Towards Simpler Pathways*

The observed evolution towards simpler pathways is in accordance with the results of the previous chapter which indicated a higher importance of reaction pathways with fewer cycles and reactions in the modern atmosphere. As was already mentioned then, finding simpler pathways seems to contradict existing research associating life and evolution with more complexity. Increase of complexity has been found for chemical species[Pea93] as well as for genetic information [AOC00].

While simple pathways are not surprising in biology [SM16; MS07], the ones evolved might also be an effect of the simple environment. This could be tested by evaluating the fitness of organisms with multiple different simulation runs. One could for example have a higher number of organisms per anorganic network and then only solve the differential equation for the network containing different subsets of all organisms. Having the organisms compete in different situations might give an incentive for more complex functioning.

*Finding complex pathways would allow to check a proposed relationship between pathways and their dissipation [SU10] [US11].*

### 5.3.2 *Network Creation and Evolution Scheme*

In systems biology and artificial life there are countless models that evolve or grow networks [JK98; JK01; PSB05; Kre+08; HA08]. Most of them do not demonstrate the evolution of more complex reaction networks [YBC11]. While the application of reaction pathway analysis proved to be successful here, the process of network creation and

evolution can certainly be improved. The largest increase in fitness is always seen in the first 50 generations and no complex interaction between different organisms emerges. The problem lies in replacing entire organisms, which is a big disturbance for the system. It could be promising to have smaller mutations. A more fine-tuned mutation operator could for example change single reactions of the least fit organism.

To ensure evolvability for all possible ecosystems, mutation would also have to be able to change chemical species that are constrained by existing reactions. The stoichiometric constraints of elementary composition leads to a trade-off between evolvability and avoiding extreme mutations. Using some rule based schemes to create and mutate the network might allow to solve the problem of the mutation operator and increase the performance of network generation at the same time [BFS03b; BFS03a; Fae+05; And+14]. A different model for network generation might also reduce the number of parameters.

### 5.3.3 *Pathway View on Dynamic Stability*

Looking at reaction pathways in steady state has one drawback. The found pathways might not include the full information of how the system got into steady state or what is keeping it there. In flow reactors, where all the species are diluted constantly, pathways would describe the buildup of the autocatalytic structure that is countering the dilution. But we are having a constant mass, that is cycling inside the network. Once the concentrations and reaction rates are in steady state, the chemistry does not need to regenerate anything. To get the pathway view of how the network focuses its mass in certain parts of the network, it might be worthwhile to analyse how the system approaches its steady state. Pathway analysis could give a detailed explanation of how the central part of the network builds up from its environment and might allow to connect this to formal definitions of autocatalysis [HS12].

## 5.4 CONCLUSION

This chapter has shown how thermodynamically constrained artificial chemistries, together with reaction pathway analysis, can be used to create and analyse model ecosystems. An artificial chemistry model was created with mass conservation and thermodynamic constraints to be driven to cycling processes by photoreactions. Its functioning has been demonstrated with simple example systems. Investigating cyclic flow as a function of the thermodynamic driving force shows two regimes. In the lower regime, the strength of the flow follows a power law. In the higher regime, the flow is limited by the mass of the system. Pathway analysis shows that transition between those two regimes goes along with a reorganisation of reaction pathways. When increasing flow, also a reorganisation of the mass of the system into a smaller part of the network is detected.

The artificial chemistry was used as building block for generating a model for artificial ecosystems. This model allows to evolve the ecosystem by selectively replacing network parts representing different organisms. The use of this simple evolution scheme together with reaction pathway analysis makes it possible to observe an evolution towards simpler pathways with less cycles and less reactions. The statistical significance of this result has been successfully tested by concluding an ensemble of simulations. This allows for the understanding of the presence of pronounced simple pathways in real systems as a consequence of an evolutionary mechanism and explain its importance for life [MS07].

More work is needed for better models of the evolution process of the ecosystem. The primitive mutation operation of substituting entire organisms leads to strong fluctuations of fitness after a few initial generations. Also it is quickly leading to very simple reaction pathways that are focused on single organisms. A more fine-tuned evolution scheme might reduce fluctuations and might also lead to pathways that connect multiple organisms. In the ecosystem interpretation, pathways covering multiple organisms represent feeding relationships which are omnipresent in real ecosystems [OH72].

## APPENDIX: LIST OF NETWORKS AND PATHWAYS

This appendix contains tabular information on the structure of Networks A (Table 5.3), B (Table 5.3), and C (Table 5.6) as well as a list of the most significant reaction pathways found in the simulation of Network C (Table 5.5).

Table 5.3: **Network A** (1 elementary component; 4+1 species).

| | | | |
|---|---|---|---|
| $h\nu + {}_2C_1$ | $\rightarrow$ | ${}_3C_1$ | (1) |
| $h\nu + {}_3C_1$ | $\rightarrow$ | $C_1$ | (2) |
| ${}_3C_1$ | $\rightarrow$ | $C_1$ | (3) |
| ${}_3C_1 + C_1$ | $\rightarrow$ | ${}_3C_1 + {}_2C_1$ | (4) |
| ${}_4C_1 + {}_4C_1$ | $\rightarrow$ | ${}_3C_1 + {}_2C_1$ | (5) |
| ${}_4C_1 + {}_3C_1$ | $\rightarrow$ | ${}_2C_1 + C_1$ | (6) |

Table 5.4: **Network B** (2 elementary components; 4+1 species).

| | | | |
|---|---|---|---|
| $h\nu + C_1N_1$ | $\rightarrow$ | $C_1 + N_1$ | (1) |
| $h\nu + {}_2N_1$ | $\rightarrow$ | $N_1$ | (2) |
| ${}_2N_1$ | $\rightarrow$ | $N_1$ | (3) |
| $C_1N_1 + {}_2N_1$ | $\rightarrow$ | $C_1N_1 + N_1$ | (4) |
| $C_1N_1$ | $\rightarrow$ | $C_1 + N_1$ | (5) |
| $C_1N_1$ | $\rightarrow$ | ${}_2N_1 + C_1$ | (6) |

Table 5.5: **Selected pathways of Network C.** Pathways that are occurring in the pathway decomposition of Network C (cf. Fig. 5.8.)

| | | | | cycles | |
|---|---|---|---|---|---|
| | $h\nu + {}_2C_1N_1$ | $\rightarrow$ | ${}_2N_1 + {}_2C_1$ | | |
| | ${}_2C_1$ | $\rightarrow$ | $C_1$ | | |
| | $C_1N_1$ | $\rightarrow$ | ${}_2C_1N_1$ | 2 | (1) |
| | ${}_2N_1 + C_1$ | $\rightarrow$ | $C_1N_1$ | | |
| net: | $h\nu$ | $\rightarrow$ | | | |
| | $h\nu + C_1N_2$ | $\rightarrow$ | ${}_2C_1N_1 + {}_2N_1$ | | |
| | $h\nu + {}_2C_1N_1$ | $\rightarrow$ | ${}_2N_1 + {}_2C_1$ | | |
| 2X | ${}_2C_1$ | $\rightarrow$ | $C_1$ | | |
| | $C_1N_1 + C_1N_1$ | $\rightarrow$ | ${}_2N_2 + {}_2C_2$ | 8 | (2) |
| | ${}_2N_2 + {}_2C_2$ | $\rightarrow$ | $C_1N_2 + {}_2C_1$ | | |
| 2X | ${}_2N_1 + C_1$ | $\rightarrow$ | $C_1N_1$ | | |
| net: | $2h\nu$ | $\rightarrow$ | | | |

(continued on next page)

Table 5.5: (continued - pathways of Network C)

| | | | | cycles | |
|---|---|---|---|---|---|
| | $h\nu + C_1N_2$ | $\rightarrow$ | $_2C_1N_1 + {}_2N_1$ | 17 | (4) |
| | $h\nu + {}_2C_1N_1$ | $\rightarrow$ | $_2N_1 + {}_2C_1$ | | |
| 2X | $_2C_1$ | $\rightarrow$ | $C_1$ | | |
| | $N_1$ | $\rightarrow$ | $_2N_1$ | | |
| | $C_1N_1 + {}_2N_2$ | $\rightarrow$ | $C_1N_2 + N_1$ | | |
| | $C_1N_1 + {}_2N_1$ | $\rightarrow$ | $_2N_2 + {}_2C_1$ | | |
| 2X | $_2N_1 + C_1$ | $\rightarrow$ | $C_1N_1$ | | |
| net: | $2h\nu$ | $\rightarrow$ | | | |
| | $h\nu + N_2$ | $\rightarrow$ | $_2N_1 + {}_2N_1$ | 10 | (5) |
| 2X | $_2C_1$ | $\rightarrow$ | $C_1$ | | |
| | $_2N_2 + C_1$ | $\rightarrow$ | $N_2 + {}_2C_1$ | | |
| | $C_1N_1 + {}_2N_1$ | $\rightarrow$ | $_2N_2 + {}_2C_1$ | | |
| | $_2N_1 + C_1$ | $\rightarrow$ | $C_1N_1$ | | |
| net: | $h\nu$ | $\rightarrow$ | | | |
| | $h\nu + N_2$ | $\rightarrow$ | $_2N_1 + {}_2N_1$ | 14 | (21) |
| | $h\nu + {}_2C_1N_1$ | $\rightarrow$ | $_2N_1 + {}_2C_1$ | | |
| 2X | $_2C_1$ | $\rightarrow$ | $C_1$ | | |
| | $C_1N_1 + {}_2N_1$ | $\rightarrow$ | $_2N_2 + {}_2C_1$ | | |
| | $_2C_1N_1 + C_1N_1$ | $\rightarrow$ | $N_2 + C_2$ | | |
| | $_2N_2 + C_2$ | $\rightarrow$ | $_2C_1N_1 + {}_2C_1N_1$ | | |
| 2X | $_2N_1 + C_1$ | $\rightarrow$ | $C_1N_1$ | | |
| net: | $2h\nu$ | $\rightarrow$ | | | |
| | $h\nu + N_2$ | $\rightarrow$ | $_2N_1 + {}_2N_1$ | 4 | (19) |
| | $_2C_1$ | $\rightarrow$ | $C_1$ | | |
| | $_2N_2$ | $\rightarrow$ | $N_2$ | | |
| | $C_1N_1 + {}_2N_1$ | $\rightarrow$ | $_2N_2 + {}_2C_1$ | | |
| | $_2N_1 + C_1$ | $\rightarrow$ | $C_1N_1$ | | |
| net: | $h\nu$ | $\rightarrow$ | | | |
| | $h\nu + C_1N_2$ | $\rightarrow$ | $_2C_1N_1 + {}_2N_1$ | 14 | (37) |
| | $h\nu + {}_2C_1N_1$ | $\rightarrow$ | $_2N_1 + {}_2C_1$ | | |
| 2X | $_2N_2$ | $\rightarrow$ | $N_2$ | | |
| 2X | $N_2 + {}_2C_1$ | $\rightarrow$ | $_2N_2 + C_1$ | | |
| | $C_1N_1 + C_1N_1$ | $\rightarrow$ | $_2N_2 + {}_2C_2$ | | |
| | $_2N_2 + {}_2C_2$ | $\rightarrow$ | $C_1N_2 + {}_2C_1$ | | |
| 2X | $_2N_1 + C_1$ | $\rightarrow$ | $C_1N_1$ | | |
| net: | $2h\nu$ | $\rightarrow$ | | | |
| | $h\nu + {}_2C_1N_1$ | $\rightarrow$ | $_2N_1 + {}_2C_1$ | 9 | (43) |
| | $N_2 + {}_2C_1$ | $\rightarrow$ | $_2N_2 + C_1$ | | |
| | $_2C_1N_1 + C_1N_1$ | $\rightarrow$ | $N_2 + C_2$ | | |
| | $_2N_2 + C_2$ | $\rightarrow$ | $_2C_1N_1 + {}_2C_1N_1$ | | |
| | $_2N_1 + C_1$ | $\rightarrow$ | $C_1N_1$ | | |
| net: | $h\nu$ | $\rightarrow$ | | | |

(continued on next page)

Table 5.5: (continued - pathways of Network C)

| | | | | cycles | |
|---|---|---|---|---|---|
| | $h\nu + C_1N_2$ | $\rightarrow$ | $_2C_1N_1 + {}_2N_1$ | | |
| | $C_2$ | $\rightarrow$ | $_2C_2$ | | |
| | $N_2 + {}_2C_1$ | $\rightarrow$ | $_2N_2 + C_1$ | | |
| | $_2C_1N_1 + C_1N_1$ | $\rightarrow$ | $N_2 + C_2$ | 9 | (44) |
| | $_2N_2 + {}_2C_2$ | $\rightarrow$ | $C_1N_2 + {}_2C_1$ | | |
| | $_2N_1 + C_1$ | $\rightarrow$ | $C_1N_1$ | | |
| net: | $h\nu$ | $\rightarrow$ | | | |

Table 5.6: **Network C** (2 elementary components; 12+1 species).

| | | | |
|---|---|---|---|
| $h\nu + {}_2C_2$ | $\rightarrow$ | $_2C_1 + C_1$ | (1) |
| $h\nu + C_1N_2$ | $\rightarrow$ | $_2C_1N_2$ | (2) |
| $h\nu + {}_2C_1N_1$ | $\rightarrow$ | $_2C_1 + N_1$ | (3) |
| $_2C_1$ | $\rightarrow$ | $C_1$ | (4) |
| $_2C_2$ | $\rightarrow$ | $C_2$ | (5) |
| $_2C_1N_1$ | $\rightarrow$ | $C_1N_1$ | (6) |
| $_2C_1N_2$ | $\rightarrow$ | $C_1N_2$ | (7) |
| $_2N_1$ | $\rightarrow$ | $N_1$ | (8) |
| $C_1N_1 + {}_2C_1$ | $\rightarrow$ | $C_2 + {}_2N_1$ | (9) |
| $_2C_2 + N_1$ | $\rightarrow$ | $C_1N_1 + C_1$ | (10) |
| $_2C_1N_2 + C_1$ | $\rightarrow$ | $C_1N_2 + {}_2C_1$ | (11) |
| $N_3 + {}_2C_2$ | $\rightarrow$ | $_2C_1N_2 + {}_2C_1N_1$ | (12) |
| $N_3 + {}_2C_2$ | $\rightarrow$ | $C_1N_2 + C_1N_1$ | (13) |
| $N_2 + {}_2C_1$ | $\rightarrow$ | $C_1N_1 + {}_2N_1$ | (14) |
| $_2C_2$ | $\rightarrow$ | $C_1 + C_1$ | (15) |
| $N_3 + {}_2C_1N_1$ | $\rightarrow$ | $C_1N_2 + N_2$ | (16) |
| $C_2$ | $\rightarrow$ | $_2C_1 + C_1$ | (17) |
| $_2C_2 + C_1N_1$ | $\rightarrow$ | $_2C_1N_1 + C_2$ | (18) |

# 6 CONCLUSION

*The conclusion refers to parts of all previous chapters.*

Methods for the analysis of reaction networks have been developed and successfully combined with a thermodynamically driven artificial chemistry. Two models of Earth's atmosphere have been compared in terms of reaction pathways and thermodynamic functioning. Using an artificial chemistry to create an evolvable artificial ecosystem has shown that evolving ecosystems leads to the formation of simpler, more pronounced reaction pathways.
**keywords:** reaction network, pathway analysis, artificial chemistry, complex network, thermodynamics, life detection

In this thesis, methods for the analysis of reaction networks through thermodynamics and reaction pathways were extended and combined with novel models for artificial chemistries that include reversible reactions. Applications in the range of origins of life and Earth system science have been tested and further connections have been discussed.

*Cf. Chapter 3.*

The generated large-scale artificial reaction networks have shown the importance of thermodynamic constraints. Without those, the reversible dynamics-induced self organisation leading to an increase in cycles is not possible. Searching for autocatalytic sets without considering this type of self organisation might miss important factors. Besides increasing the number of cycles, increasing the flow through nonlinear networks also leads to a narrower distribution of chemical potentials. Comparing different complex network types did not indicate large differences in their thermodynamic functioning. The biggest difference was found between linear and nonlinear networks. The distribution of entropy production rate of individual reactions is steeper for nonlinear networks. The power law coefficient was found to be approximately $-1.66$ for them instead of $-1.5$ for linear networks.

*See Chapter 4.*

Using pathway analysis and thermodynamic characterization, two models of Earth atmospheric chemistry could successfully be distinguished. Pathways of the modern Earth atmosphere were found to be simpler (less reactions) and contain fewer cycles than those of the Archean atmosphere. Thermodynamic analysis showed that the Archean atmosphere is driven more strongly by radiation, generally increasing the chemical energy of the matter it exchanges with its environment, while the matter fluxes of the modern atmosphere have a negative net energy balance. The estimated chemical potentials allowed for a successful quantification of methane oxidation in

*Chemicals passing the modern atmosphere (on average) decrease their chemical energy.*

the modern atmosphere in accordance with previous studies. Nevertheless, many chemical species and reaction pathways could not be quantified. Some pathways were even found to be inconsistent with the calculated thermodynamic potentials. These inconsistent pathways can be used to further improve future models. But they also provide insight into what needs to be done to achieve a comprehensive thermodynamic model analysis.

*See Chapter 5.*

For analysis with the pathway decomposition algorithm, an artificial chemistry model was created. The model ensures mass conservation and thermodynamic constraints while it is driven to thermodynamic disequilibrium through photochemistry-like reactions. It has been tested with different example systems to show how a change in the driving force leads to a reorganisation of reaction pathways. For increasing flow, the mass of the system concentrates within a smaller number of chemical species. The artificial chemistry was used as a building block for generating a model for artificial ecosystems. This model enables network evolution by selectively replacing network parts representing different organisms. The usage of this simple evolution scheme together with reaction pathway analysis provides insight into an evolution towards simpler pathways with less cycles and less reactions. This allows for an understanding of the presence of pronounced simple pathways in real systems as a consequence of an evolutionary mechanism. The primitive mutation operation of substituting entire organisms shows its limit by leading to strong fluctuations of fitness after a few initial generations with large fitness gain. Future work in this area should focus on achieving a more finely tuned evolution scheme together with an artificial chemistry whose network construction is computationally less complex.

In Chapter 3 it has been shown that Earth's atmospheric reaction network has fewer cycles than its null model with randomized reaction directions. This fits the lower cycle number found for the modern atmosphere in Chapter 4. Also in Chapter 5, a decrease of the cycle number[1] was found for evolved ecosystems, for the analysis of the anorganic part as well as of entire networks.

[1] *Weighted number of cycles in pathways.*

The simulations of large scale reaction networks in Chapter 3 indicated that when flow is increased, a larger fraction of the network is necessary to be able to explain the same fraction of dissipation (Fig. 3.6 (c)). The artificial chemistries in Chapter 5, however, concentrate their mass in smaller parts of the network with increasing flow. This is not a contradiction as both networks are simulated with different boundary conditions. The large-scale network is driven by flow of matter and increases its mass, while the artificial chemistry that is driven by photoreactions retains a constant mass.

The work presented has shown that combining pathway analysis with planetary modelling as well as artificial chemistry is full of potentials. Nonetheless, further research is needed to make the algorithm scale better for large systems and the models more easily accessible to analysis.[2] In the final chapter, possible ideas for a better combination of reaction networks with planetary modelling and artificial life are discussed.

[2] *For example through common data formats and model description languages.*

# 7 OUTLOOK

*Previous chapters (4 and 5) and questions formulated in the introduction (Chapter 1) are referred to often. The technical part requires an understanding of methods from Chapter 2.*

The potential of compiling planetary reaction networks for future studies is discussed. To demonstrate the approach, a toy model-like sample network of Earth's energy and matter flows is assembled and its pathway decomposition calculated. The potential of reaction network-based techniques is demonstrated with the help of two special artificial chemistries (bistable, oscillating dynamics). Further possible applications in the area of artificial life are discussed.
**keywords:** planetary reaction network, pathway analysis, artificial life, reaction diffusion systems, planetary life detection

In the previous chapters, reaction pathways have been used to identify atmospheric chemistries. A model of artificial ecosystems was developed and its evolving structure analysed. In the following, various ideas for further application of reaction pathway analysis in the areas of artificial life and planetary science are presented and discussed.

## 7.1 PLANETARY REACTION NETWORK

Biogeochemical cycles like the carbon cycle and the nitrogen cycle have an important function for life on Earth as they regulate planetary conditions and the recycling of matter [Fal+00; GG08]. After using reaction pathways to analyse the atmosphere in Chapter 4, the question of how the reaction pathways found are connected to the rest of the planet cannot be answered easily. Assuming we had a detailed model of reactions and flows in the geosphere and biosphere, should it not be possible to connect those to the atmospheric reaction network and get one big planetary reaction network? Extending the reaction pathway analysis to this planetary network might not only help understand how atmosphere, biosphere and geosphere interact in detail, but the important reaction pathways should directly correspond to the known biogeochemical cycles. The main structures determining the chemical composition of our planet could be automatically evaluated and compared through model analysis.

The idea of this approach is depicted in Fig. 7.1). The general idea might be applied to Earth, planets of our solar system, and even exoplanets, as far as observations allow to build a process model. The latter, however, seems to be quite far-fetched, considering that often we only have access to a noisy spectrum and an estimated size for extrasolar planets [Sea13]. Nonetheless, future progress in astronomical observations and modelling might allow us to apply the overall

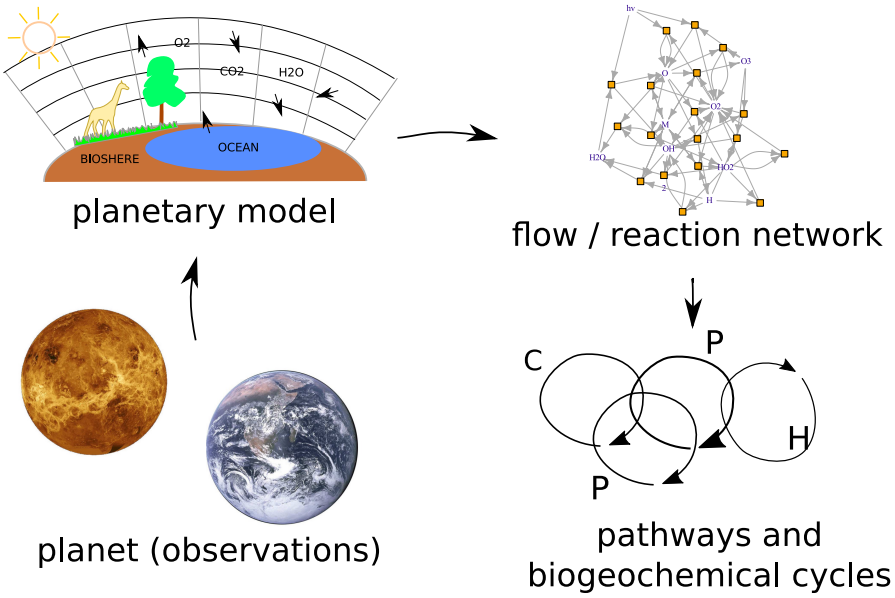



Figure 7.1: **Idea for planetary reaction network.** Starting from an integrated model, informed by extensive observations of the planet, a reaction network containing all important flow and reaction processes of the planetary system can be derived. An analysis of this network will give an integrated view on the planet's processes (biogeochemical cycles) as well as their evolution and sensitivity to changes. (Source of planet images: NASA/public domain.)

idea even for these cases. The following section, however, will focus on a discussion of the idea for Earth only. After giving an elementary example of a network that connects Earth's climate system with a simple representation of the water and the carbon cycle, we will focus on discussing how this idea might be used for larger models.

### 7.1.1 *Pathways of the Climate System*

*Geological processes and timescales are not considered here. However, doing so and considering alternative (quasi-)steady states would be an interesting extension.*

To demonstrate how such a planetary reaction network integrating different kinds of processes might work, we build a simple toy model. It is conceptually based on simple energy-balance climate models [NCC81] and contains a primitive water and (organic) carbon cycle. The main aim here is to demonstrate the concept. Thus, not all processes are directly related to physical processes. Quantities are only roughly estimated by using different sources. Nevertheless, the model is complete in the sense that the steady state condition is fulfilled (by design).

To connect carbon cycling and water cycling with energy, we calculate simple constant factors between these flows. We do not go into detail regarding the question through which physical processes these flows are connected and do not try to calculate chemical potentials for the different species in the reaction network. Estimating Earth's surface temperature at 15°C, the energy necessary for evaporating water

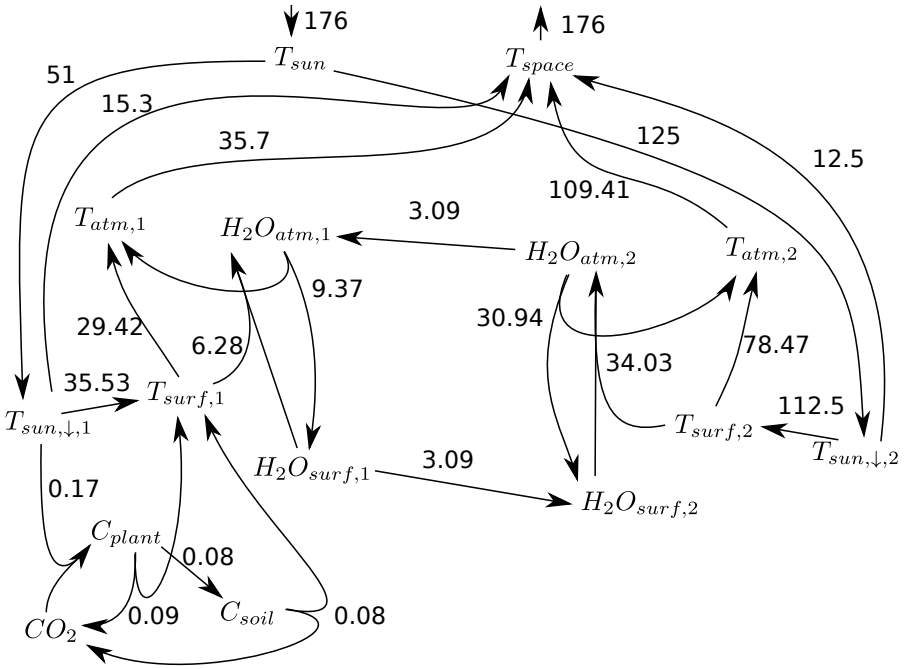



Figure 7.2: **Sample Earth-climate network inspired from simple energy balance climate model.** A terrestrial carbon cycle has been added and rates have been estimated according to data giving in the body of literature and by assuming steady state condition. The left side (subscript 1) represents terrestrial processes while processes on the right side (subscript 2) are oceanic. Units for power are PW ($10^{15}$W). For the carbon cycle, this corresponds to $2.27 \cdot 10^{10}$g/s C. For the water cycle it is $3.823 \cdot 10^{5}$m$^3$/s $H_2O$.

is the sum of the heat needed to warm it to 100°C (heat capacity is 4.19kJ/(kgK)) and the evaporation enthalpy (2260kJ/kg) [SS02]:

$$1\text{J} \equiv 3.823 \cdot 10^{-10}\text{m}^3\text{H}_2\text{O}. \tag{7.1}$$

Photosynthesis is estimated by the chemical reaction

$$\text{CO}_2 + \text{H}_2\text{O} \rightarrow \text{CH}_2\text{O} + \text{O}_2 \tag{7.2}$$

for which the change of formation enthalpy is given as $528.96 \cdot$ kJ/mol [Dou11, p. 657], leading to the following relationship between energy flow and carbon cycle:

$$1\text{J} \equiv 2.269 \cdot 10^{-5}\text{gC}. \tag{7.3}$$

*Network Structure*

The network structure is shown in Fig. 7.2. It consists of two parts describing processes happening on land (subscript *1*) and processes happening at the oceans (subscript *2*). The solar radiation hitting Earth $T_{sun}$ is split into one part for Earth's land surface $T_{sun,\downarrow,1}$ and another for its oceans $T_{sun,\downarrow,2}$. For each of these downward-facing radiative fluxes, one part is directly reflected back to space $T_{space}$ whereas another part is transformed into heat and after some intermediate[1] steps re-emitted into space as longwave radiation ($T_{atm,i} \rightarrow T_{space}$).

[1] *This may include radiative as well as convective transport.*

Part of the surface heat is the driving force behind the water cycle as it causes the evaporation of water ($\rightarrow T_{surf,i} + H_2O_{surf,i} \rightarrow H_2O_{atm,i}$). Evaporation is balanced by condensation and precipitation ($H_2O_{atm,i} \rightarrow H_2O_{surf,i} + T_{atm,i}$) and transport of water vapour landwards ($H_2O_{atm,2} \rightarrow H_2O_{atm,1}$) as well as by continental discharge into the oceans ($H_2O_{surf,1} \rightarrow H_2O_{surf,2}$).

A primitive carbon cycle is included on land, driven by part of the downward-facing radiation transforming carbon into biomass ($T_{sun,\downarrow,1} + CO_2 \rightarrow C_{plant}$). From there, organic carbon is either used for respiration ($C_{plant} \rightarrow T_{surface,1} + CO_2$) or used for plant growth and decomposed after the plant dies (via $C_{soil}$).

*Steady State Rates*

Using the solar constant of 1,370 $W/m^2$ and the Earth radius of 6,400km [SS02], we estimate the total influx of solar radiation to Earth at 176PW (petawatt, $10^{15}$W). Ignoring exact latutudinal distribution of land masses, we split the irradiation between land masses and oceans by simply using the fraction of Earth's surface they occupy[2].

[2] *29% to 71% [SS02]*

With a rough estimate for the albedo of 0.3 over land and 0.1 over oceans [SS02], we get the values for the upward flux of short wave radiation. Using the equivalency of water with energy (Eq. 7.1) and estimates for global continental discharge[3] and precipitation over land[4] and ocean[5] allows us to quantify all reactions related to the water cycle.

[3] $37288km^3/a = 1.18 \cdot 10^6 m^3/s$ *[DT02]*

[4] $1.13 \cdot 10^{14} m^3/a = 3.583 \cdot 10^6 m^3/s$ *[Tre+07]*

[5] $3.73 \cdot 10^{14} m^3/a = 1.18 \cdot 10^7 m^3/s$ *[Tre+07]*

Rates for (terrestrial) photosynthesis are calculated using estimates for the gross primary production (GPP) of 120PgC/a [Bee+10]. Respiration is calculated from the difference of GPP and net primary production (NPP). According to [Fie+98], NPP is estimated to be around 56.4PgC/a. With all these rates known, all other rates in the reaction network (Fig. 7.2) can be calculated by assuming a steady state condition, i.e. assuming that the production of each species exactly balances its consumption.

*Pathway Analysis*

The steady state vector can then be decomposed into reaction pathways (Fig. 7.3). As all pathways transform energy between the same inflow and outflow species, the coefficients $\alpha_i$ are directly proportional to the amount of energy passing through them.

$\mathbf{v} = \sum_i \alpha_i \mathbf{E_i}$ *(see Sect. 2.2.3 and Sect. 2.3)*

In the pathways with high coefficients we can see the importance of the physical climate system. The most important pathways describe how energy is transported from Earth's surface to space: through directly reflected solar radiation (pathw. (4)+(5)), through convection or long wave radiation (pathw. (1)+(3)) or through latent head (pathw. (2)+(6)+(7)). The remaining pathways are all connected to our primitive carbon cycle and are one to two magnitudes weaker than the pathways mentioned before.

This implies that pathways directly related to life on a planet do not have to be within the most important pathways found using a

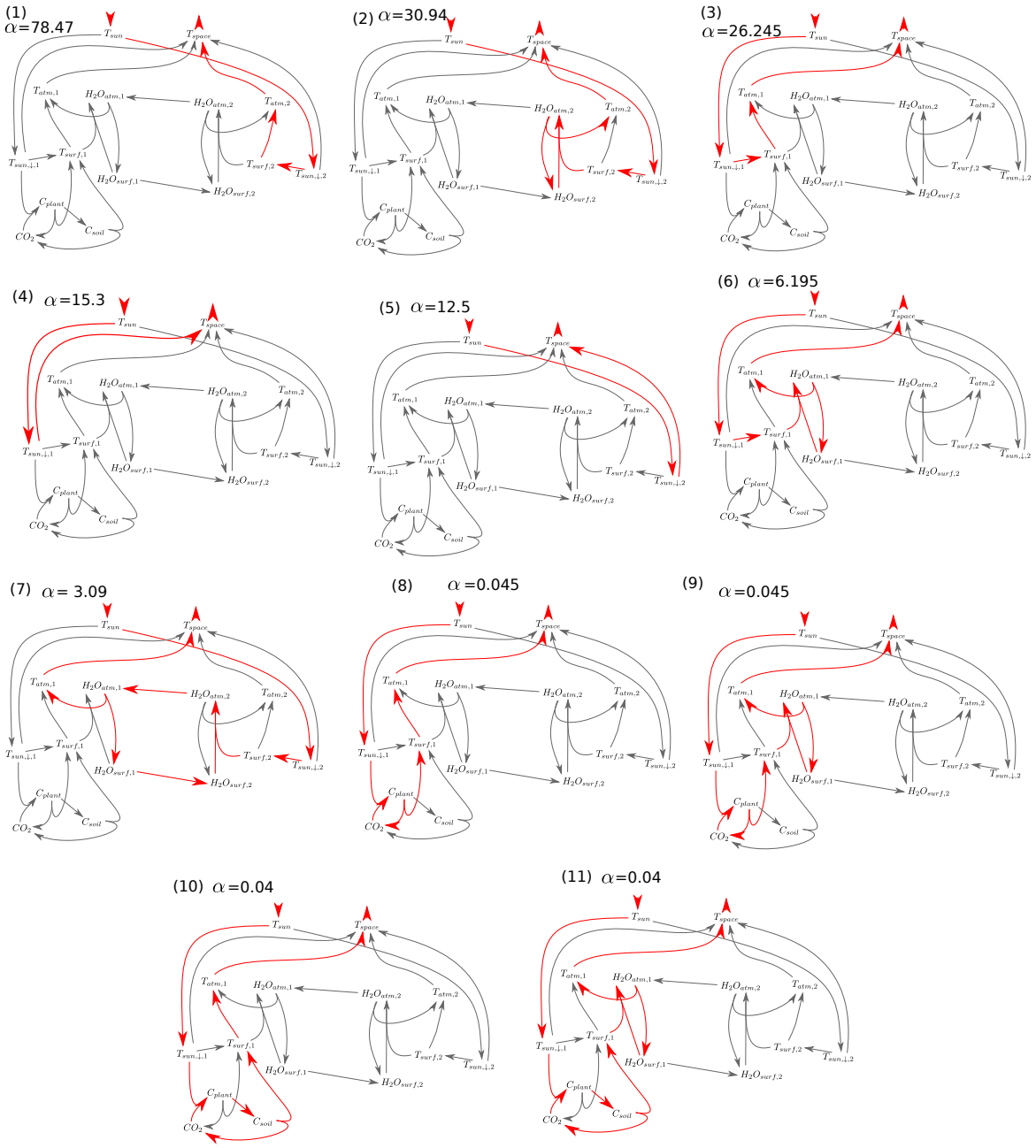



Figure 7.3: **Pathways of sample Earth climate network.** Pathway decomposition of the steady state / reaction network shown in Fig. 7.2. Units for power are PW ($10^{15}$W). For the carbon cycle, this corresponds to $2.27 \cdot 10^{10}$g/s C. For the water cycle it is $3.823 \cdot 10^{5}$m$^3$/s $H_2O$.

model of this planet. For the sake of analysis, it is necessary to compare different states of the same model or different models that are comparable in structure, like it was done in Chapter 4.

*Previous work indicates similar relations with respect to Earth's entropy budget [KMC10].*

It should be mentioned that the pathways found here are, of course, influenced by the decisions made while assembling the network. One could argue that the dominance of some pathways in the decomposition just indicates that this part of the network needs to be resolved in a more suitable way or that we could highlight any part of the network through extreme high pathways if we just design the right

networks. This is partially true if we assume the person assembling the model to be malicious. This, however, will not be the case if we assemble the network to be a meaningful model to the best of our knowledge and abilities.

### 7.1.2 *Further Steps*

A general limitation of analysing existing models derives from a lack of formal description of processes that happen outside the models' core components[6]. Future improvement in modelling and analysis could be achieved especially by using techniques from Systems Biology for handling large-scale models [Knü+13]. This would include a description of processes across model boundaries in a way similar to chemical models.

[6] *In the example of atmospheric chemistry models (Chapter 4), processes outside the gas phase chemistry are not formally described.*

Practical first steps could be explored with dynamical system models of Earth's climate [SB97; SB98; Svi99]. They allow to study feedback processes and dynamical stability. There are also models with multiple compartments for the carbon cycles for which the use of pathway analysis could be demonstrated [GZ08]. Using a network approach to spatially resolved models could be tested with the model of Dommenget and Flöter [DF11], a spatially resolved climate model for teaching purposes.

If one aims for analysing more complex models, our approach is currently limited by the inherent complexity of the algorithm for pathway decomposition.[7] The only possibility to control running time is by altering two algorithmic parameters. Even though this makes it rather easy to use this approach, it is difficult to find a good balance between running time and resolution (number of pathways). Especially the resolutions of different parts of the network can not be controlled separately. Apart from using more recent techniques for elementary mode analysis [TS08], a promising idea for future work is to extend the algorithm to allow for the fine tuning of algorithmic parameters to specific areas of the reaction network. If multiple iterations of the pathway calculation were done, the algorithm could first calculate pathways with a focus on speed and then adapt cutoff parameters to increase resolution in parts of the network for successive runs. Using a stochastic search, comparable to the way cycles are often searched [BC15], might also provide the advantages of an anytime algorithm.

[7] *A 10 variable model on a 100x100 grid would be represented by a $10^5$ species network.*

## 7.2 ADVANCED ARTIFICIAL ECOSYSTEMS

*Bistable Chemistry*

The artificial chemistry models created in this thesis all were based on the assumption that, besides the mass of the system, every simulated chemistry ends up in an deterministically decided steady state. This does not always have to be true. For biology, multistationary systems have an important function for differentiation and memory [TK01;

Pie+10]. Some theoretical work has been done on conditions for bistability [CTF06], finding minimal bistable systems [WSH97; Wil09; JS13] and analysing the relationship between thermodynamic stability and dynamic stability.

Through semiautomatic search, an example for an artificial chemistry with bistable behaviour was found.[8] The results of its simulation results are shown in Fig. 7.4. For a high enough concentration of $h\nu$, the network can end up in two different steady states. They are quantitatively distinct in the amount inflow of $h\nu$, yet also show some reactions with different directions (Fig. B.1) leading to distinct, incompatible pathways.[9]

[8] *See Sect. B.2, network shown in Fig. B.1 and Table B.3.*

[9] *Pathways 188&249 in Fig. B.1.*

*Oscillating Chemistry*

Another example for non-trivial chemistry are systems with periodic oscillation. A prominent example for this in real-life chemistry is the Belousov-Zhabotinsky (BZ) reaction [ES96; LO08].[10]. Structurally, oscillating chemistry can be seen as dynamical self-assembly that is stabilised by dissipating energy [WG02]. Chemistry with periodic dynamics is also interesting from a biological perspective. For cyanobacteria in a periodic environment it has been shown that the possession of an circadian clock (internal oscillator) gives an adaptive advantage [Woe+04].

[10] *There are periodically contracting polymer gels controlled by a BZ reaction.[Yos10]*

In Fig. 7.5 results of an artificial chemistry that exhibits oscillating behaviour is shown. Like the bistable chemistry, this artificial chemistry was found using semiautomatic search.[11] The oscillatory behaviour can be seen in the Gibbs free energy of reaction for selected reactions, but also for the concentrations of selected chemical species. Because the system does not reach a steady state, the pathway decomposition as a function of time consists of steady state pathways as well as pathways that are transforming chemical species (Fig. 7.5 **(c)** & **(e)**). For our example system, the steady state pathways seem to make up most of the reaction rates. Nevertheless, we can clearly see how the other pathways are cycling species periodically. Pathway 3125 transforms ${}_{11}C_1$ into ${}_{9}C_1$ while Pathway 3382 transforms in the reverse direction.[12] The same happens between ${}_{11}C_1$ and ${}_{5}C_1$ with Pathways 3122 and 3201. Taken together, these pairs of pathways explain how material is cycled by the periodic dynamics which cannot be done by simply averaging the steady state over time. Even though time averaging the reaction rates leads to rates that satisfy the steady state condition, it also reduces the number of reaction pathways as reactions that occur with both directions in time-resolved analysis only have one averaged direction.

[11] *See Sect. B.2. Network shown in Fig. B.2, and Table B.4.*

[12] *In the figures, isotopes ${}_{x}C_{i}$ are written as* Ci_x*!*

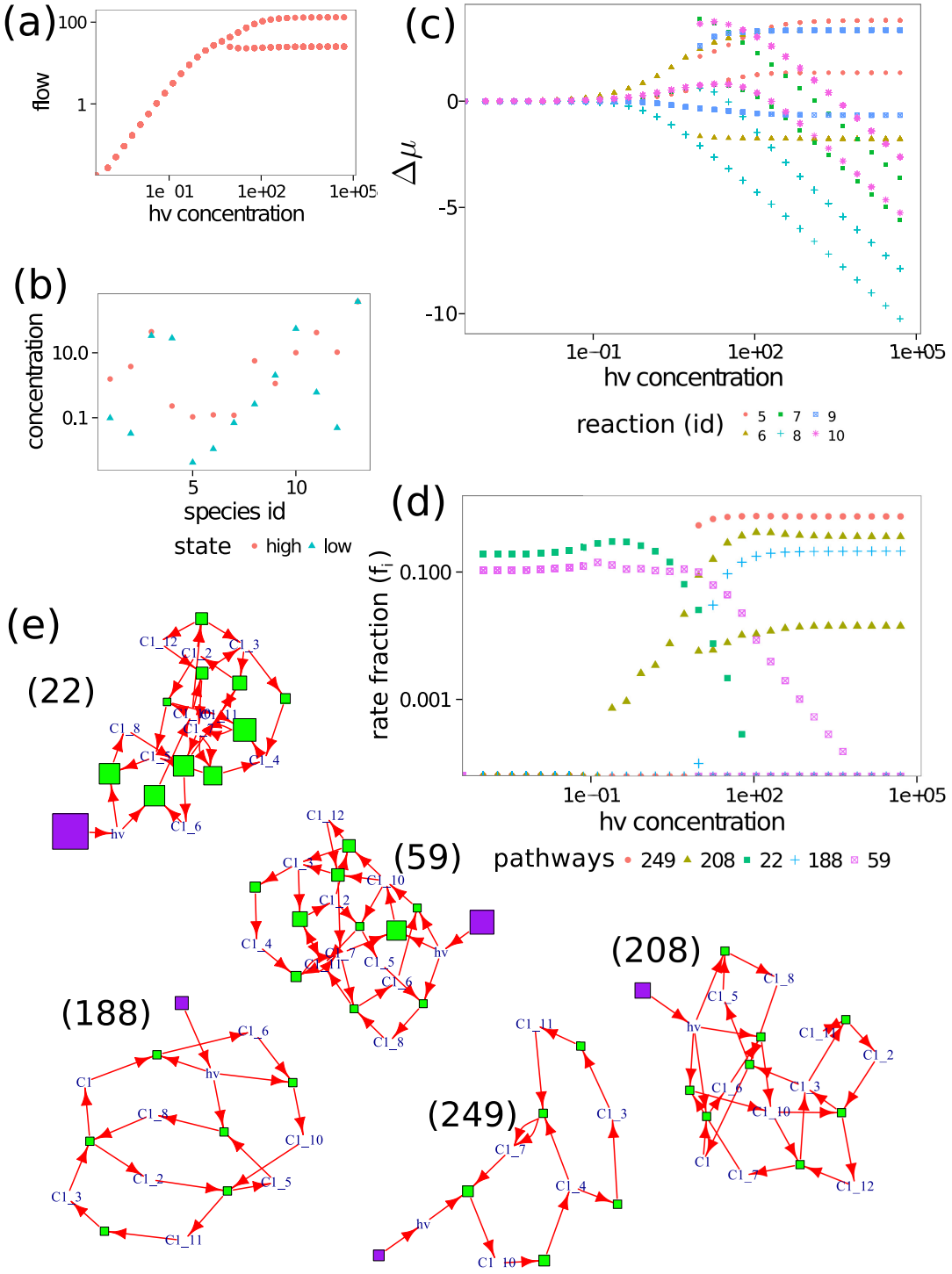



Figure 7.4: **Sample results for bistable artificial chemistry.** (Reactions shown in Fig. B.1 and Table B.3.) Initial concentration for all simulations was 10. **(a)** Energy flow (consumed hν-species) as function of driving force (hν concentration). **(b)** Species concentrations for the two different steady states at [hν] = 100. **(c)** Gibbs free energy for selected reactions. **(d)** Pathway decomposition for changing driving force. **(e)** Most important pathways in pathway decomposition. (Size of squares symbolizes coefficient of respective reaction.)

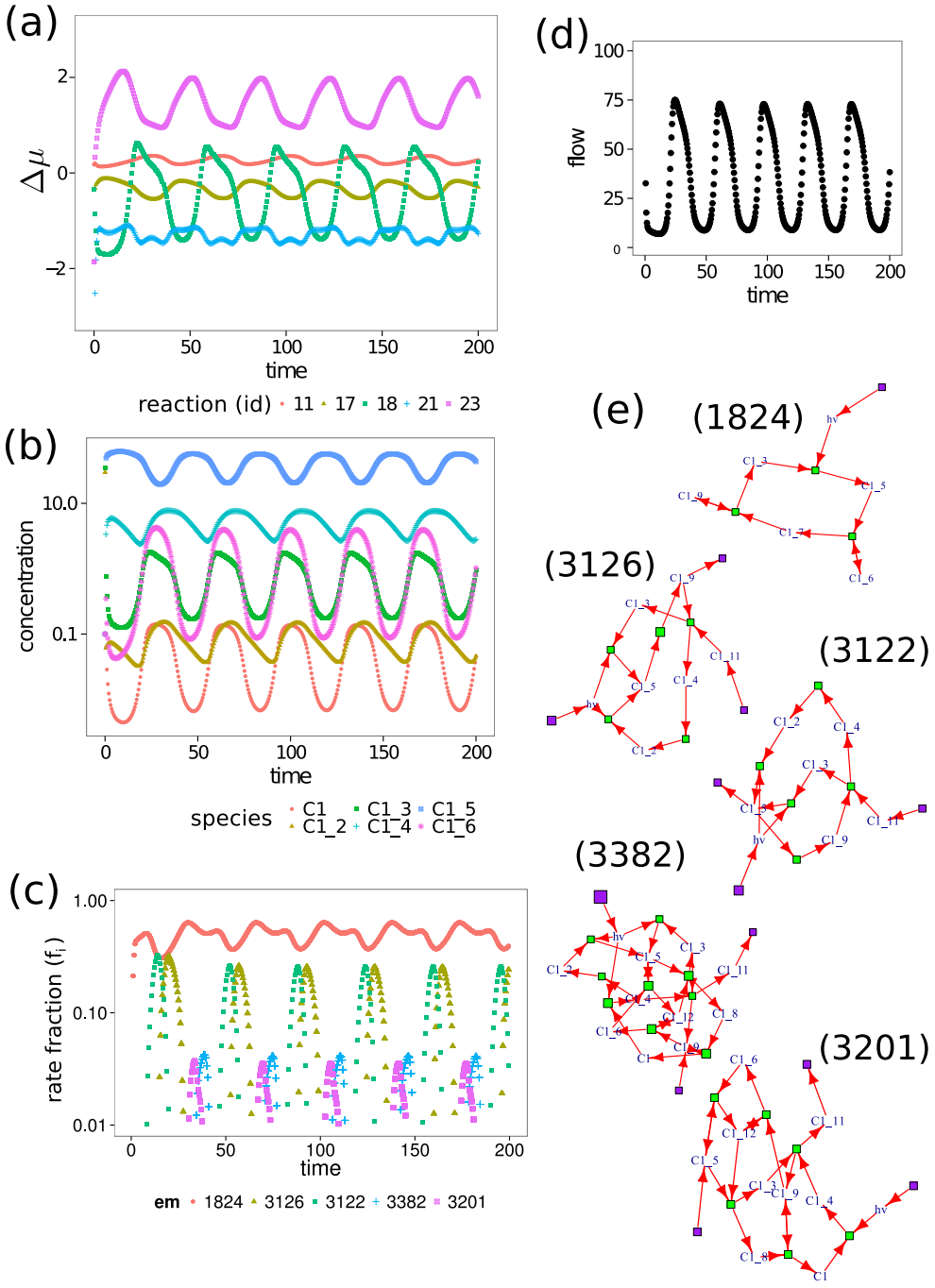



Figure 7.5: **Results for sample artificial chemistry with oscillating dynamics.** (Reactions shown in Fig. B.2 and Table B.4.) **(a)** Gibbs free energy for selected reactions. **(b)** Concentration of selected reactions. **(c)** Selected pathways in pathway decomposition. **(d)** Oscillation of energy flow through the network (consumed hv-species). **(e)** Most significant steady state pathway (1824) and the four most significant pathways that change concentrations of species (3126, 3122, 3381, 3201). (Size of squares symbolizes coefficient of respective reaction.)

*Integration into Artificial Ecosystems*

For both bistable chemistry and oscillating chemistry, a way to integrate them in our model of evolving artificial ecosystems still needs to be found. It is unlikely that artificial evolution will find such networks naturally and allow them to persist. Instead, a fitness function that allows these nontrivial networks to gain an advantage needs to be chosen. It should include multiple scenarios for which the differential equation of the networks will be solved. Some of them might require a time-dependent boundary condition. A periodic driving force could give an incentive for oscillating networks similar to the circadian clock in cyanobacteria [Woe+04]. Switching stochastically between a high and low driving force, respectively, and attributing higher fitness if the network maintains a higher flow after switching down might help developing bistable systems.

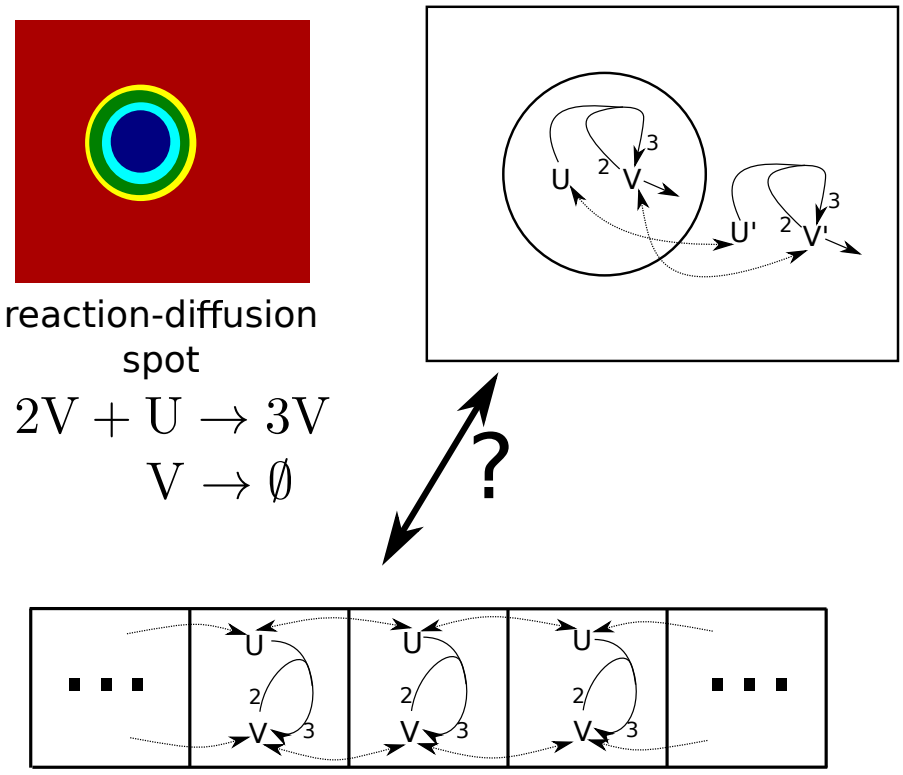



Figure 7.6: **Analysis of reaction diffusion systems through reaction networks.** Reaction diffusion systems are simple models that have been used to argue about pattern formation and morphogenesis. After shapes are identified, a simple reaction network can be assembled **(top-right)**. Simultaneously, the network of reactions and diffusion **(bottom)** can be investigated and related to the simplified network. Further, the implications of pattern types on the networks can be studied.

## 7.3 FURTHER APPLICATIONS IN ARTIFICIAL LIFE

### *Reaction Diffusion Systems*

Extending the description of oscillating chemical systems into space leads to reaction diffusion systems. The reactions are happening in the cells of a lattice and their products and educts are diffusing between next neighbour sites. Formally, the grid is the spatial discretisation in which the partial differential equation

$$\partial_t \mathbf{u} = \mathbf{D}\nabla^2 \mathbf{u} + \mathrm{R}(\mathbf{u}) \tag{7.4}$$

**D** *is the diffusion matrix and* R *the function describing sources and sinks through reactions.*

is solved [Pea93; GS83]. These systems are an important theoretical tool for the study of pattern formation as well as morphogenesis. They have also been investigated in terms of thermodynamics (entropy production)[Mah+04; MYS05]. A recent extension of reaction diffusion systems by hydrodynamics has given interesting results, showing competition between different dissipative structures [BB15]. Simple reaction diffusion-like models were also used to investigate how individuals form and how their states (i.e. their existence) develop under perturbation (ontogenesis) [AGB15].

Figure 7.6 shows how reaction networks and pathway analysis might be applied to better understand reaction diffusion systems and eventually build more complex ones. By using clustering on concentrations, the spots can be separated from their background and the concentrations and rates in the different areas integrated for constructing a simplified network model (Fig. 7.6 **(top-right)**). Simultaneously, the grid of the simulation defines a much larger network (Fig. 7.6 **(bottom)**). The difference in size of the networks implies that each pathway in the simple network corresponds to many pathways in the bigger network. Investigating this relationship might help to understand the functioning of reaction diffusion systems. Additionally, it might help to develop clustering techniques that work for pathways in the larger network without looking at the smaller network first.

### *Agent Based Models*

The last example for possible applications of thermodynamic constrained artificial chemistries are agent based models. Agent based models are an important modelling technique in ecology, economics, traffic engineering, social science, and other disciplines. In evolutionary biology and ecology, often the slightly different term "individual based models" is used. They describe ecosystem with evolving behaviour and allow for the study of species diversity and speciation [DM05; Gra+09; KG12].

Related to virtual ecosystems of individuals in ecology and artificial life is the field of evolutionary robotics [Bed03; Sim94; Pra03]. It focuses more on evolving physical machines, but relates closely to artificial life when concepts for self-replicating robots are explored [Eib14]. If such robots are electromechanical, they probably do

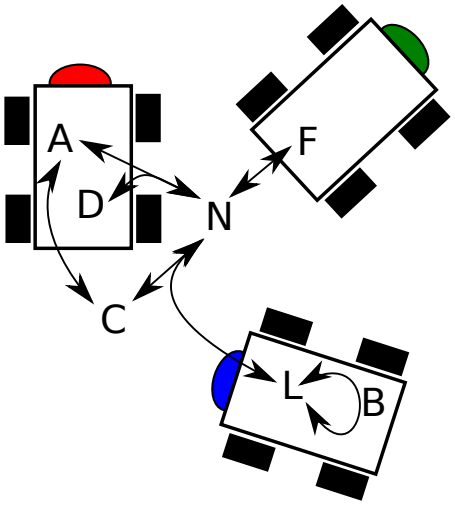



Figure 7.7: **Using artificial chemistry as physical basis for agent-based models.** Basing agents functioning on chemical thermodynamics allows for the usage of reaction networks to analyse their energetic interactions.

not contain any natural notion of chemistry, yet their self-replication process might be described with formal reaction equations.[13]

[13] *See the work of Von Neumann and Burks [VNB66].*

In these fields, the application of chemical reaction networks of agent interaction (Fig. 7.7) would directly integrate energy and matter flow into the models. For example, the interconnection between metabolic function, (co-)evolution, and behaviour [Egb13] can then be analysed easily. In a broader context, this allows for an integration of basic physics (thermodynamics) with the increasingly important field of information theory [WD13; Fri10].

# A

# APPENDIX ALGORITHMS

## A.1 SOLVING REACTION EQUATIONS

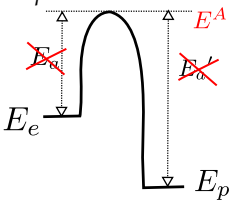
*__Bold__ symbols in formulas are used to indicate matrices and vectors.*

Solving the reaction equation of artificial chemistries as they where created for Chapter 5 comes with certain challenges. As the system does not exchange matter with its environment, but only cycles the mass it has, small errors could accumulate over time and violate mass conservation. Furthermore, for reactions close to equilibrium the forward and backward reaction rates are big compared to the effective rate:

$$v_i^f \gg |v_i^f - v_i^b| \qquad v_i^b \gg |v_i^f - v_i^b|. \tag{A.1}$$

Thus, we have to use the following reformulation that expresses the effective rates $v_i = v_i^f - v_i^b$ in terms of the standard enthalpies of formation $\boldsymbol{\mu}^0$ and the activation energies $\mathbf{E^A}$.

*Activation energy is defined in absolute values in this chapter!*




We start reformulating the reaction equation (Eq. 2.8)

$$\frac{d}{dt}\mathbf{x} = \mathbf{N} \cdot \boldsymbol{v}(\mathbf{x}) = \mathbf{N} \cdot \left(\boldsymbol{v}^f(\mathbf{x}) - \boldsymbol{v}^b(\mathbf{x})\right), \tag{A.2}$$

with $\mathbf{N} = \mathbf{N}^{out} - \mathbf{N}^{in}$ being the stoichiometric matrix given as difference of the stoichiometric matrix of the products and the educts.

Inserting mass action kinetics (Eq. 2.9)

$$v_i^f(\mathbf{x}) = k_i^f \prod_j x_j{}^{N_{jk}^{in}} \qquad v_i^b(\mathbf{x}) = k_i^b \prod_j x_j{}^{N_{jk}^{out}} \tag{A.3}$$

gives

$$\frac{d}{dt}x_i = \sum_l N_{il} \cdot \left( k_l^f \prod_j x_j{}^{N_{jl}^{in}} - k_l^b \prod_j x_j{}^{N_{jl}^{out}} \right). \tag{A.4}$$

Inserting the Arrhenius equation to calculate the reaction rate constants as defined for our artificial chemistries (Eq. 2.11)

$$\begin{aligned} k_i^f &= \exp\left(-\beta\left(E_i^A - \sum_j N_{ji}^{in}\,\mu_j^0\right)\right) \\ k_i^b &= \exp\left(-\beta\left(E_i^A - \sum_j N_{ji}^{out}\,\mu_j^0\right)\right) \end{aligned} \tag{A.5}$$

and further reformulation leads to a reaction equation in terms of thermodynamic quantities:

$$\frac{d}{dt}x_i = \sum_l N_{il}\, e^{-\beta E_l^A} \cdot \left( e^{\beta \sum_j N_{jl}^{in}\mu_j^0} \prod_j x_j{}^{N_{jl}^{in}} - e^{\beta \sum_j N_{jl}^{out}\mu_j^0} \prod_j x_j{}^{N_{jl}^{out}} \right).$$

*The values $f_{il}$, $g_l^{in}$ and $g_l^{in}$ are only depending on thermodynamic quantities that are constant. They only have to be calculated once per network.*

The abbreviations

$$\begin{aligned} f_{il} &:= N_{il}\, e^{-\beta E_l^A} \\ g_l^{in} &:= e^{\beta \sum_j N_{jl}^{in} \mu_j^0} \\ g_l^{out} &:= e^{\beta \sum_j N_{jl}^{out} \mu_j^0} \end{aligned} \tag{A.6}$$

allow to simplify the rewritten differential equation.

$$\frac{d}{dt} x_i = \dot{x_i} = \sum_l f_{il} \cdot \left( g_l^{in} \prod_j {x_j}^{N_{jl}^{in}} - g_l^{out} \prod_j {x_j}^{N_{jl}^{out}} \right) \tag{A.7}$$

The factorisation of $f_{il}$ in Eq. A.7 weakens the numeric problem mentioned above. Besides the right-hand side of the differential equation we also have to calculate the Jacobi-matrix, so that we can use an implicit solver for integration of the differential equation. The Jacobi-matrix is defined as the partial derivatives of the right-hand side of the differential equation.

$$\begin{aligned} M_{ik} &= \frac{\partial \dot{x_i}}{\partial x_k} \\ &= \sum_l f_{il} \cdot \left( g_l^{in} \cdot \frac{\partial}{\partial x_k} \prod_j {x_j}^{N_{jl}^{in}} - g_l^{out} \cdot \frac{\partial}{\partial x_k} \prod_j {x_j}^{N_{jl}^{out}} \right) \\ &= \sum_l f_{il} \cdot \left( g_l^{in} \cdot {x_k}^{(N_{kl}^{in}-1)} \prod_{j \neq k} {x_j}^{N_{jl}^{in}} \right. \\ &\qquad \left. - g_l^{out} \cdot {x_k}^{(N_{kl}^{out}-1)} \prod_{j \neq k} {x_j}^{N_{jl}^{out}} \right) \end{aligned} \tag{A.8}$$

The ODE solver used for the artificial ecosystems in Chapter 5 uses Eq. A.7 and Eq. A.8. It is written using the odeint library available through C++'s boost [AM11]. The used algorithm is called Rosenbrock 4 [Pre+07]. The program can be found in the digital appendix (Appendix C) and in Fischer [Fis16b].

## A.2 CALCULATION OF CHEMICAL POTENTIALS

To calculate thermodynamic parameters in atmospheric chemistry often the NASA polynomials from McBride, Gordon, and Reno [MGR93] are used [Ven+12; Gra+14]. For many common species this list contains coefficients $a_1 \dots a_7$ that give temperature dependent estimates for the standard enthalpy of formation $h^0(T)$

$$\frac{h^0(T)}{RT} = a_1 + a_2\frac{T}{2} + a_3\frac{T^2}{3} + a_4\frac{T^3}{4} + a_5\frac{T^4}{5} + \frac{a_6}{T} \tag{A.9}$$

and the standard entropy $s^0(T)$

$$\frac{s^0(T)}{R} = a_1 \log T + a_2 T + a_3\frac{T^2}{2} + a_4\frac{T^3}{3} + a_5\frac{T^4}{4} + a_7. \tag{A.10}$$

By assuming all components of the simulated atmosphere behave like an ideal gas, we can calculate the chemical potentials for all the components in the different layers of the atmosphere using the equation (see Eq. (B.3) in [Ven+12])

$$\mu = h^0 - Ts^0 + RT \log\left(\frac{p}{p_0}\right). \tag{A.11}$$

Here, $p_0$ is the standard pressure and $p$ the partial pressure of the chemical species. The latter is calculated from the number density $N/V$ by using the ideal gas law ($pV = NRT$).

# B

# APPENDIX NETWORKS AND PATHWAYS

## B.1 ATMOSPHERIC CHEMISTRY MODELS

### B.1.1 *Model Structure and Application*

For the work in Chapter 4, two freely available[1] atmospheric chemistry models from the group of James Kasting were analysed. One model describes the modern oxidising atmosphere with high $O_2$ concentration [KHP85; PK02]. The other model describes a reducing, pre photosynthetic atmosphere dominated by $CO_2$ and $H_2$ [PBK01; KKS05; Seg+07]. Both models are one dimensional photochemical models that describe the concentrations of the atmospheric components by a set of homogeneous layers connected by vertical mixing and influenced by parametrisations of fluxes and further processes. These models play an important role in connecting geological observations with probable atmospheric conditions. The Archean model for example, was used to show that mass independent fractionation in sulphur isotopes found in sediments is in fact an indicator for extremely low oxygen concentrations in the past atmosphere [PK02].

[1] *Obtained in August 2013 from* `http://vpl.astro.washington.edu/sci/AntiModels/models09.html`

As parts of the structure of the models are almost 30 years old and computational resources were more limited then, these models treat chemical species differently depending on their lifetime and concentration. Species with a long lifetime and a high concentration are kept at the same constant fraction in all layers. Short lived species are treated as being in photochemical equilibrium in the individual layers. For the remaining species combined vertical transport and continuity equations are solved. Furthermore, a constant deposition flux of some species in the lowest layer and simple parametrisations of aerosol chemistries and rainout are integrated in the model equations.

For both models the concentrations of the most abundant species of the investigated steady state is given as follows: For the first model (modern atmosphere) the dominating species and their respective number fraction are $N_2$ (78.5%), $O_2$ (20.9%), $H_2O$ (0.5%), $CO_2$ (353 ppm), and $CH_4$ (1.5 ppm). For the second model (Archean atmosphere) these are $N_2$ (97.2%), $CO_2$ (2.5%), $H_2O$ (0.2%), $CH_4$ (846 ppm), CO (75 ppm), and $H_2$ (58 ppm).

### B.1.2 *Preprocessing for Metaanalysis*

Before we perform our analysis, we condense the model output into a box model representation by vertically integrating the reaction rates. This implies that some spatial resolution is lost. For example, we can no longer distinguish the flux from Earth's surface and the fluxes by parameterised processes in the different columns of the model.

From the source code of the models we are extracting a list of all the reactions that are simulated by the respective models. From the model output we are getting vertically integrated reaction rates $v_i$, concentration profiles $c_{jk}$, temperature profiles $T_k$ and the position of the layer's upper boundaries $x_k$. Please note, for clarity specific letters are used for indices in this section. The letter $i$ selects the chemical reaction, $j$ the chemical species and $k$ the layer of the atmosphere. Indices start with one, except $x_0$ which is defined to be the lower boundary of the first atmospheric layer.

In an additional step, before both reaction models are analysed we replace the few reactions that are modelled reversibly by their effective reactions. This means removing the reaction of the reverse pair that has the lower rate and setting the rate of the other reaction to the effective reaction rate. This is in accordance with our understanding of the non-reversibly modelled reactions just representing the effective reactions of reversible reactions that are far from equilibrium.

[2] *This quantity is sometimes called* column density.

The concentration $c_j$ in integrated representation[2] is calculated by the formula

$$c_j = \sum_{k=1}^{k_{max}} c_{jk} \left( x_k - x_{k-1} \right) . \tag{B.1}$$

This is analogous to the formula used by the models to calculate the integrated reaction rates $v_i$. A graphic overview of inflow and outflow as well as concentration in the integrated representation of both models is given in Fig. 4.7. The full list of reactions for the model of the modern atmosphere and the integrated steady state flux is given in Table B.1. The reaction equations and their steady state rates for the Archean atmosphere model can be seen in Table B.2.

Table B.1: **Reaction network of the model of the modern atmosphere.** Reaction equations have been extracted from the model source code. Vertically integrated rates are taken from the model output.

| | | | rates ($1/cm^2$) | |
|---|---|---|---|---|
| $H_2O + O(_1D)$ | $\rightarrow$ | $2OH$ | 2.522e+11 | (1) |
| $H_2 + O(_1D)$ | $\rightarrow$ | $OH + H$ | 7.729e+08 | (2) |
| $H_2 + O$ | $\rightarrow$ | $OH + H$ | 5.072e+06 | (3) |
| $H_2 + OH$ | $\rightarrow$ | $H_2O + H$ | 5.931e+10 | (4) |
| $H + O_3$ | $\rightarrow$ | $OH + O_2$ | 4.688e+11 | (5) |
| $H + O_2 + M$ | $\rightarrow$ | $HO_2 + M$ | 3.251e+12 | (6) |
| $H + HO_2$ | $\rightarrow$ | $H_2 + O_2$ | 9.605e+07 | (7) |
| $H + HO_2$ | $\rightarrow$ | $H_2O + O$ | 2.401e+07 | (8) |
| $H + HO_2$ | $\rightarrow$ | $2OH$ | 1.081e+09 | (9) |
| $OH + O$ | $\rightarrow$ | $H + O_2$ | 3.144e+12 | (10) |
| $OH + HO_2$ | $\rightarrow$ | $H_2O + O_2$ | 1.166e+11 | (11) |
| $OH + O_3$ | $\rightarrow$ | $HO_2 + O_2$ | 9.055e+11 | (12) |
| $HO_2 + O$ | $\rightarrow$ | $OH + O_2$ | 2.914e+12 | (13) |
| $HO_2 + O_3$ | $\rightarrow$ | $OH + 2O_2$ | 2.887e+11 | (14) |
| $2HO_2$ | $\rightarrow$ | $H_2O_2 + O_2$ | 1.239e+11 | (15) |
| $H_2O_2 + OH$ | $\rightarrow$ | $HO_2 + H_2O$ | 4.492e+10 | (16) |

(continued on next page)

Table B.1: (continued - modern atm. reaction network)

| | | | rates ($1/cm^2$) | |
|---|---|---|---|---|
| $2O + M$ | $\rightarrow$ | $O_2 + M$ | 8.853e+08 | (17) |
| $O + O_2 + M$ | $\rightarrow$ | $O_3 + M$ | 3.405e+15 | (18) |
| $O + O_3$ | $\rightarrow$ | $2O_2$ | 1.326e+12 | (19) |
| $2OH$ | $\rightarrow$ | $H_2O + O$ | 1.381e+09 | (20) |
| $O(_1D) + N_2$ | $\rightarrow$ | $O + N_2$ | 6.576e+14 | (21) |
| $O(_1D) + O_2$ | $\rightarrow$ | $O + O_2$ | 2.646e+14 | (22) |
| $O_2 + hv$ | $\rightarrow$ | $O + O(_1D)$ | 3.076e+07 | (23) |
| $O_2 + hv$ | $\rightarrow$ | $2O$ | 8.934e+12 | (24) |
| $H_2O + hv$ | $\rightarrow$ | $H + OH$ | 1.374e+10 | (25) |
| $O_3 + hv$ | $\rightarrow$ | $O_2 + O(_1D)$ | 9.225e+14 | (26) |
| $O_3 + hv$ | $\rightarrow$ | $O_2 + O$ | 2.445e+15 | (27) |
| $H_2O_2 + hv$ | $\rightarrow$ | $2OH$ | 5.457e+10 | (28) |
| $CO_2 + hv$ | $\rightarrow$ | $CO + O$ | 1.395e+09 | (29) |
| $CO + OH$ | $\rightarrow$ | $CO_2 + H$ | 4.646e+11 | (30) |
| $CO + O + M$ | $\rightarrow$ | $CO_2 + M$ | 2.552e+05 | (31) |
| $H + CO + M$ | $\rightarrow$ | $HCO + M$ | 7.931e+02 | (32) |
| $H + HCO$ | $\rightarrow$ | $H_2 + CO$ | 8.27e-01 | (33) |
| $2HCO$ | $\rightarrow$ | $H_2CO + CO$ | 0e+00 | (34) |
| $OH + HCO$ | $\rightarrow$ | $H_2O + CO$ | 2.217e+01 | (35) |
| $O + HCO$ | $\rightarrow$ | $H + CO_2$ | 3.602e+03 | (36) |
| $O + HCO$ | $\rightarrow$ | $OH + CO$ | 3.602e+03 | (37) |
| $H_2CO + hv$ | $\rightarrow$ | $H_2 + CO$ | 6.789e+10 | (38) |
| $H_2CO + hv$ | $\rightarrow$ | $HCO + H$ | 3.931e+10 | (39) |
| $HCO + hv$ | $\rightarrow$ | $H + CO$ | 3.161e+02 | (40) |
| $H_2CO + H$ | $\rightarrow$ | $H_2 + HCO$ | 1.558e+04 | (41) |
| $CO_2 + hv$ | $\rightarrow$ | $CO + O(_1D)$ | 2.496e+05 | (42) |
| $2H + M$ | $\rightarrow$ | $H_2 + M$ | 4.531e+02 | (43) |
| $HCO + O_2$ | $\rightarrow$ | $HO_2 + CO$ | 9.356e+10 | (44) |
| $H_2CO + OH$ | $\rightarrow$ | $H_2O + HCO$ | 5.374e+10 | (45) |
| $H + OH + M$ | $\rightarrow$ | $H_2O + M$ | 2.445e+05 | (46) |
| $2OH + M$ | $\rightarrow$ | $H_2O_2 + M$ | 4.712e+07 | (47) |
| $H_2CO + O$ | $\rightarrow$ | $HCO + OH$ | 3.343e+08 | (48) |
| $H_2O_2 + O$ | $\rightarrow$ | $OH + HO_2$ | 9.42e+06 | (49) |
| $HO_2 + hv$ | $\rightarrow$ | $OH + O$ | 5.803e+09 | (50) |
| $CH_4 + hv$ | $\rightarrow$ | $_1CH_2 + H_2$ | 2.635e+04 | (51) |
| $CH_3OOH + hv$ | $\rightarrow$ | $H_3CO + OH$ | 2.009e+10 | (52) |
| $N_2O + hv$ | $\rightarrow$ | $N_2 + O$ | 9.737e+08 | (53) |
| $HNO_2 + hv$ | $\rightarrow$ | $NO + OH$ | 7.812e+09 | (54) |
| $HNO_3 + hv$ | $\rightarrow$ | $NO_2 + OH$ | 2.446e+10 | (55) |
| $NO + hv$ | $\rightarrow$ | $N + O$ | 1.863e+08 | (56) |
| $NO_2 + hv$ | $\rightarrow$ | $NO + O$ | 3.123e+13 | (57) |
| $CH_4 + OH$ | $\rightarrow$ | $CH_3 + H_2O$ | 1.733e+11 | (58) |
| $CH_4 + O(_1D)$ | $\rightarrow$ | $CH_3 + OH$ | 1.769e+09 | (59) |
| $CH_4 + O(_1D)$ | $\rightarrow$ | $H_2CO + H_2$ | 1.769e+08 | (60) |
| $_1CH_2 + CH_4$ | $\rightarrow$ | $2CH_3$ | 7.472e-04 | (61) |
| $_1CH_2 + O_2$ | $\rightarrow$ | $H_2CO + O$ | 2.479e+04 | (62) |
| $_1CH_2 + N_2$ | $\rightarrow$ | $_3CH_2 + N_2$ | 1.554e+03 | (63) |
| $_3CH_2 + H_2$ | $\rightarrow$ | $CH_3 + H$ | 3.986e-05 | (64) |
| $_3CH_2 + CH_4$ | $\rightarrow$ | $2CH_3$ | 2.465e-05 | (65) |
| $_3CH_2 + O_2$ | $\rightarrow$ | $H_2CO + O$ | 1.554e+03 | (66) |
| $CH_3 + O_2 + M$ | $\rightarrow$ | $CH_3O_2 + M$ | 1.762e+11 | (67) |

(continued on next page)

Table B.1: (continued - modern atm. reaction network)

| | | | rates ($1/cm^2$) | |
|---|---|---|---|---|
| $CH_3 + OH$ | $\rightarrow$ | $H_2CO + H_2$ | 1.514e+04 | (68) |
| $CH_3 + O$ | $\rightarrow$ | $H_2CO + H$ | 2.231e+06 | (69) |
| $CH_3 + O_3$ | $\rightarrow$ | $H_2CO + HO_2$ | 3.439e+06 | (70) |
| $CH_3O_2 + HO_2$ | $\rightarrow$ | $CH_3OOH + O_2$ | 1.177e+11 | (71) |
| $2CH_3O_2$ | $\rightarrow$ | $2H_3CO + O_2$ | 4.028e+09 | (72) |
| $CH_3O_2 + NO$ | $\rightarrow$ | $H_3CO + NO_2$ | 1.182e+11 | (73) |
| $H_3CO + O_2$ | $\rightarrow$ | $H_2CO + HO_2$ | 1.65e+11 | (74) |
| $H_3CO + O$ | $\rightarrow$ | $H_2CO + OH$ | 1.918e+03 | (75) |
| $H_3CO + OH$ | $\rightarrow$ | $H_2CO + H_2O$ | 1.241e+05 | (76) |
| $N_2O + O(_1D)$ | $\rightarrow$ | $2NO$ | 1.048e+08 | (77) |
| $N_2O + O(_1D)$ | $\rightarrow$ | $N_2 + O_2$ | 7.667e+07 | (78) |
| $N + O_2$ | $\rightarrow$ | $NO + O$ | 1.602e+08 | (79) |
| $N + O_3$ | $\rightarrow$ | $NO + O_2$ | 2.179e+04 | (80) |
| $N + OH$ | $\rightarrow$ | $NO + H$ | 3.887e+06 | (81) |
| $N + NO$ | $\rightarrow$ | $N_2 + O$ | 2.222e+07 | (82) |
| $NO + O_3$ | $\rightarrow$ | $NO_2 + O_2$ | 3.318e+13 | (83) |
| $NO + O + M$ | $\rightarrow$ | $NO_2 + M$ | 8.275e+09 | (84) |
| $NO + HO_2$ | $\rightarrow$ | $NO_2 + OH$ | 7.795e+11 | (85) |
| $NO + OH + M$ | $\rightarrow$ | $HNO_2 + M$ | 7.813e+09 | (86) |
| $NO_2 + O$ | $\rightarrow$ | $NO + O_2$ | 3.6e+12 | (87) |
| $NO_2 + OH + M$ | $\rightarrow$ | $HNO_3 + M$ | 3.917e+10 | (88) |
| $NO_2 + H$ | $\rightarrow$ | $NO + OH$ | 3.406e+08 | (89) |
| $HNO_3 + OH$ | $\rightarrow$ | $H_2O + NO_3$ | 9.514e+09 | (90) |
| $HO_2 + NO_2 + M$ | $\rightarrow$ | $HO_2NO_2 + M$ | 1.098e+11 | (91) |
| $HO_2NO_2 + OH$ | $\rightarrow$ | $NO_2 + H_2O + O_2$ | 1.005e+10 | (92) |
| $HO_2NO_2 + O$ | $\rightarrow$ | $NO_2 + OH + O_2$ | 2.846e+05 | (93) |
| $HO_2NO_2 + M$ | $\rightarrow$ | $HO_2 + NO_2 + M$ | 9.247e+10 | (94) |
| $HO_2NO_2 + hv$ | $\rightarrow$ | $HO_2 + NO_2$ | 7.115e+09 | (95) |
| $CH_3OOH + OH$ | $\rightarrow$ | $CH_3O_2 + H_2O$ | 8.657e+10 | (96) |
| $CH_3O_2 + OH$ | $\rightarrow$ | $H_3CO + HO_2$ | 1.868e+10 | (97) |
| $O_3 + NO_2$ | $\rightarrow$ | $O_2 + NO_3$ | 4.197e+10 | (98) |
| $NO_2 + NO_3$ | $\rightarrow$ | $NO + NO_2 + O_2$ | 5.149e+05 | (99) |
| $O + NO_3$ | $\rightarrow$ | $O_2 + NO_2$ | 6.84e+08 | (100) |
| $CH_3Cl + hv$ | $\rightarrow$ | $CH_3 + Cl$ | 8.041e+03 | (101) |
| $NO + NO_3$ | $\rightarrow$ | $2NO_2$ | 3.725e+10 | (102) |
| $OH + NO_3$ | $\rightarrow$ | $HO_2 + NO_2$ | 1.211e+08 | (103) |
| $CH_3Cl + OH$ | $\rightarrow$ | $Cl + H_2O + {}_1CH_2$ | 4.395e+08 | (104) |
| $Cl + O_3$ | $\rightarrow$ | $ClO + O_2$ | 1.174e+12 | (105) |
| $Cl + H_2$ | $\rightarrow$ | $HCl + H$ | 5.198e+07 | (106) |
| $Cl + CH_4$ | $\rightarrow$ | $HCl + CH_3$ | 1.18e+09 | (107) |
| $Cl + CH_3Cl$ | $\rightarrow$ | $Cl + HCl + {}_1CH_2$ | 1.458e+06 | (108) |
| $Cl + H_2CO$ | $\rightarrow$ | $HCl + HCO$ | 1.721e+08 | (109) |
| $Cl + H_2O_2$ | $\rightarrow$ | $HCl + HO_2$ | 2.631e+06 | (110) |
| $Cl + HO_2$ | $\rightarrow$ | $HCl + O_2$ | 1.518e+08 | (111) |
| $Cl + HO_2$ | $\rightarrow$ | $ClO + OH$ | 3.118e+07 | (112) |
| $Cl + ClONO_2$ | $\rightarrow$ | $2Cl + NO_2 + O$ | 1.936e+07 | (113) |
| $Cl + NO + M$ | $\rightarrow$ | $NOCl + M$ | 2.803e+06 | (114) |
| $Cl + NO_2 + M$ | $\rightarrow$ | $ClONO + M$ | 5.124e+07 | (115) |
| $Cl + NOCl$ | $\rightarrow$ | $NO + Cl_2$ | 6.585e+03 | (116) |
| $Cl + O_2 + M$ | $\rightarrow$ | $ClO_2 + M$ | 1.232e+14 | (117) |
| $Cl + ClO_2$ | $\rightarrow$ | $Cl_2 + O_2$ | 7.231e+02 | (118) |

(continued on next page)

Table B.1: (continued - modern atm. reaction network)

| | | | rates ($1/cm^2$) | |
|---|---|---|---|---|
| $Cl + ClO_2$ | $\rightarrow$ | $2ClO$ | 3.772e+01 | (119) |
| $ClO + O$ | $\rightarrow$ | $Cl + O_2$ | 4.392e+11 | (120) |
| $ClO + NO$ | $\rightarrow$ | $Cl + NO_2$ | 7.128e+11 | (121) |
| $ClO + NO_2 + M$ | $\rightarrow$ | $ClONO_2 + M$ | 1.091e+10 | (122) |
| $ClO + HO_2$ | $\rightarrow$ | $HOCl + O_2$ | 5.282e+09 | (123) |
| $ClO + OH$ | $\rightarrow$ | $Cl + HO_2$ | 5.908e+09 | (124) |
| $HCl + OH$ | $\rightarrow$ | $Cl + H_2O$ | 1.122e+09 | (125) |
| $HOCl + OH$ | $\rightarrow$ | $ClO + H_2O$ | 7.694e+07 | (126) |
| $ClONO_2 + OH$ | $\rightarrow$ | $Cl + HO_2 + NO_2$ | 1.399e+08 | (127) |
| $HCl + O$ | $\rightarrow$ | $Cl + OH$ | 1.2e+06 | (128) |
| $HOCl + O$ | $\rightarrow$ | $ClO + OH$ | 8.088e+06 | (129) |
| $ClONO_2 + O$ | $\rightarrow$ | $Cl + O_2 + NO_2$ | 3.647e+08 | (130) |
| $Cl_2 + OH$ | $\rightarrow$ | $HOCl + Cl$ | 5.608e+02 | (131) |
| $Cl_2 + hv$ | $\rightarrow$ | $2Cl$ | 1.662e+06 | (132) |
| $ClO_2 + hv$ | $\rightarrow$ | $ClO + O$ | 1.864e+04 | (133) |
| $HCl + hv$ | $\rightarrow$ | $H + Cl$ | 5.364e+06 | (134) |
| $HOCl + hv$ | $\rightarrow$ | $OH + Cl$ | 5.192e+09 | (135) |
| $NOCl + hv$ | $\rightarrow$ | $Cl + NO$ | 2.797e+06 | (136) |
| $ClONO + hv$ | $\rightarrow$ | $Cl + NO_2$ | 5.646e+07 | (137) |
| $ClONO_2 + hv$ | $\rightarrow$ | $Cl + NO_3$ | 1.037e+10 | (138) |
| $ClO_2 + hv$ | $\rightarrow$ | $Cl + O_2$ | 3.235e+13 | (139) |
| $HO_2 + NO_3$ | $\rightarrow$ | $HNO_3 + O_2$ | 2.319e+08 | (140) |
| $2ClO + M$ | $\rightarrow$ | $Cl_2O_2 + M$ | 5.743e+06 | (141) |
| $Cl_2O_2 + hv$ | $\rightarrow$ | $ClO_2 + Cl$ | 2.749e+06 | (142) |
| $Cl_2O_2 + M$ | $\rightarrow$ | $2ClO + M$ | 2.994e+06 | (143) |
| $ClO_2 + M$ | $\rightarrow$ | $Cl + O_2 + M$ | 9.084e+13 | (144) |
| $Cl + NO_3$ | $\rightarrow$ | $ClO + NO_2$ | 4.998e+05 | (145) |
| $Cl + HOCl$ | $\rightarrow$ | $Cl_2 + OH$ | 1.655e+06 | (146) |
| $ClO + NO_3$ | $\rightarrow$ | $ClONO + O_2$ | 5.217e+06 | (147) |
| $ClONO + OH$ | $\rightarrow$ | $HOCl + NO_2$ | 2.678e+03 | (148) |
| $ClO_2 + O$ | $\rightarrow$ | $ClO + O_2$ | 1.583e+03 | (149) |
| $NO_2 + O + M$ | $\rightarrow$ | $NO_3 + M$ | 8.082e+09 | (150) |
| $NO_3 + hv$ | $\rightarrow$ | $NO_2 + O$ | 3.14e+10 | (151) |
| $NO_3 + NO_2 + M$ | $\rightarrow$ | $N_2O_5 + M$ | 2.065e+09 | (152) |
| $N_2O_5 + hv$ | $\rightarrow$ | $NO_2 + NO_3$ | 1.788e+09 | (153) |
| $N_2O_5 + M$ | $\rightarrow$ | $NO_2 + NO_3 + M$ | 2.951e+07 | (154) |
| $N_2O_5 + H_2O$ | $\rightarrow$ | $2HNO_3$ | 2.474e+08 | (155) |
| $SO + hv$ | $\rightarrow$ | $S + O$ | 0e+00 | (156) |
| $SO_2 + hv$ | $\rightarrow$ | $SO + O$ | 8.386e+04 | (157) |
| $H_2S + hv$ | $\rightarrow$ | $HS + H$ | 5.96e+01 | (158) |
| $SO + O_2$ | $\rightarrow$ | $O + SO_2$ | 3.48e+08 | (159) |
| $SO + HO_2$ | $\rightarrow$ | $SO_2 + OH$ | 1.492e+04 | (160) |
| $SO + O$ | $\rightarrow$ | $SO_2$ | 9.906e+00 | (161) |
| $SO + OH$ | $\rightarrow$ | $SO_2 + H$ | 1.721e+03 | (162) |
| $SO_2 + OH$ | $\rightarrow$ | $HSO_3$ | 2.71e+09 | (163) |
| $SO_2 + O$ | $\rightarrow$ | $SO_3$ | 5.666e+04 | (164) |
| $SO_3 + H_2O$ | $\rightarrow$ | $H_2SO_4$ | 2.71e+09 | (165) |
| $HSO_3 + O_2$ | $\rightarrow$ | $HO_2 + SO_3$ | 2.71e+09 | (166) |
| $HSO_3 + OH$ | $\rightarrow$ | $H_2O + SO_3$ | 4.901e-02 | (167) |
| $HSO_3 + H$ | $\rightarrow$ | $H_2 + SO_3$ | 3.344e-07 | (168) |
| $HSO_3 + O$ | $\rightarrow$ | $OH + SO_3$ | 1.505e-03 | (169) |

(continued on next page)

Table B.1: (continued - modern atm. reaction network)

| | | | rates ($1/cm^2$) | |
|---|---|---|---|---|
| $H_2S + OH$ | $\rightarrow$ | $H_2O + HS$ | 3.499e+08 | (170) |
| $H_2S + H$ | $\rightarrow$ | $H_2 + HS$ | 2.704e+00 | (171) |
| $H_2S + O$ | $\rightarrow$ | $OH + HS$ | 1.779e+03 | (172) |
| $HS + O$ | $\rightarrow$ | $H + SO$ | 9.816e+01 | (173) |
| $HS + O_2$ | $\rightarrow$ | $OH + SO$ | 1.31e+08 | (174) |
| $HS + HO_2$ | $\rightarrow$ | $H_2S + O_2$ | 4.637e+05 | (175) |
| $2HS$ | $\rightarrow$ | $H_2S + S$ | 1.14e-01 | (176) |
| $HS + HCO$ | $\rightarrow$ | $H_2S + CO$ | 1.965e-05 | (177) |
| $HS + H$ | $\rightarrow$ | $H_2 + S$ | 1.387e-04 | (178) |
| $HS + S$ | $\rightarrow$ | $H + S_2$ | 9.387e-14 | (179) |
| $S + O_2$ | $\rightarrow$ | $SO + O$ | 3.594e+04 | (180) |
| $S + OH$ | $\rightarrow$ | $SO + H$ | 1.463e-03 | (181) |
| $SO_2 + hv$ | $\rightarrow$ | $S + O_2$ | 3.594e+04 | (182) |
| $S + HO_2$ | $\rightarrow$ | $HS + O_2$ | 1.638e-04 | (183) |
| $S + HO_2$ | $\rightarrow$ | $SO + OH$ | 1.638e-04 | (184) |
| $HS + H_2CO$ | $\rightarrow$ | $H_2S + HCO$ | 2.049e+05 | (185) |
| $SO_2 + hv$ | $\rightarrow$ | $_1SO_2$ | 2.098e+11 | (186) |
| $SO_2 + hv$ | $\rightarrow$ | $_3SO_2$ | 1.504e+09 | (187) |
| $H_2SO_4 + hv$ | $\rightarrow$ | $SO_2 + 2OH$ | 4.135e+02 | (188) |
| $SO_3 + hv$ | $\rightarrow$ | $SO_2 + O$ | 0e+00 | (189) |
| $_1SO_2 + M$ | $\rightarrow$ | $_3SO_2 + M$ | 1.907e+10 | (190) |
| $_1SO_2 + M$ | $\rightarrow$ | $SO_2 + M$ | 1.907e+11 | (191) |
| $_1SO_2 + hv$ | $\rightarrow$ | $_3SO_2 + hv$ | 2.626e+06 | (192) |
| $_1SO_2 + hv$ | $\rightarrow$ | $SO_2 + hv$ | 3.851e+07 | (193) |
| $_1SO_2 + O_2$ | $\rightarrow$ | $SO_3 + O$ | 4.004e+05 | (194) |
| $_1SO_2 + SO_2$ | $\rightarrow$ | $SO_3 + SO$ | 9.696e+00 | (195) |
| $_3SO_2 + M$ | $\rightarrow$ | $SO_2 + M$ | 2.056e+10 | (196) |
| $_3SO_2 + hv$ | $\rightarrow$ | $SO_2 + hv$ | 1.419e+07 | (197) |
| $_3SO_2 + SO_2$ | $\rightarrow$ | $SO_3 + SO$ | 1.218e+00 | (198) |
| $SO + NO_2$ | $\rightarrow$ | $SO_2 + NO$ | 4.348e+04 | (199) |
| $SO + O_3$ | $\rightarrow$ | $SO_2 + O_2$ | 2.227e+05 | (200) |
| $SO_2 + HO_2$ | $\rightarrow$ | $SO_3 + OH$ | 0e+00 | (201) |
| $HS + O_3$ | $\rightarrow$ | $HSO + O_2$ | 2.152e+08 | (202) |
| $HS + NO_2$ | $\rightarrow$ | $HSO + NO$ | 3.013e+06 | (203) |
| $S + O_3$ | $\rightarrow$ | $SO + O_2$ | 2.108e+00 | (204) |
| $2SO$ | $\rightarrow$ | $SO_2 + S$ | 6.594e-07 | (205) |
| $SO_3 + SO$ | $\rightarrow$ | $2SO_2$ | 6.682e-06 | (206) |
| $S + CO_2$ | $\rightarrow$ | $SO + CO$ | 2.641e-07 | (207) |
| $SO + HO_2$ | $\rightarrow$ | $HSO + O_2$ | 0e+00 | (208) |
| $SO + HCO$ | $\rightarrow$ | $HSO + CO$ | 1.111e-07 | (209) |
| $H + SO$ | $\rightarrow$ | $HSO$ | 4.655e-04 | (210) |
| $HSO + hv$ | $\rightarrow$ | $HS + O$ | 1.753e+00 | (211) |
| $HSO + OH$ | $\rightarrow$ | $H_2O + SO$ | 2.167e+08 | (212) |
| $HSO + H$ | $\rightarrow$ | $HS + OH$ | 8.633e+00 | (213) |
| $HSO + H$ | $\rightarrow$ | $H_2 + SO$ | 7.674e-01 | (214) |
| $HSO + HS$ | $\rightarrow$ | $H_2S + SO$ | 8.49e+01 | (215) |
| $HSO + O$ | $\rightarrow$ | $OH + SO$ | 4.581e+05 | (216) |
| $HSO + S$ | $\rightarrow$ | $HS + SO$ | 6.294e-10 | (217) |
| $N_2 + O(_1D)$ | $\rightarrow$ | $N_2O$ | 4.031e+06 | (218) |
| $N_2O + H + O_2$ | $\rightarrow$ | $2NO + OH$ | 6.226e-18 | (219) |
| $N_2O + NO$ | $\rightarrow$ | $NO_2 + N_2$ | 1.378e-22 | (220) |

Table B.2: **Reaction network of the model of the Archean atmosphere.** Reaction equations have been extracted from the model source code. Vertically integrated rates are taken from the model output.

| | | | rates ($1/cm^2$) | |
|---|---|---|---|---|
| $H_2O + O(_1D)$ | $\rightarrow$ | $2OH$ | 5.48e+06 | (1) |
| $H_2 + O(_1D)$ | $\rightarrow$ | $OH + H$ | 1.443e+09 | (2) |
| $H_2 + O$ | $\rightarrow$ | $OH + H$ | 1.391e+08 | (3) |
| $H_2 + OH$ | $\rightarrow$ | $H_2O + H$ | 1.974e+09 | (4) |
| $H + O_3$ | $\rightarrow$ | $OH + O_2$ | 4.481e+10 | (5) |
| $H + O_2 + M$ | $\rightarrow$ | $HO_2 + M$ | 1.818e+12 | (6) |
| $H + HO_2$ | $\rightarrow$ | $H_2 + O_2$ | 4.955e+10 | (7) |
| $H + HO_2$ | $\rightarrow$ | $H_2O + O$ | 1.239e+10 | (8) |
| $H + HO_2$ | $\rightarrow$ | $2OH$ | 5.574e+11 | (9) |
| $OH + O$ | $\rightarrow$ | $H + O_2$ | 5.585e+11 | (10) |
| $OH + HO_2$ | $\rightarrow$ | $H_2O + O_2$ | 1.333e+08 | (11) |
| $OH + O_3$ | $\rightarrow$ | $HO_2 + O_2$ | 1.571e+04 | (12) |
| $HO_2 + O$ | $\rightarrow$ | $OH + O_2$ | 1.401e+12 | (13) |
| $HO_2 + O_3$ | $\rightarrow$ | $OH + 2O_2$ | 6.313e+04 | (14) |
| $2HO_2$ | $\rightarrow$ | $H_2O_2 + O_2$ | 3.268e+08 | (15) |
| $H_2O_2 + OH$ | $\rightarrow$ | $HO_2 + H_2O$ | 1.42e+05 | (16) |
| $2O + M$ | $\rightarrow$ | $O_2 + M$ | 6.844e+10 | (17) |
| $O + O_2 + M$ | $\rightarrow$ | $O_3 + M$ | 2.231e+11 | (18) |
| $O + O_3$ | $\rightarrow$ | $2O_2$ | 3.764e+06 | (19) |
| $2OH$ | $\rightarrow$ | $H_2O + O$ | 5.025e+06 | (20) |
| $O(_1D) + M$ | $\rightarrow$ | $O + M$ | 6.467e+11 | (21) |
| $O(_1D) + O_2$ | $\rightarrow$ | $O + O_2$ | 6.841e+07 | (22) |
| $O_2 + hv$ | $\rightarrow$ | $O + O(_1D)$ | 3.613e+10 | (23) |
| $O_2 + hv$ | $\rightarrow$ | $2O$ | 4.643e+09 | (24) |
| $H_2O + hv$ | $\rightarrow$ | $H + OH$ | 1.082e+11 | (25) |
| $O_3 + hv$ | $\rightarrow$ | $O_2 + O(_1D)$ | 1.42e+11 | (26) |
| $O_3 + hv$ | $\rightarrow$ | $O_2 + O$ | 3.627e+10 | (27) |
| $H_2O_2 + hv$ | $\rightarrow$ | $2OH$ | 3.266e+08 | (28) |
| $CO_2 + hv$ | $\rightarrow$ | $CO + O$ | 2.251e+12 | (29) |
| $CO + OH$ | $\rightarrow$ | $CO_2 + H$ | 2.4e+12 | (30) |
| $CO + O + M$ | $\rightarrow$ | $CO_2 + M$ | 6.659e+10 | (31) |
| $H + CO + M$ | $\rightarrow$ | $HCO + M$ | 2.035e+12 | (32) |
| $H + HCO$ | $\rightarrow$ | $H_2 + CO$ | 9.554e+10 | (33) |
| $2HCO$ | $\rightarrow$ | $H_2CO + CO$ | 2.058e+11 | (34) |
| $OH + HCO$ | $\rightarrow$ | $H_2O + CO$ | 4.751e+06 | (35) |
| $O + HCO$ | $\rightarrow$ | $H + CO_2$ | 2.074e+11 | (36) |
| $O + HCO$ | $\rightarrow$ | $OH + CO$ | 2.074e+11 | (37) |
| $H_2CO + hv$ | $\rightarrow$ | $H_2 + CO$ | 8.474e+10 | (38) |
| $H_2CO + hv$ | $\rightarrow$ | $HCO + H$ | 9.75e+10 | (39) |
| $HCO + hv$ | $\rightarrow$ | $H + CO$ | 8.537e+11 | (40) |
| $H_2CO + H$ | $\rightarrow$ | $H_2 + HCO$ | 4.475e+10 | (41) |
| $CO_2 + hv$ | $\rightarrow$ | $CO + O(_1D)$ | 4.713e+11 | (42) |
| $2H + M$ | $\rightarrow$ | $H_2 + M$ | 2.866e-04 | (43) |
| $HCO + O_2$ | $\rightarrow$ | $HO_2 + CO$ | 2.154e+11 | (44) |
| $H_2CO + OH$ | $\rightarrow$ | $H_2O + HCO$ | 1.23e+09 | (45) |
| $H + OH + M$ | $\rightarrow$ | $H_2O + M$ | 5.648e+08 | (46) |
| $2OH + M$ | $\rightarrow$ | $H_2O_2 + M$ | 1.113e+04 | (47) |

(continued on next page)

Table B.2: (continued - Archean atm. reaction network)

| | | | rates (1/cm$^2$) | |
|---|---|---|---|---|
| $H_2CO + O$ | $\rightarrow$ | $HCO + OH$ | 9.081e+09 | (48) |
| $H_2O_2 + O$ | $\rightarrow$ | $OH + HO_2$ | 8.214e+04 | (49) |
| $HO_2 + hv$ | $\rightarrow$ | $OH + O$ | 5.501e+09 | (50) |
| $CH_4 + hv$ | $\rightarrow$ | $_1CH_2 + H_2$ | 2.665e+10 | (51) |
| $C_2H_6 + hv$ | $\rightarrow$ | $2_3CH_2 + H_2$ | 0e+00 | (52) |
| $C_2H_6 + hv$ | $\rightarrow$ | $CH_4 + _1CH_2$ | 4.73e+06 | (53) |
| $HNO_2 + hv$ | $\rightarrow$ | $NO + OH$ | 1.432e+06 | (54) |
| $HNO_3 + hv$ | $\rightarrow$ | $NO_2 + OH$ | 1.229e+05 | (55) |
| $NO + hv$ | $\rightarrow$ | $N + O$ | 1.096e+05 | (56) |
| $NO_2 + hv$ | $\rightarrow$ | $NO + O$ | 6.589e+09 | (57) |
| $CH_4 + OH$ | $\rightarrow$ | $CH_3 + H_2O$ | 2.63e+10 | (58) |
| $CH_4 + O(_1D)$ | $\rightarrow$ | $CH_3 + OH$ | 1.084e+09 | (59) |
| $CH_4 + O(_1D)$ | $\rightarrow$ | $H_2CO + H_2$ | 1.084e+08 | (60) |
| $_1CH_2 + CH_4$ | $\rightarrow$ | $2CH_3$ | 3.665e+07 | (61) |
| $_1CH_2 + O_2$ | $\rightarrow$ | $HCO + OH$ | 1.182e+07 | (62) |
| $_1CH_2 + M$ | $\rightarrow$ | $_3CH_2 + M$ | 2.659e+10 | (63) |
| $_3CH_2 + H_2$ | $\rightarrow$ | $CH_3 + H$ | 0e+00 | (64) |
| $_3CH_2 + CH_4$ | $\rightarrow$ | $2CH_3$ | 0e+00 | (65) |
| $_3CH_2 + O_2$ | $\rightarrow$ | $HCO + OH$ | 3.721e+08 | (66) |
| $CH_3 + O_2 + M$ | $\rightarrow$ | $H_2CO + OH + M$ | 2.369e+09 | (67) |
| $CH_3 + OH$ | $\rightarrow$ | $H_2CO + H_2$ | 3.147e+07 | (68) |
| $CH_3 + O$ | $\rightarrow$ | $H_2CO + H$ | 8.379e+10 | (69) |
| $CH_3 + O_3$ | $\rightarrow$ | $H_2CO + HO_2$ | 1.29e+06 | (70) |
| $2CH_3 + M$ | $\rightarrow$ | $C_2H_6 + M$ | 5.265e+08 | (71) |
| $CH_3 + hv$ | $\rightarrow$ | $_1CH_2 + H$ | 3.905e+08 | (72) |
| $CH_3 + H + M$ | $\rightarrow$ | $CH_4 + M$ | 1.296e+09 | (73) |
| $CH_3 + HCO$ | $\rightarrow$ | $CH_4 + CO$ | 1.744e+10 | (74) |
| $CH_3 + HNO$ | $\rightarrow$ | $CH_4 + NO$ | 2.576e+06 | (75) |
| $CH_3 + H_2CO$ | $\rightarrow$ | $CH_4 + HCO$ | 3.688e+08 | (76) |
| $H + NO + M$ | $\rightarrow$ | $HNO + M$ | 3.69e+08 | (77) |
| $2N + M$ | $\rightarrow$ | $N_2 + M$ | 0e+00 | (78) |
| $N + O_2$ | $\rightarrow$ | $NO + O$ | 5.542e+00 | (79) |
| $N + O_3$ | $\rightarrow$ | $NO + O_2$ | 0e+00 | (80) |
| $N + OH$ | $\rightarrow$ | $NO + H$ | 1.386e+04 | (81) |
| $N + NO$ | $\rightarrow$ | $N_2 + O$ | 8.186e+05 | (82) |
| $NO + O_3$ | $\rightarrow$ | $NO_2 + O_2$ | 1.462e+05 | (83) |
| $NO + O + M$ | $\rightarrow$ | $NO_2 + M$ | 1.284e+10 | (84) |
| $NO + HO_2$ | $\rightarrow$ | $NO_2 + OH$ | 2.514e+09 | (85) |
| $NO + OH + M$ | $\rightarrow$ | $HNO_2 + M$ | 1.432e+06 | (86) |
| $NO_2 + O$ | $\rightarrow$ | $NO + O_2$ | 5.873e+09 | (87) |
| $NO_2 + OH + M$ | $\rightarrow$ | $HNO_3 + M$ | 1.229e+05 | (88) |
| $NO_2 + H$ | $\rightarrow$ | $NO + OH$ | 1.128e+09 | (89) |
| $HNO_3 + OH$ | $\rightarrow$ | $H_2O + NO_2 + O$ | 1.352e+01 | (90) |
| $HCO + NO$ | $\rightarrow$ | $HNO + CO$ | 2.11e+10 | (91) |
| $HNO + hv$ | $\rightarrow$ | $NO + H$ | 4.613e+10 | (92) |
| $H + HNO$ | $\rightarrow$ | $H_2 + NO$ | 1.649e+06 | (93) |
| $O + HNO$ | $\rightarrow$ | $OH + NO$ | 1.668e+07 | (94) |
| $OH + HNO$ | $\rightarrow$ | $H_2O + NO$ | 2.012e+06 | (95) |
| $HNO_2 + OH$ | $\rightarrow$ | $H_2O + NO_2$ | 1.711e+01 | (96) |
| $CH_4 + O$ | $\rightarrow$ | $CH_3 + OH$ | 1.886e+10 | (97) |
| $_1CH_2 + H_2$ | $\rightarrow$ | $CH_3 + H$ | 2.924e+08 | (98) |

(continued on next page)

Table B.2: (continued - Archean atm. reaction network)

| | | | rates ($1/cm^2$) | |
|---|---|---|---|---|
| $_1CH_2 + CO_2$ | $\rightarrow$ | $H_2CO + CO$ | 7.555e+07 | (99) |
| $_3CH_2 + O$ | $\rightarrow$ | $HCO + H$ | 1.529e+09 | (100) |
| $_3CH_2 + CO_2$ | $\rightarrow$ | $H_2CO + CO$ | 4.946e+09 | (101) |
| $C_2H_6 + OH$ | $\rightarrow$ | $C_2H_5 + H_2O$ | 1.06e+09 | (102) |
| $C_2H_6 + O$ | $\rightarrow$ | $C_2H_5 + OH$ | 1.847e+09 | (103) |
| $C_2H_6 + O(_1D)$ | $\rightarrow$ | $C_2H_5 + OH$ | 3.131e+05 | (104) |
| $C_2H_5 + H$ | $\rightarrow$ | $2CH_3$ | 1.665e+09 | (105) |
| $C_2H_5 + O$ | $\rightarrow$ | $CH_3 + HCO + H$ | 8.116e+08 | (106) |
| $C_2H_5 + OH$ | $\rightarrow$ | $CH_3 + HCO + H_2$ | 1.461e+05 | (107) |
| $C_2H_5 + HCO$ | $\rightarrow$ | $C_2H_6 + CO$ | 8.916e+08 | (108) |
| $C_2H_5 + HNO$ | $\rightarrow$ | $C_2H_6 + NO$ | 2.104e+05 | (109) |
| $C_2H_5 + O_2 + M$ | $\rightarrow$ | $CH_3 + HCO + OH + M$ | 6.317e+08 | (110) |
| $SO + hv$ | $\rightarrow$ | $S + O$ | 0e+00 | (111) |
| $SO_2 + hv$ | $\rightarrow$ | $SO + O$ | 1.025e+11 | (112) |
| $H_2S + hv$ | $\rightarrow$ | $HS + H$ | 1.905e+11 | (113) |
| $SO + O_2$ | $\rightarrow$ | $O + SO_2$ | 8.547e+04 | (114) |
| $SO + HO_2$ | $\rightarrow$ | $SO_2 + OH$ | 3.923e+09 | (115) |
| $SO + O + M$ | $\rightarrow$ | $SO_2 + M$ | 9.424e+10 | (116) |
| $SO + OH$ | $\rightarrow$ | $SO_2 + H$ | 5.229e+07 | (117) |
| $SO_2 + OH + M$ | $\rightarrow$ | $HSO_3 + M$ | 2.878e+06 | (118) |
| $SO_2 + O + M$ | $\rightarrow$ | $SO_3 + M$ | 9.054e+07 | (119) |
| $SO_3 + H_2O$ | $\rightarrow$ | $H_2SO_4$ | 9.342e+07 | (120) |
| $HSO_3 + O_2$ | $\rightarrow$ | $HO_2 + SO_3$ | 4.612e+05 | (121) |
| $HSO_3 + OH$ | $\rightarrow$ | $H_2O + SO_3$ | 6.281e+01 | (122) |
| $HSO_3 + H$ | $\rightarrow$ | $H_2 + SO_3$ | 1.245e+05 | (123) |
| $HSO_3 + O$ | $\rightarrow$ | $OH + SO_3$ | 2.292e+06 | (124) |
| $H_2S + OH$ | $\rightarrow$ | $H_2O + HS$ | 6.704e+06 | (125) |
| $H_2S + H$ | $\rightarrow$ | $H_2 + HS$ | 1.078e+09 | (126) |
| $H_2S + O$ | $\rightarrow$ | $OH + HS$ | 2.164e+07 | (127) |
| $HS + O$ | $\rightarrow$ | $H + SO$ | 2.241e+10 | (128) |
| $HS + O_2$ | $\rightarrow$ | $OH + SO$ | 6.89e+00 | (129) |
| $HS + HO_2$ | $\rightarrow$ | $H_2S + O_2$ | 7.366e+07 | (130) |
| $2HS$ | $\rightarrow$ | $H_2S + S$ | 1.514e+10 | (131) |
| $HS + HCO$ | $\rightarrow$ | $H_2S + CO$ | 1.262e+11 | (132) |
| $HS + H$ | $\rightarrow$ | $H_2 + S$ | 3.684e+08 | (133) |
| $HS + S$ | $\rightarrow$ | $H + S_2$ | 4.326e+09 | (134) |
| $S + O_2$ | $\rightarrow$ | $SO + O$ | 1.827e+07 | (135) |
| $S + OH$ | $\rightarrow$ | $SO + H$ | 1.355e+05 | (136) |
| $S + HCO$ | $\rightarrow$ | $HS + CO$ | 8.659e+09 | (137) |
| $S + HO_2$ | $\rightarrow$ | $HS + O_2$ | 1.493e+06 | (138) |
| $S + HO_2$ | $\rightarrow$ | $SO + OH$ | 1.493e+06 | (139) |
| $2S$ | $\rightarrow$ | $S_2$ | 1.097e+06 | (140) |
| $S_2 + OH$ | $\rightarrow$ | $HSO + S$ | 0e+00 | (141) |
| $S_2 + O$ | $\rightarrow$ | $S + SO$ | 3.017e+08 | (142) |
| $HS + H_2CO$ | $\rightarrow$ | $H_2S + HCO$ | 4.733e+10 | (143) |
| $SO_2 + hv$ | $\rightarrow$ | $_1SO_2$ | 1.592e+12 | (144) |
| $SO_2 + hv$ | $\rightarrow$ | $_3SO_2$ | 9.135e+08 | (145) |
| $S_2 + hv$ | $\rightarrow$ | $2S$ | 4.014e+09 | (146) |
| $S_2 + hv$ | $\rightarrow$ | $S_2$ | 0e+00 | (147) |
| $H_2SO_4 + hv$ | $\rightarrow$ | $SO_2 + 2OH$ | 8.72e+05 | (148) |
| $SO_3 + hv$ | $\rightarrow$ | $SO_2 + O$ | 0e+00 | (149) |

(continued on next page)

Table B.2: (continued - Archean atm. reaction network)

| | | | rates ($1/cm^2$) | |
|---|---|---|---|---|
| $_1SO_2 + M$ | → | $_3SO_2 + M$ | 1.447e+11 | (150) |
| $_1SO_2 + M$ | → | $SO_2 + M$ | 1.447e+12 | (151) |
| $hv + _1SO_2$ | → | $_3SO_2 + hv$ | 2.762e+07 | (152) |
| $hv + _1SO_2$ | → | $SO_2 + hv$ | 4.052e+08 | (153) |
| $_1SO_2 + O_2$ | → | $SO_3 + O$ | 9.188e-03 | (154) |
| $_1SO_2 + SO_2$ | → | $SO_3 + SO$ | 7.954e+01 | (155) |
| $_3SO_2 + M$ | → | $SO_2 + M$ | 1.455e+11 | (156) |
| $hv + _3SO_2$ | → | $SO_2 + hv$ | 1.393e+08 | (157) |
| $_3SO_2 + SO_2$ | → | $SO_3 + SO$ | 9.331e+00 | (158) |
| $SO + NO_2$ | → | $SO_2 + NO$ | 1.722e+09 | (159) |
| $SO + O_3$ | → | $SO_2 + O_2$ | 7.409e+05 | (160) |
| $SO_2 + HO_2$ | → | $SO_3 + OH$ | 0e+00 | (161) |
| $HS + O_3$ | → | $HSO + O_2$ | 4.477e+03 | (162) |
| $HS + NO_2$ | → | $HSO + NO$ | 4.148e+07 | (163) |
| $S + O_3$ | → | $SO + O_2$ | 3.436e+04 | (164) |
| $2SO$ | → | $SO_2 + S$ | 2.345e+09 | (165) |
| $SO_3 + SO$ | → | $2SO_2$ | 3.137e+03 | (166) |
| $S + CO_2$ | → | $SO + CO$ | 1.005e+10 | (167) |
| $SO + HO_2$ | → | $HSO + O_2$ | 0e+00 | (168) |
| $SO + HCO$ | → | $HSO + CO$ | 1.032e+11 | (169) |
| $H + SO + M$ | → | $HSO + M$ | 9.929e+08 | (170) |
| $HSO + hv$ | → | $HS + O$ | 2.353e+10 | (171) |
| $HSO + NO$ | → | $HNO + SO$ | 2.468e+10 | (172) |
| $HSO + OH$ | → | $H_2O + SO$ | 7.627e+06 | (173) |
| $HSO + H$ | → | $HS + OH$ | 7.118e+09 | (174) |
| $HSO + H$ | → | $H_2 + SO$ | 6.327e+08 | (175) |
| $HSO + HS$ | → | $H_2S + SO$ | 2.961e+09 | (176) |
| $HSO + O$ | → | $OH + SO$ | 4.22e+10 | (177) |
| $HSO + S$ | → | $HS + SO$ | 3.126e+09 | (178) |
| $S + S_2 + M$ | → | $S_3 + M$ | 1.277e+07 | (179) |
| $2S_2 + M$ | → | $S_4 + M$ | 3.22e+08 | (180) |
| $S + S_3 + M$ | → | $S_4 + M$ | 4.273e+05 | (181) |
| $2S_4 + M$ | → | $S_8(aer) + M$ | 2.878e+06 | (182) |
| $S_4 + hv$ | → | $2S_2$ | 3.167e+08 | (183) |
| $S_3 + hv$ | → | $S_2 + S$ | 1.235e+07 | (184) |
| $NH_3 + hv$ | → | $NH_2 + H$ | 5.375e+09 | (185) |
| $NH_3 + OH$ | → | $NH_2 + H_2O$ | 2.39e+04 | (186) |
| $NH_3 + O(_1D)$ | → | $NH_2 + OH$ | 2.466e-07 | (187) |
| $NH_2 + H + M$ | → | $NH_3 + M$ | 1.249e+08 | (188) |
| $NH_2 + NO$ | → | $N_2 + H_2O$ | 1.67e+08 | (189) |
| $2NH_2 + M$ | → | $N_2H_4 + M$ | 1.14e+09 | (190) |
| $NH_2 + O$ | → | $NH + OH$ | 5.291e+06 | (191) |
| $NH_2 + O$ | → | $HNO + H$ | 5.291e+06 | (192) |
| $NH + NO$ | → | $N_2 + O + H$ | 6.686e+06 | (193) |
| $NH + O$ | → | $N + OH$ | 2.219e+05 | (194) |
| $N_2H_4 + hv$ | → | $N_2H_3 + H$ | 2.116e+09 | (195) |
| $N_2H_4 + H$ | → | $N_2H_3 + H_2$ | 4.978e+06 | (196) |
| $N_2H_3 + H$ | → | $2NH_2$ | 6.486e+06 | (197) |
| $2N_2H_3$ | → | $N_2H_4 + N_2 + H_2$ | 1.053e+09 | (198) |
| $NH + H + M$ | → | $NH_2 + M$ | 1.627e+06 | (199) |
| $NH + hv$ | → | $N + H$ | 5.011e+05 | (200) |

(continued on next page)

Table B.2: (continued - Archean atm. reaction network)

| | | | rates ($1/cm^2$) | |
|---|---|---|---|---|
| $NH_2$ + hv | → | NH + H | 3.489e+07 | (201) |
| $NH_2$ + hv | → | $NH_2(s)$ | 7.735e+09 | (202) |
| hv + $NH_2(s)$ | → | $NH_2$ + hv | 1.404e+06 | (203) |
| $NH_2(s)$ + M | → | $NH_2$ + M | 7.733e+09 | (204) |
| $NH_2(s)$ + $H_2$ | → | $NH_3$ + H | 4.352e+05 | (205) |
| $NH_2$ + HCO | → | $NH_3$ + CO | 2.795e+09 | (206) |
| NH + HCO | → | $NH_2$ + CO | 3.114e+07 | (207) |
| $_1CH_2$ + $O_2$ | → | $H_2CO$ + O | 0e+00 | (208) |
| $_3CH_2$ + $O_2$ | → | $H_2CO$ + O | 0e+00 | (209) |
| $C_2H_2$ + hv | → | $C_2H$ + H | 4.156e+06 | (210) |
| $C_2H_2$ + hv | → | $C_2$ + $H_2$ | 3.253e+06 | (211) |
| $C_2H_4$ + hv | → | $C_2H_2$ + $H_2$ | 8.592e+07 | (212) |
| $_3CH_2$ + $CH_3$ | → | $C_2H_4$ + H | 5.587e+04 | (213) |
| $C_2H_5$ + $CH_3$ + M | → | $C_3H_8$ + M | 4.095e+07 | (214) |
| $C_3H_8$ + OH | → | $C_3H_7$ + $H_2O$ | 2.181e+06 | (215) |
| $C_3H_8$ + O | → | $C_3H_7$ + OH | 3.877e+07 | (216) |
| $C_3H_8$ + $O(_1D)$ | → | $C_3H_7$ + OH | 1.575e+01 | (217) |
| $C_3H_7$ + H | → | $CH_3$ + $C_2H_5$ | 2.927e+06 | (218) |
| $2_3CH_2$ | → | $C_2H_2$ + 2H | 0e+00 | (219) |
| $C_2H_2$ + OH | → | CO + $CH_3$ | 1.296e+07 | (220) |
| $C_2H_2$ + H + M | → | $C_2H_3$ + M | 2.249e+08 | (221) |
| $C_2H_3$ + H | → | $C_2H_2$ + $H_2$ | 2.16e+08 | (222) |
| $C_2H_3$ + $H_2$ | → | $C_2H_4$ + H | 1.579e+03 | (223) |
| $C_2H_3$ + $CH_4$ | → | $C_2H_4$ + $CH_3$ | 6.685e+02 | (224) |
| $C_2H_3$ + $C_2H_6$ | → | $C_2H_4$ + $C_2H_5$ | 6.189e-04 | (225) |
| $C_2H_4$ + OH | → | $H_2CO$ + $CH_3$ | 4.551e+08 | (226) |
| $C_2H_4$ + O | → | HCO + $CH_3$ | 8.797e+09 | (227) |
| $C_2H_4$ + H + M | → | $C_2H_5$ + M | 1.96e+09 | (228) |
| $C_2H$ + $O_2$ | → | CO + HCO | 3.661e+06 | (229) |
| $C_2H$ + $H_2$ | → | $C_2H_2$ + H | 5.671e+04 | (230) |
| $C_2H$ + $CH_4$ | → | $C_2H_2$ + $CH_3$ | 2.08e+06 | (231) |
| $C_2H$ + $C_2H_6$ | → | $C_2H_2$ + $C_2H_5$ | 1.079e+04 | (232) |
| $C_2H$ + H + M | → | $C_2H_2$ + M | 8.008e+05 | (233) |
| $C_3H_8$ + hv | → | $C_3H_6$ + $H_2$ | 8.329e+02 | (234) |
| $C_3H_8$ + hv | → | $C_2H_6$ + $_1CH_2$ | 2.223e+02 | (235) |
| $C_3H_8$ + hv | → | $C_2H_4$ + $CH_4$ | 9.632e+02 | (236) |
| $C_3H_8$ + hv | → | $C_2H_5$ + $CH_3$ | 4.951e+02 | (237) |
| $C_2H_6$ + hv | → | $C_2H_2$ + $2H_2$ | 4.825e+06 | (238) |
| $C_2H_6$ + hv | → | $C_2H_4$ + 2H | 5.72e+06 | (239) |
| $C_2H_6$ + hv | → | $C_2H_4$ + $H_2$ | 2.669e+06 | (240) |
| $C_2H_6$ + hv | → | $2CH_3$ | 1.515e+06 | (241) |
| $C_2H_4$ + hv | → | $C_2H_2$ + 2H | 8.255e+07 | (242) |
| $C_3H_6$ + hv | → | $C_2H_2$ + $CH_3$ + H | 4.753e+01 | (243) |
| $CH_4$ + hv | → | $_3CH_2$ + 2H | 2.711e+10 | (244) |
| $CH_4$ + hv | → | $CH_3$ + H | 5.531e+10 | (245) |
| CH + hv | → | C + H | 4.332e+04 | (246) |
| $CH_2CO$ + hv | → | $_3CH_2$ + CO | 2.116e+09 | (247) |
| $CH_3CHO$ + hv | → | $CH_3$ + HCO | 1.545e+08 | (248) |
| $CH_3CHO$ + hv | → | $CH_4$ + CO | 1.545e+08 | (249) |
| $C_2H_5CHO$ + hv | → | $C_2H_5$ + HCO | 3.789e+07 | (250) |
| $C_3H_3$ + hv | → | $C_3H_2$ + H | 2.261e+03 | (251) |

(continued on next page)

Table B.2: (continued - Archean atm. reaction network)

| | | | rates (1/cm$^2$) | |
|---|---|---|---|---|
| $CH_3C_2H$ + hv | → | $C_3H_3$ + H | 1.176e+06 | (252) |
| $CH_3C_2H$ + hv | → | $C_3H_2$ + $H_2$ | 4.41e+05 | (253) |
| $CH_3C_2H$ + hv | → | $CH_3$ + $C_2H$ | 5.88e+04 | (254) |
| $CH_2CCH_2$ + hv | → | $C_3H_3$ + H | 2.137e-02 | (255) |
| $CH_2CCH_2$ + hv | → | $C_3H_2$ + $H_2$ | 8.013e-03 | (256) |
| $CH_2CCH_2$ + hv | → | $C_2H_2$ + $_3CH_2$ | 3.205e-03 | (257) |
| $C_3H_6$ + hv | → | $CH_2CCH_2$ + $H_2$ | 7.968e+01 | (258) |
| $C_3H_6$ + hv | → | $C_2H_4$ + $_3CH_2$ | 2.796e+00 | (259) |
| $C_3H_6$ + hv | → | $C_2H$ + $CH_4$ + H | 6.989e+00 | (260) |
| C + OH | → | CO + H | 6.829e+04 | (261) |
| C + $H_2$ + M | → | $_3CH_2$ + M | 4.691e+06 | (262) |
| C + $O_2$ | → | CO + O | 2.347e+09 | (263) |
| CH + H | → | C + $H_2$ | 2.351e+09 | (264) |
| CH + O | → | CO + H | 4.958e+09 | (265) |
| CH + $H_2$ | → | $_3CH_2$ + H | 8.866e+06 | (266) |
| CH + $H_2$ + M | → | $CH_3$ + M | 4.338e+06 | (267) |
| CH + $O_2$ | → | CO + OH | 4.343e+09 | (268) |
| CH + $CO_2$ | → | HCO + CO | 1.354e+10 | (269) |
| CH + $CH_4$ | → | $C_2H_4$ + H | 1.125e+10 | (270) |
| CH + $C_2H_2$ | → | $C_3H_2$ + H | 2.707e+06 | (271) |
| CH + $C_2H_4$ | → | $CH_3C_2H$ + H | 1.976e+06 | (272) |
| CH + $C_2H_4$ | → | $CH_2CCH_2$ + H | 1.976e+06 | (273) |
| $_3CH_2$ + O | → | CH + OH | 1.223e+09 | (274) |
| $_3CH_2$ + O | → | CO + 2H | 1.269e+10 | (275) |
| $_3CH_2$ + H + M | → | $CH_3$ + M | 9.434e+06 | (276) |
| $_3CH_2$ + H | → | CH + $H_2$ | 3.524e+10 | (277) |
| $_3CH_2$ + CO + M | → | $CH_2CO$ + M | 3.99e+06 | (278) |
| $2_3CH_2$ | → | $C_2H_2$ + $H_2$ | 1.941e+03 | (279) |
| $_3CH_2$ + $C_2H_2$ + M | → | $CH_3C_2H$ + M | 1.103e+05 | (280) |
| $_3CH_2$ + $C_2H_3$ | → | $CH_3$ + $C_2H_2$ | 6.594e+00 | (281) |
| $_3CH_2$ + $C_2H_5$ | → | $CH_3$ + $C_2H_4$ | 6.491e+01 | (282) |
| $CH_2CO$ + H | → | $CH_3$ + CO | 3.367e+05 | (283) |
| $CH_2CO$ + O | → | $H_2CO$ + CO | 3.105e+08 | (284) |
| $CH_2CCH_2$ + H + M | → | $CH_3$ + $C_2H_2$ + M | 1.177e+04 | (285) |
| $CH_2CCH_2$ + H + M | → | $C_3H_5$ + M | 2.214e+04 | (286) |
| $CH_3$ + $O_2$ + M | → | $CH_3O_2$ + M | 0e+00 | (287) |
| $CH_3$ + CO + M | → | $CH_3CO$ + M | 5.443e+09 | (288) |
| $CH_3$ + $H_2CO$ | → | $CH_4$ + HCO | 0e+00 | (289) |
| $CH_3$ + OH | → | CO + $2H_2$ | 2.267e+06 | (290) |
| $CH_3$ + $C_2H_3$ | → | $C_3H_5$ + H | 3.199e+00 | (291) |
| $CH_3O_2$ + H | → | $CH_4$ + $O_2$ | 0e+00 | (292) |
| $CH_3O_2$ + H | → | $H_2O$ + $H_2CO$ | 0e+00 | (293) |
| $CH_3O_2$ + O | → | $H_2CO$ + $HO_2$ | 0e+00 | (294) |
| $CH_3CO$ + H | → | $CH_4$ + CO | 1.225e+09 | (295) |
| $CH_3CO$ + O | → | $H_2CO$ + HCO | 7.363e+08 | (296) |
| $CH_3CO$ + $CH_3$ | → | $C_2H_6$ + CO | 1.508e+09 | (297) |
| $CH_3CO$ + $CH_3$ | → | $CH_4$ + $CH_2CO$ | 2.402e+09 | (298) |
| $CH_3CHO$ + H | → | $CH_3CO$ + $H_2$ | 2.329e+07 | (299) |
| $CH_3CHO$ + O | → | $CH_3CO$ + OH | 4.035e+08 | (300) |
| $CH_3CHO$ + OH | → | $CH_3CO$ + $H_2O$ | 1.905e+06 | (301) |
| $CH_3CHO$ + $CH_3$ | → | $CH_3CO$ + $CH_4$ | 1.537e+05 | (302) |

(continued on next page)

Table B.2: (continued - Archean atm. reaction network)

| | | | rates ($1/cm^2$) | |
|---|---|---|---|---|
| $CH_3C_2H + H + M$ | $\rightarrow$ | $CH_3 + C_2H_2 + M$ | 5.153e+06 | (303) |
| $CH_3C_2H + H + M$ | $\rightarrow$ | $C_3H_5 + M$ | 5.153e+06 | (304) |
| $C_2 + O$ | $\rightarrow$ | $C + CO$ | 3.235e+05 | (305) |
| $C_2 + O_2$ | $\rightarrow$ | $2CO$ | 4.024e+04 | (306) |
| $C_2 + H_2$ | $\rightarrow$ | $C_2H + H$ | 4.209e+04 | (307) |
| $C_2 + CH_4$ | $\rightarrow$ | $C_2H + CH_3$ | 2.847e+06 | (308) |
| $C_2H + O$ | $\rightarrow$ | $CO + CH$ | 4.824e+05 | (309) |
| $C_2H + C_3H_8$ | $\rightarrow$ | $C_2H_2 + C_3H_7$ | 2.178e-01 | (310) |
| $C_2H_2 + O$ | $\rightarrow$ | $_3CH_2 + CO$ | 1.39e+08 | (311) |
| $C_2H_2 + OH$ | $\rightarrow$ | $C_2H_2OH$ | 1.016e+05 | (312) |
| $C_2H_2 + OH + M$ | $\rightarrow$ | $CH_2CO + H + M$ | 1.185e+07 | (313) |
| $C_2H_2OH + H$ | $\rightarrow$ | $H_2O + C_2H_2$ | 1.016e+05 | (314) |
| $C_2H_2OH + H$ | $\rightarrow$ | $H_2 + CH_2CO$ | 1.004e+00 | (315) |
| $C_2H_2OH + O$ | $\rightarrow$ | $OH + CH_2CO$ | 2.874e-02 | (316) |
| $C_2H_2OH + OH$ | $\rightarrow$ | $H_2O + CH_2CO$ | 3.24e-03 | (317) |
| $C_2H_3 + O$ | $\rightarrow$ | $CH_2CO + H$ | 8.85e+06 | (318) |
| $C_2H_3 + OH$ | $\rightarrow$ | $C_2H_2 + H_2O$ | 1.23e+03 | (319) |
| $C_2H_3 + CH_3$ | $\rightarrow$ | $C_2H_2 + CH_4$ | 4.532e+02 | (320) |
| $C_2H_3 + CH_3 + M$ | $\rightarrow$ | $C_3H_6 + M$ | 1.597e+03 | (321) |
| $2C_2H_3$ | $\rightarrow$ | $C_2H_4 + C_2H_2$ | 1.382e-01 | (322) |
| $C_2H_3 + C_2H_5$ | $\rightarrow$ | $2C_2H_4$ | 4.225e-01 | (323) |
| $C_2H_3 + C_2H_5$ | $\rightarrow$ | $CH_3 + C_3H_5$ | 9.278e-01 | (324) |
| $C_2H_4 + OH + M$ | $\rightarrow$ | $C_2H_4OH + M$ | 2.098e+06 | (325) |
| $C_2H_4OH + H$ | $\rightarrow$ | $H_2O + C_2H_4$ | 2.098e+06 | (326) |
| $C_2H_4OH + H$ | $\rightarrow$ | $H_2 + CH_3CHO$ | 2.184e+01 | (327) |
| $C_2H_4OH + O$ | $\rightarrow$ | $OH + CH_3CHO$ | 2.365e+00 | (328) |
| $C_2H_4OH + OH$ | $\rightarrow$ | $H_2O + CH_3CHO$ | 9.121e-02 | (329) |
| $C_2H_5 + OH$ | $\rightarrow$ | $CH_3CHO + H_2$ | 1.328e+05 | (330) |
| $C_2H_5 + O$ | $\rightarrow$ | $CH_3CHO + H$ | 7.378e+08 | (331) |
| $C_2H_5 + CH_3$ | $\rightarrow$ | $C_2H_4 + CH_4$ | 1.639e+06 | (332) |
| $C_2H_5 + C_2H_3$ | $\rightarrow$ | $C_2H_6 + C_2H_2$ | 8.451e-01 | (333) |
| $2C_2H_5$ | $\rightarrow$ | $C_2H_6 + C_2H_4$ | 1.476e+05 | (334) |
| $C_2H_5 + H + M$ | $\rightarrow$ | $C_2H_6 + M$ | 1.265e+05 | (335) |
| $C_2H_5 + H$ | $\rightarrow$ | $C_2H_4 + H_2$ | 1.271e+08 | (336) |
| $C_3H_2 + H + M$ | $\rightarrow$ | $C_3H_3 + M$ | 3.15e+06 | (337) |
| $C_3H_3 + H + M$ | $\rightarrow$ | $CH_3C_2H + M$ | 2.162e+06 | (338) |
| $C_3H_3 + H + M$ | $\rightarrow$ | $CH_2CCH_2 + M$ | 2.162e+06 | (339) |
| $C_3H_5 + H$ | $\rightarrow$ | $CH_3C_2H + H_2$ | 1.723e+06 | (340) |
| $C_3H_5 + H + M$ | $\rightarrow$ | $C_3H_6 + M$ | 6.928e+03 | (341) |
| $C_3H_5 + H$ | $\rightarrow$ | $CH_4 + C_2H_2$ | 1.723e+06 | (342) |
| $C_3H_5 + CH_3$ | $\rightarrow$ | $CH_3C_2H + CH_4$ | 1.276e+01 | (343) |
| $C_3H_5 + CH_3$ | $\rightarrow$ | $CH_2CCH_2 + CH_4$ | 1.276e+01 | (344) |
| $C_3H_6 + OH$ | $\rightarrow$ | $CH_3CHO + CH_3$ | 1.007e+03 | (345) |
| $C_3H_6 + O$ | $\rightarrow$ | $2CH_3 + CO$ | 1.363e+05 | (346) |
| $C_3H_6 + H + M$ | $\rightarrow$ | $C_3H_7 + M$ | 2.974e+04 | (347) |
| $C_3H_7 + CH_3$ | $\rightarrow$ | $C_3H_6 + CH_4$ | 1.578e+05 | (348) |
| $C_3H_7 + OH$ | $\rightarrow$ | $C_2H_5CHO + H_2$ | 4.027e+03 | (349) |
| $C_3H_7 + O$ | $\rightarrow$ | $C_2H_5CHO + H$ | 3.789e+07 | (350) |
| $H + CH_2CCH_2$ | $\rightarrow$ | $CH_3C_2H + H$ | 6.011e+06 | (351) |
| $O + H_2CO$ | $\rightarrow$ | $HCO + OH$ | 9.081e+09 | (352) |
| $_3CH_2 + C_2H_2 + M$ | $\rightarrow$ | $CH_2CCH_2 + M$ | 1.842e+05 | (353) |

(continued on next page)

Table B.2: (continued - Archean atm. reaction network)

| | | | rates ($1/cm^2$) | |
|---|---|---|---|---|
| $C_2H + C_2H_2$ | $\rightarrow$ | $HC(aer) + H$ | 1.236e+04 | (354) |
| $_1CH_2 + H_2$ | $\rightarrow$ | $_3CH_2 + H_2$ | 3.987e+07 | (355) |
| $C_3H_5 + H$ | $\rightarrow$ | $CH_2CCH_2 + H_2$ | 1.723e+06 | (356) |
| $HCO + H_2CO$ | $\rightarrow$ | $CH_3O + CO$ | 1.126e+07 | (357) |
| $CH_3O + CO$ | $\rightarrow$ | $CH_3 + CO_2$ | 1.126e+07 | (358) |
| $C_2H + CH_2CCH_2$ | $\rightarrow$ | $HC(aer) + H$ | 1.678e+00 | (359) |

## B.2 SELECTED ARTIFICIAL NETWORKS

In Sect. 7.2 two artificial chemistries with especially interesting properties were presented and discussed in relation to the idea of further applications of reaction network analysis. We can not investigate the details of how probable it is to find such networks with scientific rigour here, but besides stating the existence of these networks a short explanation of how they were found should be given.

Roughly 10.000 artificial chemistries[3] are generated to semi-automatically search for especially interesting networks. All networks contain 12+1 chemical species and 24 reactions. From these reactions 8 are linear reactions and 4 are photochemical reactions. The parameter for the number of elementary components of the chemical species (one or two) is randomly chosen for each network. The other parameters that control how duplicate reactions are handled are also randomly chosen.[4].

The concentration of all species (except $h\nu$) is set to 10 initially and then reactions are applied randomly to randomise initial conditions before starting the simulation. Each network is simulated to $1 \cdot 10^7$ time units twenty times, each simulation with different initial conditions. The concentration of $h\nu$ (driving force) is set to 50 for all simulations.

To ensure that the networks approach their steady states as much as possible in the simulated time frame the activation energy is cutoff at a value of three. This prohibits higher activation energies and thus slower dynamics that takes longer to relaxate.

[3] *See Sect. 5.1.1.*

*12 normal species plus one for controlling "photochemical" reactions ($h\nu$).*

[4] *Parameters are explained in the source code of the function* jrnf_create_artificial_ecosystem *in* jrnf_R_tools *(see Appendix C)*

### B.2.1 *Bistable Artificial Chemistry*

To identify bistable artificial chemistries in the networks simulated to $1 \cdot 10^7$ time units, for each network the statistics of inflow of $h\nu$ for the last time step is investigated. If a network has multiple steady states it is reasonable that those steady states also have a different inflow values. Because of this, we start looking at the inflow distribution manually, starting with networks for which the quotient between the standard deviation of inflow values and the average inflow is high. This means that the spread between the inflows is relatively large. We have to identify those networks that have two distinct inflow values manually, because slow relaxation also may lead to a relative big standard deviation. These network then are investigated for different values for the strength of their driving force ($h\nu$ concentration). The network topology of the found bistable artificial chemistry can be seen in Fig. B.1, a full list of reactions is given in Table B.3.

### B.2.2 *Oscillating Artificial Chemistry*

The (successful) candidate for an artificial chemistry with periodic dynamics was easy to identify. As the implicit solver[5] for differential equations (Sect. A.1) uses an adaptive step size it can increase step

[5] Implicit *solvers are especially suited for stiff equations (multiple timescales).*

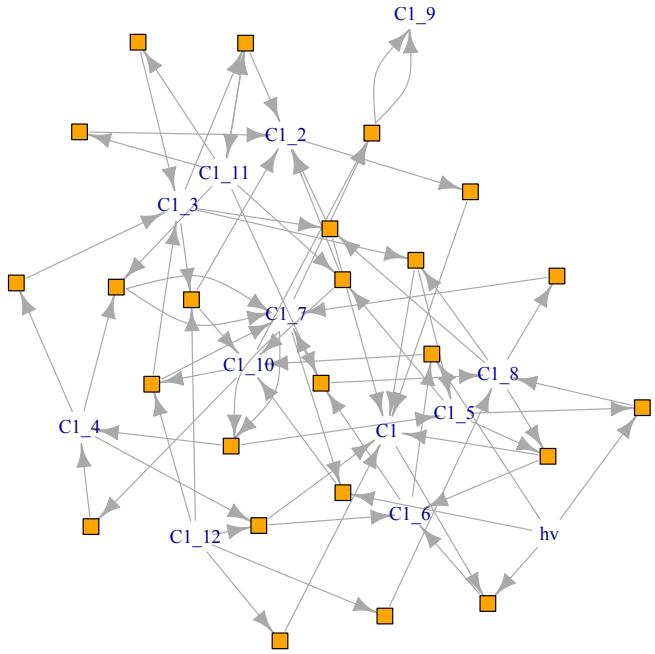



Figure B.1: **Plot of sample bistable artificial chemistry.** (Cf. Table B.3.)

size when approaching steady state. Thus, the artificial chemistries (as described above) take normally less than one second to integrate to $1 \cdot 10^7$ time units. This is not true for periodic systems (which do not approach any steady state). When integrating thousands of candidate networks, the one network that did not finish integration for hours was investigated manually and identified as having a periodic dynamics. The network topology can be seen in Fig. B.2, a full list of reactions is given in Table B.4.

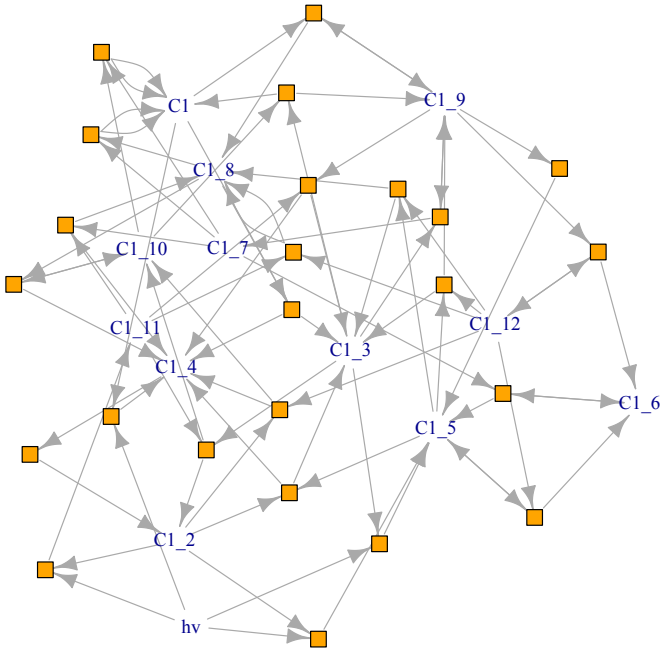



Figure B.2: **Plot of sample artificial chemistry with oscillating dynamics.** (Cf. Table B.4).

Table B.3: **Sample for an artificial reaction chemistry that has two stable steady states.** Chemistry was found by generating and simulating a large number of artificial reaction networks ($\approx$ 10000) and manually looking at heuristically preselected candidates. Forward reaction constants $k^f$ and backward reaction constants $k^b$ have been calculated from random energies as shown in Sect. 3.1.2.

| | | | $k^f$ | $k^b$ | |
|---|---|---|---|---|---|
| $hv + C_1$ | $\rightarrow$ | $_6C_1$ | 4.978707e-02 | 1.492498e-24 | (1) |
| $hv + _5C_1$ | $\rightarrow$ | $_8C_1$ | 7.452432e-01 | 7.727118e-22 | (2) |
| $hv + _7C_1$ | $\rightarrow$ | $_{10}C_1$ | 7.142340e-01 | 6.735825e-22 | (3) |
| $hv + _6C_1$ | $\rightarrow$ | $_{10}C_1$ | 6.737922e-01 | 1.476818e-21 | (4) |
| $_{12}C_1$ | $\rightarrow$ | $_8C_1$ | 9.131044e-02 | 6.341485e-01 | (5) |
| $_4C_1$ | $\rightarrow$ | $_3C_1$ | 3.847280e-01 | 5.513001e-02 | (6) |
| $_2C_1$ | $\rightarrow$ | $C_1$ | 1.210867e-01 | 1.324041e-01 | (7) |
| $_8C_1$ | $\rightarrow$ | $_7C_1$ | 7.304459e-02 | 1.536627e-02 | (8) |
| $_{11}C_1$ | $\rightarrow$ | $_3C_1$ | 5.125628e-01 | 2.566127e-01 | (9) |
| $_{12}C_1$ | $\rightarrow$ | $C_1$ | 2.240622e-01 | 9.062480e-01 | (10) |
| $_{10}C_1$ | $\rightarrow$ | $_4C_1$ | 4.269291e-01 | 1.602344e-01 | (11) |
| $_{11}C_1$ | $\rightarrow$ | $_2C_1$ | 3.181214e-02 | 1.533339e-01 | (12) |
| $_{12}C_1 + _{10}C_1$ | $\rightarrow$ | $_7C_1 + _3C_1$ | 2.060491e-01 | 1.619034e-02 | (13) |
| $_8C_1 + _3C_1$ | $\rightarrow$ | $_2C_1 + C_1$ | 5.617656e-02 | 3.149758e-01 | (14) |
| $_{11}C_1 + _5C_1$ | $\rightarrow$ | $_{10}C_1 + _2C_1$ | 1.547249e-02 | 4.123857e-01 | (15) |
| $_{11}C_1 + _3C_1$ | $\rightarrow$ | $_{11}C_1 + _2C_1$ | 7.166242e-02 | 6.899306e-01 | (16) |
| $_8C_1 + _3C_1$ | $\rightarrow$ | $_5C_1 + C_1$ | 5.642374e-02 | 1.104936e-01 | (17) |
| $_{10}C_1 + _7C_1$ | $\rightarrow$ | $_9C_1 + _9C_1$ | 5.669881e-01 | 5.319954e-01 | (18) |
| $_{11}C_1 + _4C_1$ | $\rightarrow$ | $_7C_1 + _7C_1$ | 5.292460e-01 | 5.490419e-01 | (19) |
| $_8C_1 + _5C_1$ | $\rightarrow$ | $_6C_1 + C_1$ | 8.021296e-02 | 2.273137e-02 | (20) |
| $_7C_1 + _7C_1$ | $\rightarrow$ | $_5C_1 + _4C_1$ | 3.222952e-01 | 5.230033e-01 | (21) |
| $_{12}C_1 + _3C_1$ | $\rightarrow$ | $_{10}C_1 + _2C_1$ | 9.906861e-03 | 6.813578e-01 | (22) |
| $_{12}C_1 + _4C_1$ | $\rightarrow$ | $_6C_1 + C_1$ | 6.811983e-01 | 6.459914e-01 | (23) |
| $_{11}C_1 + _6C_1$ | $\rightarrow$ | $_8C_1 + _7C_1$ | 3.109788e-02 | 6.540693e-01 | (24) |

Table B.4: **Sample artificial chemistry with oscillating dynamics.** Chemistry was found by generating and simulating a large number of artificial reaction networks ($\approx$ 10000) and manually looking at heuristically preselected candidates. Forward reaction constants $k^f$ and backward reaction constants $k^b$ have been calculated from random energies as shown in Sect. 3.1.2.

| | | | $k^f$ | $k^b$ | |
|---|---|---|---|---|---|
| $hv + {}_3C_1$ | $\rightarrow$ | ${}_5C_1$ | 3.493301e-01 | 3.515938e-23 | (1) |
| $hv + C_1$ | $\rightarrow$ | ${}_4C_1$ | 5.915973e-01 | 2.339289e-22 | (2) |
| $hv + {}_2C_1$ | $\rightarrow$ | ${}_5C_1$ | 3.303608e-01 | 2.845655e-22 | (3) |
| $hv + {}_2C_1$ | $\rightarrow$ | ${}_{11}C_1$ | 1.166197e-01 | 1.583525e-23 | (4) |
| ${}_{10}C_1 + {}_8C_1$ | $\rightarrow$ | ${}_{10}C_1 + {}_4C_1$ | 8.491597e-02 | 3.687288e-01 | (5) |
| ${}_9C_1 + {}_3C_1$ | $\rightarrow$ | ${}_9C_1 + {}_7C_1$ | 4.186161e-01 | 1.650730e-01 | (6) |
| ${}_9C_1 + C_1$ | $\rightarrow$ | ${}_9C_1 + {}_8C_1$ | 3.284513e-01 | 1.550726e-01 | (7) |
| ${}_{12}C_1 + {}_9C_1$ | $\rightarrow$ | ${}_{12}C_1 + {}_6C_1$ | 1.629401e-01 | 7.104541e-01 | (8) |
| ${}_7C_1 + {}_6C_1$ | $\rightarrow$ | ${}_6C_1 + {}_5C_1$ | 5.519999e-01 | 7.304799e-01 | (9) |
| ${}_9C_1$ | $\rightarrow$ | ${}_5C_1$ | 1.340816e-02 | 7.331016e-02 | (10) |
| ${}_{12}C_1 + {}_5C_1$ | $\rightarrow$ | ${}_6C_1 + {}_5C_1$ | 6.272412e-01 | 7.496179e-01 | (11) |
| ${}_4C_1$ | $\rightarrow$ | ${}_2C_1$ | 5.095917e-01 | 5.181876e-02 | (12) |
| ${}_{10}C_1 + {}_7C_1$ | $\rightarrow$ | $C_1 + C_1$ | 2.690245e-01 | 6.716334e-01 | (13) |
| ${}_{12}C_1 + {}_2C_1$ | $\rightarrow$ | ${}_{10}C_1 + {}_4C_1$ | 1.802030e-02 | 1.624015e-01 | (14) |
| ${}_{11}C_1 + {}_3C_1$ | $\rightarrow$ | ${}_{10}C_1 + {}_2C_1$ | 2.058409e-02 | 9.330057e-03 | (15) |
| ${}_{12}C_1 + {}_{11}C_1$ | $\rightarrow$ | ${}_8C_1 + {}_8C_1$ | 6.899998e-03 | 1.686871e-02 | (16) |
| ${}_{12}C_1 + {}_5C_1$ | $\rightarrow$ | ${}_9C_1 + {}_3C_1$ | 5.269599e-01 | 2.767870e-01 | (17) |
| ${}_{11}C_1 + {}_9C_1$ | $\rightarrow$ | ${}_4C_1 + {}_3C_1$ | 5.275851e-03 | 7.721767e-01 | (18) |
| ${}_8C_1 + {}_7C_1$ | $\rightarrow$ | $C_1 + C_1$ | 2.467435e-01 | 7.428251e-01 | (19) |
| ${}_{12}C_1 + {}_5C_1$ | $\rightarrow$ | ${}_8C_1 + {}_3C_1$ | 5.095817e-02 | 7.421245e-02 | (20) |
| ${}_{11}C_1 + {}_7C_1$ | $\rightarrow$ | ${}_8C_1 + {}_4C_1$ | 1.581391e-03 | 1.482405e-02 | (21) |
| ${}_5C_1 + {}_2C_1$ | $\rightarrow$ | ${}_4C_1 + {}_3C_1$ | 9.786585e-03 | 1.844324e-01 | (22) |
| ${}_{10}C_1 + {}_3C_1$ | $\rightarrow$ | ${}_9C_1 + C_1$ | 2.485208e-01 | 4.166127e-02 | (23) |
| ${}_8C_1 + C_1$ | $\rightarrow$ | ${}_4C_1 + {}_3C_1$ | 8.697111e-04 | 6.737947e-03 | (24) |

## B.3 NETWORK FILE FORMAT

This section contains the definition of the file format that is used to process reaction networks for this thesis. The definition and respective source code is also contained in the digital appendix (Appendix C) and available online on GitHub. [6]

[6] https://github.com/jakob-fischer/jrnf_R_tools

The *jrnf file format* describes a reaction network consisting out of a set of chemical species, reactions between those species and (thermo)dynamic parameters. The networks are contained in plain Unix-format text files (line feed character for new line) with only ASCII characters. Individual elements are unsigned integer or floating point numbers in plain text or character strings. They are separated by whitespaces.

The first line of a file should contain the string "jrnf0003" to identify the file format and version:

| | |
|---|---|
| **1** | "jrnf0003" |

The second line contains two integers, the number of chemical species N and the number of reactions M.

| | |
|---|---|
| **2** | ⟨N⟩⟨M⟩ |

The following N lines each have 4 values (separated by whitespaces). Each line represents one chemical species and has the following form:

| | |
|---|---|
| **3 . . . 2 +** ⟨**N**⟩ | ⟨type⟩⟨name⟩⟨energy⟩⟨constant⟩ |

The entries each consist of a number or a string.

| entry | type | description |
|---|---|---|
| **type** | unsigned integer | Contains the type of the species. Standard is 0, but users may use / define additional types. A not recognized type should just be ignored by programs (treated as type 0). |
| **name** | string | Name of the species. May contain alphanumeric characters and underscores. Name should be unique in the network. |
| **energy** | double | Value containing the formation enthalpy of the species. |
| **constant** | unsigned integer | Indicating if the species is constant (buffered by environment). Value is 1 if the species is constant and 0 else. |

The next M lines each describe one of the M reactions. The exact number of entries for each line depends on the complexity of the reaction. Each line has at least 7 entries.

| | |
|---|---|
| $3 + \mathbf{N} \ldots$ | ⟨reversible⟩ ⟨c⟩ ⟨$k^f$⟩ ⟨$k^b$⟩ ⟨$E_a$⟩⟨educt_number⟩ |
| $2 + \mathbf{N} + \mathbf{M}$ | ⟨product_number⟩ (⟨educt_id⟩ ⟨educt_mul⟩)$*$ (⟨product_id⟩ ⟨product_mul⟩)$*$ |

The pattern in the first bracket occurs exactly ⟨educt_number⟩ times, the pattern in the second brackets occurs ⟨product_number⟩ times.

| entry | type | description |
|---|---|---|
| **reversible** | unsigned integer | The value is '1' if reaction is simulated reversible. Else the value has to be 'o'. |
| **c** | double (positive) | Reaction constant for stochastic simulations. Not used at the moment. |
| $\mathbf{k^f}$ | double (positive) | Forward reaction rate. |
| $\mathbf{k^b}$ | double (positive) | Backward reaction rate. |
| $\mathbf{E_a}$ | double (positive) | Activation energy. (Please note remark below!) |
| **educt_number** | unsigned integer | Number of educts in the reaction equation. |
| **product_number** | unsigned integer | Number of products in the reaction equation. |
| **educt_id** | unsigned integer | Id of the species occurring on educt side of this reaction equation. Species are o-indexed. |
| **educt_mul** | unsigned integer | Number of occurrences of respective species on the educt side of this reaction equation. |
| **product_id** | unsigned integer | Id of the species occurring on product side of this reaction equation. Species are o-indexed. |
| **product_mul** | unsigned integer | Number of occurrences of respective species on the product side of this reaction equation. |

Reaction rates ($k^f$ and $k^b$) and energies (formation enthalpy & activation energy) are complementary. The specification does not require for them to have the same dynamics. Applications are free to specify which one to use. For reversible reactions the activation energy refers to that reaction direction which respective products have a higher Gibbs energy of formation (formation enthalpy). See Sect. 3.1.2 for details.

# C

# DIGITAL APPENDIX

## C.1 DIGITAL CONTENT

Selected data of this thesis as well as the thesis itself in digital form is contained on the accompanying data medium. The directory structure of the medium is shown in Fig. C.1. The data is ordered according to the chapters it is related to, with one extra folder (*code/*) for the tools and scripts that are used multiple times. These tools have

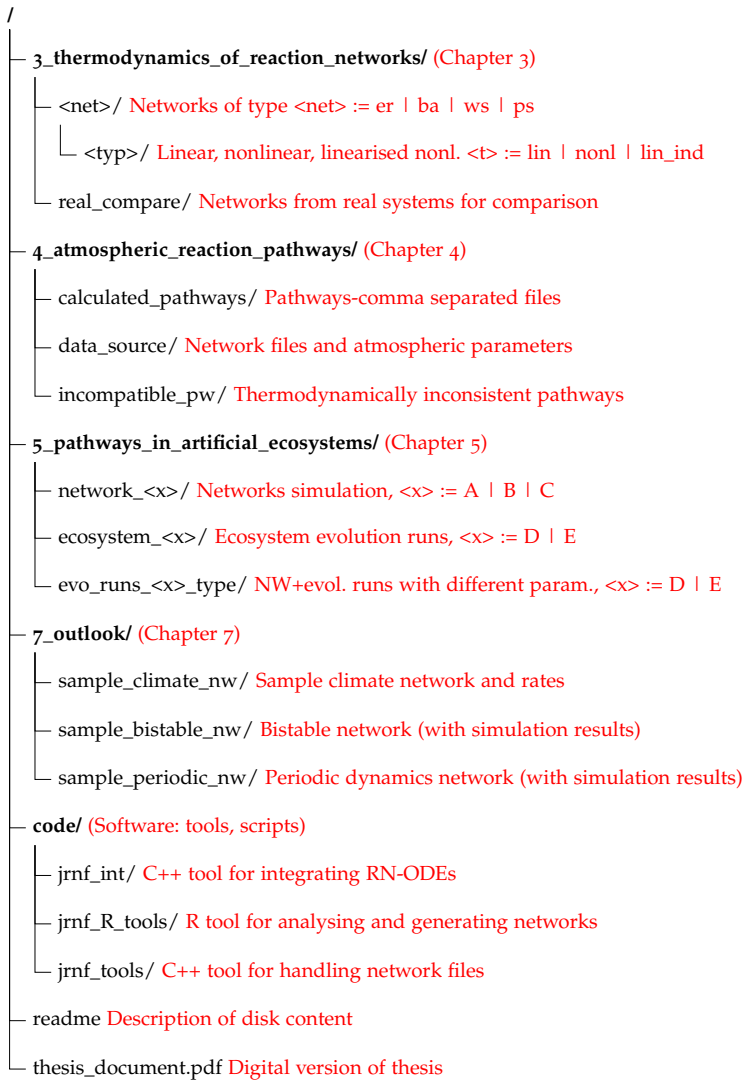



Figure C.1: **Directory structure of digital appendix.** Directories starting with numbers contain reaction networks, simulation data and scripts related to the respective chapter. Scripts are mostly in R language, networks in jrnf format (see Sect. B.3) and Data as R data files (*.Rdata*). Every directories for data of the individual chapters contain a separate readme file describing its content.

been published previously ([Fis16a; Fis16b; Fis16c]). Future releases should be available through the author's GitHub page[1]. All tools use the jrnf file format for exchanging reaction networks (Sect. B.3).

[1] http://www.github.com/jakob-fischer/

Generation of large artificial networks for Chapter 3 was done with the C++-tool in *code/jrnf_tools*. The artificial networks and ecosystems for Chapter 5 were generated with the R-tool *code/jrnf_R_tools*. Generally, simulation results (intermediates) are contained in the medium. But because of space limitation for the simulations of large-scale reaction networks of Chapter 3 only the network files are contained in the digital appendix.

The two C++-tools (*jrnf_tools*, *jrnf_int*) have been developed and tested with gcc (version 6.3). All the R-scripts were tested with version 3.3 of *R* and require the *ggplot2* and *igraph* libraries (tested with version 1.0.1 of *igraph*).

# BIBLIOGRAPHY


[Ada02] C Adami. "What is complexity?" In: *BioEssays* 24.12 (2002), pp. 1085–1094. DOI: 10.1002/bies.10192.

[AOC00] C Adami, C Ofria, and T C Collier. "Evolution of biological complexity". In: *P Natl Acad Sci* 97.9 (2000), pp. 4463–4468. DOI: 10.1073/pnas.97.9.4463.

[AGB15] E Agmon, A J Gates, and R D Beer. "Ontogeny and adaptivity in a model protocell". In: *Adv Artif Life, ECAL*. Vol. 13. 2015, pp. 216–223. DOI: 10.7551/978-0-262-33027-5-ch043.

[AM11] K Ahnert and M Mulansky. "Odeint - Solving Ordinary Differential Equations in C++". In: *American Institute of Physics Conference Series*. Ed. by T E Simos, G Psihoyios, C Tsitouras, and Z Anastassi. Vol. 1389. American Institute of Physics Conference Series. Sept. 2011, pp. 1586–1589. DOI: 10.1063/1.3637934.

[AB02] R Albert and A L Barabási. "Statistical mechanics of complex networks". In: *Rev Mod Phys* 74.January (2002), pp. 47–97. DOI: 10.1103/RevModPhys.74.47.

[And+14] J L Andersen, C Flamm, D Merkle, and P F Stadler. "50 Shades of Rule Composition". In: *International Conference on Formal Methods in Macro-Biology*. Ed. by Piazza C Fages F. Vol. 8738. Springer. 2014, pp. 117–135. DOI: 10.1007/978-3-319-10398-3_9.

[Auf+11] A K Aufdenkampe, E Mayorga, P A Raymond, J M Melack, S C Doney, S R Alin, R E Aalto, and K Yoo. "Riverine coupling of biogeochemical cycles between land, oceans, and atmosphere". In: *Front Ecol Environ* 9.1 (2011), pp. 53–60. DOI: 10.1890/100014.

[BF92] R J Bagley and J D Farmer. "Spontaneous Emergence of a Metabolism". In: *Artificial Life II*. Ed. by C G Langton, Charles Taylor, J D Farmer, and S Rasmussen. Redwood City, CA: Addison-Wesley, 1992, pp. 93–140.

[BC15] R Banisch and N D Conrad. "Cycle-flow–based module detection in directed recurrence networks". In: *Europhys Lett* 108.6 (2015), p. 68008. DOI: 10.1209/0295-5075/108/68008.

[Bar01] G W Barrett. "Closing the ecological cycle: the emergence of integrative science". In: *Ecosystem Health* 7.2 (2001), pp. 79–84. DOI: 10.1046/j.1526-0992.2001.007002079.x.

[BB15] S Bartlett and S Bullock. "Emergence of Competition between Different Dissipative Structures for the Same Free Energy Source". In: *Adv Artif Life, ECAL*. Vol. 13. 2015, pp. 415–422. DOI: 10.7551/978-0-262-33027-5-ch0743.

[BLQ02] D A Beard, S Liang, and H Qian. "Energy balance for analysis of complex metabolic networks". In: *Biophys J* 83.1 (2002), pp. 79–86. DOI: 10.1016/S0006-3495(02)75150-3.

[BQ05] D A Beard and H Qian. "Thermodynamic-based computational profiling of cellular regulatory control in hepatocyte metabolism". In: *Am J Physiol Endocrinol Metab* 288.3 (2005), E633–E644. DOI: 10.1152/ajpendo.00239.2004.

[BQ07] D A Beard and H Qian. "Relationship between thermodynamic driving force and one-way fluxes in reversible processes". In: *PLOS ONE* 2.1 (2007), e144. DOI: 10.1371/journal.pone.0000144.

[Bea+04] D A Beard, E Babson, E Curtis, and H Qian. "Thermodynamic constraints for biochemical networks". In: *J Theor Biol* 228.3 (2004), pp. 327–333. DOI: 10.1016/j.jtbi.2004.01.008.

[BPW06] A P Beckerman, O L Petchey, and P H Warren. "Foraging biology predicts food web complexity". In: *P Natl Acad Sci* 103.37 (2006), pp. 13745–13749. DOI: 10.1073/pnas.0603039103.

[Bed03] M A Bedau. "Artificial life: organization, adaptation and complexity from the bottom up". In: *Trends Cogn Sci* 7.11 (2003), pp. 505–512.

[Bed+00] M A Bedau, J S McCaskill, N H Packard, S Rasmussen, C Adami, D G Green, T Ikegami, K Kaneko, and T S Ray. "Open problems in artificial life". In: *Artif Life* 6.4 (2000), pp. 363–376. DOI: 10.1162/106454600300103683.

[Bee+10] C Beer, M Reichstein, E Tomelleri, P Ciais, M Jung, N Carvalhais, C Rödenbeck, M A Arain, D Baldocchi, G B Bonan, et al. "Terrestrial gross carbon dioxide uptake: global distribution and covariation with climate". In: *Science* 329.5993 (2010), pp. 834–838. DOI: 10.1126/science.1184984.

[BFS03a] G Benkö, C Flamm, and P F Stadler. "A graph-based toy model of chemistry". In: *J Chem Inf Comput Sci* 43.4 (2003), pp. 1085–1093. DOI: 10.1021/ci0200570.

[BFS03b] G Benkö, C Flamm, and P F Stadler. "Generic properties of chemical networks: Artificial chemistry based on graph rewriting". In: *Adv Artif Life, ECAL*. Springer. 2003, pp. 10–19. DOI: 10.1007/978-3-540-39432-7_2.

[Ben10] S A Benner. "Defining life". In: *Astrobiology* 10.10 (2010), pp. 1021–1030. DOI: 10.1089/ast.2010.0524.

[Ber99] R A Berner. "A new look at the long-term carbon cycle". In: *GSA Today* 9.11 (1999), pp. 1–6.

[Ber03] R A Berner. "The long-term carbon cycle, fossil fuels and atmospheric composition". In: *Nature* 426.6964 (2003), pp. 323–326. DOI: 10.1038/nature02131.

[BS02] H G Beyer and H P Schwefel. "Evolution strategies–A comprehensive introduction". In: *Nat Comp* 1.1 (2002), pp. 3–52.

[Can14] W R Cannon. "Simulating metabolism with statistical thermodynamics". In: *PLOS ONE* 9.8 (2014), e103582. DOI: 10.1371/journal.pone.0103582.

[CN06a] A G Cantú and G Nicolis. "Toward a Thermodynamic Characterization of Chemical Reaction Networks". In: *J Non-Equil Thermodyn* 31.1 (2006), pp. 23–46. DOI: 10.1515/JNETDY.2006.003.

[CK07] D Catling and J F Kasting. "Planetary atmospheres and life". In: *Planets and Life: The Emerging Science of Astrobiology*. Ed. by W T Sullivan and J A Baross. Cambridge, UK: Cambridge University Press, 2007, pp. 91–116. DOI: 10.1016/b978-0-444-53802-4.00185-8.

[Cec12] T R Cech. "The RNA worlds in context". In: *Cold Spring Harb Perspect Biol* 4.7 (2012), a006742. DOI: 10.1101/cshperspect.a006742.

[CD07] F Centler and P Dittrich. "Chemical organizations in atmospheric photochemistries—A new method to analyze chemical reaction networks". In: *Planet Space Sci* 55.4 (2007), pp. 413–428. DOI: 10.1016/j.pss.2006.08.002.

[Cha+13] A Chakrabarti, L Miskovic, K C Soh, and V Hatzimanikatis. "Towards kinetic modeling of genome-scale metabolic networks without sacrificing stoichiometric, thermodynamic and physiological constraints". In: *Biotechnol J* 8.9 (2013), pp. 1043–1057. DOI: 10.1002/biot.201300091.

[CI+08] F S Chapin III, J T Randerson, A D McGuire, J A Foley, and C B Field. "Changing feedbacks in the climate-biosphere system". In: *Front Ecol Environ* 6.6 (2008), pp. 313–320. DOI: 10.1890/080005.

[CL16] A Chopra and C H Lineweaver. "The Case for a Gaian Bottleneck: The Biology of Habitability". In: *Astrobiology* 16.1 (2016), pp. 7–22. DOI: 10.1089/ast.2015.1387.

[Cia+14] P Ciais, C Sabine, G Bala, L Bopp, V Brovkin, J Canadell, A Chhabra, R DeFries, J Galloway, M Heimann, et al. "Carbon and other biogeochemical cycles". In: *Climate change 2013: the physical science basis. Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change*. Cambridge University Press, 2014, pp. 465–570.

[CHS17] E B Clark, S J Hickinbotham, and S Stepney. "Semantic closure demonstrated by the evolution of a universal constructor architecture in an artificial chemistry". In: *J R Soc Interface* 14.130 (2017), p. 20161033. DOI: 10.1098/rsif.2016.1033.

[Cla88] B L Clarke. "Stoichiometric network analysis". In: *Cell Biochem Biophys* 12 (1988), pp. 237–253. DOI: 10.1007/bf02918360.

[CC02] C E Cleland and C F Chyba. "Defining 'life'". In: *Origins Life Evol B* 32.4 (2002), pp. 387–393.

[CSM07] S D Copley, E Smith, and H J Morowitz. "The origin of the RNA world: co-evolution of genes and metabolism". In: *Bioorg Chem* 35.6 (2007), pp. 430–443. DOI: 10.1016/j.bioorg.2007.08.001.

[CTF06] G Craciun, Y Tang, and M Feinberg. "Understanding bistability in complex enzyme-driven reaction networks". In: *P Natl Acad Sci* 103.23 (2006), pp. 8697–8702. DOI: 10.1073/pnas.0602767103.

[CN06b] G Csardi and T Nepusz. "The igraph software package for complex network research". In: *InterJournal, Complex Systems* 1695.5 (2006), pp. 1–9.

[DT02] A Dai and K E Trenberth. "Estimates of freshwater discharge from continents: Latitudinal and seasonal variations." In: *J Hydrometeorol* 3.6 (2002). DOI: 10.1175/1525-7541(2002)003<0660:EOFDFC>2.0.CO;2.

[DM13] D De Martino. "Thermodynamics of biochemical networks and duality theorems". In: *Phys Rev E* 87.5 (2013), p. 052108. DOI: 10.1103/PhysRevE.87.052108.

[De +12] D De Martino, M Figliuzzi, A De Martino, and E Marinari. "A scalable algorithm to explore the Gibbs energy landscape of genome-scale metabolic networks." In: *PLoS Comput Biol* 8.6 (2012), e1002562. DOI: 10.1371/journal.pcbi.1002562.

[DM05] D L DeAngelis and W M Mooij. "Individual-based modeling of ecological and evolutionary processes 1". In: *Annu Rev Ecol Evol Syst* 36 (2005), pp. 147–168. DOI: 10.1146/annurev.ecolsys.36.102003.152644.

[DF07] P Dittrich and P S di Fenizio. "Chemical organisation theory." In: *Bull Math Biol* 69.4 (2007), pp. 1199–231. DOI: 10.1007/s11538-006-9130-8.

[DZB01] P Dittrich, J Ziegler, and W Banzhaf. "Artificial chemistries—a review". In: *Artif Life* 7.3 (2001), pp. 225–275. DOI: 10.1162/106454601753238636.

[Djo+17] T Djokic, M J Van Kranendonk, K A Campbell, M R Walter, and C R Ward. "Earliest signs of life on land preserved in ca. 3.5 Ga hot spring deposits." In: *Nat Commun* 8 (2017), p. 15263. DOI: 10.1038/ncomms15263.

[Dod+17] M S Dodd, D Papineau, T Grenne, J F Slack, M Rittner, F Pirajno, J O'Neil, and C T S Little. "Evidence for early life in Earth's oldest hydrothermal vent precipitates". In: *Nature* 543.7643 (2017), pp. 60–64. DOI: 10.1038/nature21377.

[DF11] D Dommenget and J Flöter. "Conceptual understanding of climate change with a globally resolved energy balance model". In: *Clim Dynam* 37.11-12 (2011), pp. 2143–2165. DOI: 10.1007/s00382-011-1026-0.

[Don+09] J F Donges, Y Zou, N Marwan, and J Kurths. "Complex networks in climate dynamics". In: *Eur Phys J Spec Top* 174.1 (2009), pp. 157–179. DOI: 10.1140/epjst/e2009-01098-2.

[Dön+15] M Döntgen, Marie-Dominique Przybylski-Freund, Leif C Kröger, Wassja A Kopp, Ahmed E Ismail, and Kai Leonhard. "Automated Discovery of Reaction Pathways, Rate Constants, and Transition States Using Reactive Molecular Dynamics Simulations". In: *J Chem Theory Comput* 11.6 (2015), pp. 2517–2524. DOI: 10.1021/acs.jctc.5b00201.

[DK07] A Dorin and K B Korb. "Building virtual ecosystems from artificial chemistry". In: *Adv Artif Life, ECAL*. Springer, 2007, pp. 103–112. DOI: 10.1007/978-3-540-74913-4_11.

[DKG08] A Dorin, K B Korb, and V Grimm. "Artificial-Life Ecosystems-What are they and what could they become?" In: *Artificial Life XI*. 2008, pp. 173–180.

[Dou11] A P Douce. *Thermodynamics of the Earth and Planets*. Cambridge, UK: Cambridge University Press, 2011. DOI: 10.1017/cbo9780511974854.

[Dow02] K L Downing. "The simulated emergence of distributed environmental control in evolving microcosms". In: *Artif Life* 8.2 (2002), pp. 123–153. DOI: 10.1162/106454602320184211.

[DZ99] K Downing and P Zvirinsky. "The simulated evolution of biochemical guilds: reconciling Gaia theory and natural selection". In: *Artif Life* 5.4 (1999), pp. 291–318. DOI: 10.1162/106454699568791.

[Duf+07] J E Duffy, B J Cardinale, K E France, P B McIntyre, E Thébault, and M Loreau. "The functional role of biodiversity in ecosystems: incorporating trophic complexity". In: *Ecol Lett* 10.6 (2007), pp. 522–538. DOI: 10.1111/j.1461-0248.2007.01037.x.

[Dyk08] J G Dyke. "Entropy production in an energy balance Daisyworld model". In: *Artificial Life XI*. 2008, p. 189.

[DK10] J G Dyke and A Kleidon. "The maximum entropy production principle: Its theoretical foundations and applications to the earth system". In: *Entropy* 12.3 (2010), pp. 613–630. DOI: 10.3390/e12030613.

[Egb13] M D Egbert. "Bacterial chemotaxis: Introverted or extroverted? A comparison of the advantages and disadvantages of basic forms of metabolism-based and metabolism-independent behavior using a computational model". In: *PLOS ONE* 8.5 (2013), e63617. DOI: 10.1371/journal.pone.0063617.

[Eib14] A E Eiben. "Grand challenges for evolutionary robotics". In: *Front Robot AI* 1 (2014), p. 4. DOI: 10.3389/frobt.2014.00004.

[ES78] M Eigen and P Schuster. "The hypercycle". In: *Naturwissenschaften* 65.1 (1978), pp. 7–41.

[ES96] I R Epstein and K Showalter. "Nonlinear chemical dynamics: oscillations, patterns, and chaos". In: *J Phys Chem* 100.31 (1996), pp. 13132–13147. DOI: 10.1021/jp953547m.

[ER59] P Erdős and A Rényi. "On random graphs I." In: *Publ Math-Debrecen* 6 (1959), pp. 290–297.

[Est12] E Estrada. "Returnability as a criterion of disequilibrium in atmospheric reactions networks". In: *J Math Chem* 50.6 (2012), pp. 1363–1372. DOI: 10.1007/s10910-012-9977-x.

[Fae+05] J R Faeder, M L Blinov, B Goldstein, and W S Hlavacek. "Rule-based modeling of biochemical networks". In: *Complexity* 10.4 (2005), pp. 22–41. DOI: 10.1002/cplx.20074.

[Fal+00] P Falkowski, R J Scholes, E E A Boyle, J Canadell, D Canfield, J Elser, N Gruber, K Hibbard, P Högberg, S Linder, et al. "The global carbon cycle: a test of our knowledge of earth as a system". In: *Science* 290.5490 (2000), pp. 291–296. DOI: 10.1126/science.290.5490.291.

[Fie+98] C B Field, M J Behrenfeld, J T Randerson, and P Falkowski. "Primary production of the biosphere: integrating terrestrial and oceanic components". In: *Science* 281.5374 (1998), pp. 237–240. DOI: 10.1126/science.281.5374.237.

[Fis16a] J Fischer. *jrnf_R_tools: A tool for handling of reaction network analysis and simulation*. Version 1.0.0. Dec. 2016. DOI: 10.5281/zenodo.197056.

[Fis16b] J Fischer. *jrnf_int: Integrator for reaction networks in jrnf format*. Version 1.0.0. Dec. 2016. DOI: 10.5281/zenodo.197058.

[Fis16c] J Fischer. *jrnf_tools: Command line tool for generating large random reaction networks*. Version 1.0.0. Dec. 2016. DOI: 10.5281/zenodo.197060.

[FKD15] J Fischer, A Kleidon, and P Dittrich. "Thermodynamics of Random Reaction Networks". In: *PLOS ONE* 10.2 (2015), e0117312. DOI: 10.1371/journal.pone.0117312.

[Fis01] D Fiscus. "The ecosystemic life hypothesis I: introduction and definitions". In: *Bull Ecol Soc Am* 82.4 (2001), pp. 248–250.

[Fri10] K Friston. "The free-energy principle: a unified brain theory?" In: *Nat Rev Neurosci* 11.2 (2010), pp. 127–138. DOI: 10.1038/nrn2787.

[Geb+17] S Gebauer, J L Grenfell, J W Stock, R Lehmann, M Godolt, P von Paris, and H Rauer. "Evolution of Earth-like Extrasolar Planetary Atmospheres: Assessing the Atmospheres and Biospheres of Early Earth Analog Planets with a Coupled Atmosphere Biogeochemical Model". In: *Astrobiology* 17.1 (2017), pp. 27–54. DOI: 10.1089/ast.2015.1384.

[GZ08] A S Ginzburg and N N Zavalishin. "Dynamics of a closed low-parameter compartment model of the global carbon cycle". In: *Izv Atmos Ocean Phys* 44.6 (2008), pp. 684–700. DOI: 10.1134/S0001433808060029.

[Gle+01] P M Gleiss, P F Stadler, A Wagner, and D A Fell. "Relevant cycles in chemical reaction networks". In: *Adv Complex Syst* 4.02n03 (2001), pp. 207–226. DOI: 10.1142/s0219525901000140.

[GLW06] C Goldblatt, T M Lenton, and Andrew J Watson. "Bistability of atmospheric oxygen and the Great Oxidation". In: *Nature* 443.7112 (2006), pp. 683–686. DOI: 10.1038/nature05169.

[Gor91] E Gorham. "Biogeochemistry: its origins and development". In: *Biogeochemistry* 13.3 (1991), pp. 199–239. DOI: 10.1007/BF00002942.

[Gra+09] R Gras, D Devaurs, A Wozniak, and A Aspinall. "An individual-based evolving predator-prey ecosystem simulation using a fuzzy cognitive map as the behavior model". In: *Artif Life* 15.4 (2009), pp. 423–463. DOI: 10.1162/artl.2009.Gras.012.

[Gra+14] T Grassi, S Bovino, D R G Schleicher, J Prieto, D Seifried, E Simoncini, and F A Gianturco. "KROME–a package to embed chemistry in astrophysical simulations". In: *Mon Not R Astron Soc* 439.3 (2014), pp. 2386–2419. DOI: 10.1093/mnras/stu114.

[GS83] P Gray and S K Scott. "Autocatalytic reactions in the isothermal, continuous stirred tank reactor: isolas and other forms of multistability". In: *Chem Eng Sci* 38.1 (1983), pp. 29–43. DOI: 10.1016/0009-2509(83)80132-8.

[GRP13] J L Grenfell, H Rauer, and P von Paris. "Exoplanets: criteria for their habitability and possible biospheres". In: *Habitability of Other Planets and Satellites*. Springer, 2013, pp. 13–29. DOI: 10.1007/978-94-007-6546-7_2.

[Gre+06] J L Grenfell, R Lehmann, P Mieth, U Langematz, and B Steil. "Chemical reaction pathways affecting stratospheric and mesospheric ozone". In: *J Geophys Res-Atmos* 111.D17 (2006). DOI: 10.1029/2004jd005713.

[Gre+13] J L Grenfell, S Gebauer, M Godolt, K Palczynski, H Rauer, J Stock, P Von Paris, R Lehmann, and F Selsis. "Potential biosignatures in super-Earth atmospheres II. Photochemical responses". In: *Astrobiology* 13.5 (2013), pp. 415–438. DOI: 10.1089/ast.2012.0926.

[GG08] N Gruber and J N Galloway. "An Earth-system perspective of the global nitrogen cycle". In: *Nature* 451.7176 (2008), pp. 293–296. DOI: 10.1038/nature06592.

[HTS15] M Harada, E Tajika, and Y Sekine. "Transition to an oxygen-rich atmosphere with an extensive overshoot triggered by the Paleoproterozoic snowball Earth". In: *Earth Planet Sci Lett* 419 (2015), pp. 178–186. DOI: 10.1016/j.epsl.2015.03.005.

[Hed92] J I Hedges. "Global biogeochemical cycles: progress and problems". In: *Mar Chem* 39.1-3 (1992), pp. 67–93. DOI: 10.1016/0304-4203(92)90096-S.

[HBHk07] C S Henry, L J Broadbelt, and V Hatzimanikatis. "Thermodynamics-based metabolic flux analysis". In: *Biophys J* 92.5 (2007), pp. 1792–1805. DOI: 10.1529/biophysj.106.093138.

[HA08] A Hintze and C Adami. "Evolution of complex modular biological networks". In: *PLoS Comput Biol* 4.2 (2008), e23. DOI: 10.1371/journal.pcbi.0040023.

[HL67] D R Hitchcock and J E Lovelock. "Life detection by atmospheric analysis". In: *Icarus* 159 (1967), pp. 149–159. DOI: 10.1016/0019-1035(67)90059-0.

[Hol06] H D Holland. "The oxygenation of the atmosphere and oceans". In: *Philos T Roy Soc B* 361.1470 (June 2006), pp. 903–15. DOI: 10.1098/rstb.2006.1838.

[HHS10] W Hordijk, J Hein, and M Steel. "Autocatalytic sets and the origin of life". In: *Entropy* 12.7 (2010), pp. 1733–1742. DOI: 10.3390/e12071733.

[HSS15] W Hordijk, J I Smith, and M Steel. "Algorithms for detecting and analysing autocatalytic sets". In: *Algorithms Mol Biol* 10.1 (2015), p. 1. DOI: 10.1186/s13015-015-0042-8.

[HS04] W Hordijk and M Steel. "Detecting autocatalytic, self-sustaining sets in chemical reaction systems". In: *J Theor Biol* 227.4 (2004), pp. 451–461. DOI: 10.1016/j.jtbi.2003.11.020.

[HS12] W Hordijk and M Steel. "Autocatalytic sets extended: Dynamics, inhibition, and a generalization". In: *J Syst Chem* 3.1 (2012), pp. 1–12. DOI: 10.1186/1759-2208-3-5.

[HSK13] W Hordijk, M Steel, and S Kauffman. "The Origin of Life, Evolution, and Functional Organization". In: *Evolutionary Biology: Exobiology and Evolutionary Mechanisms*. Springer, 2013, pp. 49–60. DOI: 10.1007/978-3-642-38212-3_4.

[HJ72] F Horn and R Jackson. "General mass action kinetics". In: *Arch Rational Mech Anal* 47.2 (1972), pp. 81–116. DOI: 10.1007/BF00251225.

[IG96] R Ihaka and R Gentleman. "R: a language for data analysis and graphics". In: *J Comput Graph Stat* 5.3 (1996), pp. 299–314. DOI: 10.2307/1390807.

[JK98] S Jain and S Krishna. "Autocatalytic sets and the growth of complexity in an evolutionary model". In: *Phys Rev Lett* 81.25 (1998), p. 5684. DOI: 10.1103/PhysRevLett.81.5684.

[JK01] S Jain and S Krishna. "A model for the emergence of cooperation, interdependence, and structure in evolving networks". In: *P Natl Acad Sci* 98.2 (2001), pp. 543–547. DOI: 10.1073/pnas.98.2.543.

[JD10] C C Jolley and T Douglas. "a Network-Theoretical Approach To Understanding Interstellar Chemistry". In: *Astrophys J* 722.2 (2010), pp. 1921–1931. DOI: 10.1088/0004-637X/722/2/1921.

[JD12] C Jolley and T Douglas. "Topological biosignatures: Large-scale structure of chemical networks from biology and astrochemistry". In: *Astrobiology* 12.1 (2012), pp. 29–39. DOI: 10.1089/ast.2011.0674.

[JS13] B Joshi and A Shiu. "Atoms of multistationarity in chemical reaction networks". In: *J Math Chem* 51.1 (2013), pp. 153–178. DOI: 10.1007/s10910-012-0072-0.

[Kai+00] E W Kaiser, T J Wallington, M D Hurley, J Platz, H J Curran, W J Pitz, and C K Westbrook. "Experimental and modeling study of premixed atmospheric-pressure dimethyl ether-air flames". In: *J Phys Chem A* 104.35 (2000), pp. 8194–8206. DOI: 10.1021/jp994074c.

[Kas97] J F Kasting. "Habitable zones around low mass stars and the search for extraterrestrial life". In: *Planetary and interstellar processes relevant to the origins of life*. Springer, 1997, pp. 291–307. DOI: 10.1007/978-94-015-8907-9_15.

[KHP85] J F Kasting, H D Holland, and J P Pinto. "Oxidant abundances in rainwater and the evolution of atmospheric oxygen". In: *J Geophys Res-Atmos* 90.10 (1985), pp. 497–510. DOI: 10.1029/jd090id06p10497.

[Kau86] S A Kauffman. "Autocatalytic sets of proteins". In: *J Theor Biol* 119.1 (1986), pp. 1–24. DOI: 10.1016/S0022-5193(86)80047-9.

[KKS05] P Kharecha, J Kasting, and J Siefert. "A coupled atmosphere–ecosystem model of the early Archean Earth". In: *Geobiology* 3.2 (2005), pp. 53–76. DOI: 10.1111/j.1472-4669.2005.00049.x.

[KG12] M Khater and R Gras. "Adaptation and genomic evolution in ecosim". In: *International Conference on Simulation of Adaptive Behavior*. Springer. 2012, pp. 219–229. DOI: 10.1007/978-3-642-33093-3_22.

[Kle09] A Kleidon. "Nonequilibrium thermodynamics and maximum entropy production in the Earth system". In: *Naturwissenschaften* 96.6 (2009), pp. 1–25. DOI: 10.1007/s00114-009-0509-x.

[Kle10a] A Kleidon. "A basic introduction to the thermodynamics of the Earth system far from equilibrium and maximum entropy production". In: *Philos T Roy Soc B* 365.1545 (2010), pp. 1303–1315. DOI: 10.1098/rstb.2009.0310.

[Kle10b] A Kleidon. "Life, hierarchy, and the thermodynamic machinery of planet Earth". In: *Phys Life Rev* 7.4 (2010), pp. 424–460. DOI: 10.1016/j.plrev.2010.10.002.

[Kle16] A Kleidon. *Thermodynamic Foundations of the Earth System*. Cambridge University Press, 2016. DOI: 10.1017/CBO9781139342742.

[KMC10] A Kleidon, Y Malhi, and P M Cox. "Maximum entropy production in environmental and ecological systems". In: *Philos T Roy Soc B* 365.1545 (2010), pp. 1297–1302. DOI: 10.1098/rstb.2010.0018.

[Kno03] A H Knoll. "The geological consequences of evolution". In: *Geobiology* 1.1 (2003), pp. 3–14. DOI: 10.1046/j.1472-4669.2003.00002.x.

[Knü+13] C Knüpfer, C Beckstein, P Dittrich, and N Le Novère. "Structure, function, and behaviour of computational models in systems biology". In: *BMC Syst Biol* 7.1 (2013), p. 1. DOI: 10.1186/1752-0509-7-43.

[KP98] D Kondepudi and I Prigogine. *Modern thermodynamics / from heat engines to dissipative structures*. Chichester ; Weinheim [u.a.]: Wiley, 1998. DOI: 10.1002/9781118698723.

[Kre+08] A Kreimer, E Borenstein, U Gophna, and E Ruppin. "The evolution of modularity in bacterial metabolic networks". In: *P Natl Acad Sci* 105.19 (2008), pp. 6976–6981. DOI: 10.1073/pnas.0712149105.

[Kre+12] P Kreyssig, G Escuela, B Reynaert, T Veloz, B Ibrahim, and P Dittrich. "Cycles and the qualitative evolution of chemical systems." In: *PLOS ONE* 7.10 (2012), e45772. DOI: 10.1371/journal.pone.0045772.

[KTBC16] J Krissansen-Totton, D S Bergsman, and D C Catling. "On detecting biospheres from chemical thermodynamic disequilibrium in planetary atmospheres". In: *Astrobiology* 16.1 (2016), pp. 39–67. DOI: 10.1089/ast.2015.1327.

[Ksc10] M Kschischo. "A gentle introduction to the thermodynamics of biochemical stoichiometric networks in steady state". In: *Eur Phys J Special Topics* 187 (2010), pp. 225–274. DOI: 10.1140/epjst/e2010-01290-3.

[KPH06] A Kümmel, S Panke, and M Heinemann. "Putative regulatory sites unraveled by network-embedded thermodynamic analysis of metabolome data". In: *Mol Syst Biol* 2.1 (2006). DOI: 10.1038/msb4100074.

[Lam+09] H Lammer, J H Bredehöft, A Coustenis, M L Khodachenko, L Kaltenegger, O Grasset, D Prieur, F Raulin, P Ehrenfreund, M Yamauchi, et al. "What makes a planet habitable?" In: *Astron Astrophys Rev* 17.2 (2009), pp. 181–249. DOI: 10.1007/s00159-009-0019-z.

[Lan90] C G Langton. "Computation at the edge of chaos: phase transitions and emergent computation". In: *Physica D* 42.1 (1990), pp. 12–37. DOI: 10.1016/0167-2789(90)90064-V.

[Lan97] C G Langton. *Artificial life: An overview*. MIT Press, 1997. DOI: 10.1002/(SICI)1099-1743(199701/02)14:1<78::AID-SRES139>3.0.CO;2-0.

[Lay+12] A Layton, J Reap, B Bras, and M Weissburg. "Correlation between thermodynamic efficiency and ecological cyclicity for thermodynamic power cycles". In: *PLOS ONE* 7.12 (2012), e51841. DOI: 10.1371/journal.pone.0051841.

[Leh04] R Lehmann. "An algorithm for the determination of all significant pathways in chemical reaction systems". In: *J Atmos Chem* 47.1 (2004), pp. 45–78. DOI: 10.1023/b:joch.0000012284.28801.b1.

[Len98] T M Lenton. "Gaia and natural selection". In: *Nature* 394.6692 (1998), pp. 439–447. DOI: 10.1038/28792.

[LL00] T M Lenton and J E Lovelock. "Daisyworld is Darwinian: constraints on adaptation are important for planetary self-regulation". In: *J Theor Biol* 206.1 (2000), pp. 109–114. DOI: 10.1006/jtbi.2000.2105.

[LPW16] T M Lenton, P-P Pichler, and H Weisz. "Revolutions in energy input and material cycling in Earth history and human history". In: *Earth Syst Dynam* 7.2 (2016), pp. 353–370. DOI: 10.5194/esd-7-353-2016.

[Lev92] S A Levin. "The problem of pattern and scale in ecology". In: *Ecology* 73.6 (1992), pp. 1943–1967. DOI: 10.2307/1941447.

[Lev98] S A Levin. "Ecosystems and the biosphere as complex adaptive systems". In: *Ecosystems* 1.5 (1998), pp. 431–436. DOI: 10.1007/s100219900037.

[Li+10] C Li, M Donizelli, N Rodriguez, H Dharuri, L Endler, V Chelliah, L Li, E He, A Henry, M I Stefan, et al. "BioModels Database: An enhanced, curated and annotated resource for published quantitative kinetic models". In: *BMC Syst Biol* 4.1 (2010), p. 92. DOI: 10.1186/1752-0509-4-92.

[Li+11] X Li, F Wu, F Qi, and D A Beard. "A database of thermodynamic properties of the reactions of glycolysis, the tricarboxylic acid cycle, and the pentose phosphate pathway". In: *Database* 2011 (2011), bar005. DOI: 10.1093/database/bar005.

[LE08] C H Lineweaver and C A Egan. "Life, gravity and the second law of thermodynamics". In: *Phys Life Rev* 5.4 (2008), pp. 225–242. DOI: 10.1016/j.plrev.2008.08.002.

[Lov65] J E Lovelock. "A physical basis for life detection experiments". In: *Nature* 207.997 (1965), pp. 568–570. DOI: 10.1038/207568a0.

[LL72] J E Lovelock and J P Lodge. "Oxygen in the contemporary atmosphere". In: *Atmos Environ (1967)* 6.8 (1972), pp. 575–578. DOI: 10.1016/0004-6981(72)90075-3.

[LM74] J E Lovelock and L Margulis. "Atmospheric homeostasis by and for the biosphere: the Gaia hypothesis". In: *Tellus* 26.1-2 (1974), pp. 2–10. DOI: 10.3402/tellusa.v26i1-2.9731.

[LO08] R F Ludlow and S Otto. "Systems chemistry". In: *Chem Soc Rev* 37.1 (2008), pp. 101–108. DOI: 10.1039/B611921M.

[LRP14] T W Lyons, C T Reinhard, and N J Planavsky. "The rise of oxygen in Earth's early ocean and atmosphere". In: *Nature* 506.7488 (2014), pp. 307–315. DOI: 10.1038/nature13068.

[MOI03] D Madina, N Ono, and T Ikegami. "Cellular evolution in a 3D lattice artificial chemistry". In: *Adv Artif Life, ECAL*. Springer, 2003, pp. 59–68. DOI: 10.1007/978-3-540-39432-7_7.

[MYS05] H Mahara, T Yamaguchi, and M Shimomura. "Entropy production in a two-dimensional reversible Gray-Scott system". In: *Chaos* 15.4 (2005), p. 047508. DOI: 10.1063/1.2140303.

[Mah+04] H Mahara, N J Suematsu, T Yamaguchi, K Ohgane, Y Nishiura, and M Shimomura. "Three-variable reversible Gray-Scott model." In: *J Chem Phys* 121.18 (2004), pp. 8968–72. DOI: 10.1063/1.1803531.

[MM93] B Å Månsson and J M McGlade. "Ecology, thermodynamics and HT Odum's conjectures". In: *Oecologia* 93.4 (1993), pp. 582–596. DOI: 10.1007/BF00328969.

[ML74] L Margulis and J E Lovelock. "Biological modulation of the Earth's atmosphere". In: *Icarus* 21.4 (1974), pp. 471–489. DOI: 10.1016/0019-1035(74)90150-X.

[Mar99] N M Marinov. "A detailed chemical kinetic model for high temperature ethanol oxidation". In: *Int J Chem Kinet* 31.3 (1999), pp. 183–220. DOI: 10.1002/(SICI)1097-4601(1999)31:3<183::AID-KIN3>3.0.CO;2-X.

[MQN14] V S Martínez, L-E Quek, and L K Nielsen. "Network Thermodynamic Curation of Human and Yeast Genome-Scale Metabolic Models". In: *Biophys J* 107.2 (2014), pp. 493–503. DOI: 10.1016/j.bpj.2014.05.029.

[MS06] L M Martyushev and V D Seleznev. "Maximum entropy production principle in physics, chemistry and biology". In: *Phys Rep* 426.1 (2006), pp. 1–45. DOI: 10.1016/j.physrep.2005.12.001.

[MV87] H R Maturana and F J Varela. *The tree of knowledge: The biological roots of human understanding*. New Science Library/Shambhala Publications, 1987.

[MGR93] B J McBride, S Gordon, and M A Reno. "Coefficients for calculating thermodynamic and transport properties of individual species". In: *NASA Technical Memorandum 4513* (1993).

[MB07] F J R Meysman and S Bruers. "A thermodynamic perspective on food webs: quantifying entropy production within detrital-based ecosystems". In: *J Theor Biol* 249.1 (2007), pp. 124–139. DOI: 10.1016/j.jtbi.2007.07.015.

[Mic05] K Michaelian. "Thermodynamic stability of ecosystems". In: *J Theor Biol* 237.3 (2005), pp. 323–335. DOI: 10.1016/j.jtbi.2005.04.019.

[Mic12] K Michaelian. "The biosphere: a thermodynamic imperative". In: *The Biosphere*. Ed. by N Ishwaran. INTECH, Paris, 2012, pp. 51–60. DOI: 10.5772/34620.

[Mor66] H J Morowitz. "Physical background of cycles in biological systems". In: *J Theor Biol* 13 (1966), pp. 60–62. DOI: 10.1016/0022-5193(66)90007-5.

[MS07] H Morowitz and E Smith. "Energy flow and the organization of life". In: *Complexity* 13.1 (2007), pp. 51–59. DOI: 10.1002/cplx.20191.

[New05] M E J Newman. "Power laws, Pareto distributions and Zipf's law". In: *Contemp Phys* 46.5 (2005), pp. 323–351. DOI: 10.1080/00107510500052444.

[NCC81] G R North, R F Cahalan, and J A Coakley. "Energy balance climate models". In: *Rev Geophys* 19.1 (1981), pp. 91–121. DOI: 10.1029/RG019i001p00091.

[Odu68] E P Odum. "Energy flow in ecosystems: a historical review". In: *Integr Comp Biol* 8.1 (1968), pp. 11–18. DOI: 10.1093/icb/8.1.11.

[OH72] W E Odum and E J Heald. "Trophic analyses of an estuarine mangrove community". In: *Bull Mar Sci* 22.3 (1972), pp. 671–738.

[OTP10] J D Orth, I Thiele, and B Ø Palsson. "What is flux balance analysis?" In: *Nat Biotechnol* 28.3 (2010), pp. 245–248. DOI: 10.1038/nbt.1614.

[OPK71] G Oster, A Perelson, and A Katchalsky. "Network thermodynamics". In: *Nature* 234 (1971). DOI: 10.1038/234393a0.

[Oza+03] H Ozawa, A Ohmura, R D Lorenz, and T Pujol. "The second law of thermodynamics and the global climate system: a review of the maximum entropy production principle". In: *Rev Geophys* 41.4 (2003). DOI: 10.1029/2002RG000113.

[Pac01] N R Pace. "The universal nature of biochemistry". In: *P Natl Acad Sci* 98.3 (2001), pp. 805–808. DOI: 10.1017/cbo9780511730191.015.

[PS07] R K Pan and S Sinha. "Modular networks emerge from multiconstraint optimization". In: *Phys Rev E* 76.4 (2007), pp. 1–4. DOI: 10.1103/PhysRevE.76.045103.

[PS09] R K Pan and S Sinha. "Modular networks with hierarchical organization: The dynamical implications of complex structure". In: *Pramana* 71.2 (2009), pp. 331–340. DOI: 10.1007/s12043-008-0166-1.

[Pap+03] J A Papin, N D Price, S J Wiback, D A Fell, and B O Palsson. "Metabolic pathways in the post-genome era". In: *Trends Biochem Sci* 28.5 (2003), pp. 250–258. DOI: 10.1016/S0968-0004(03)00064-1.

[Pap+04] J A Papin, J Stelling, N D Price, S Klamt, S Schuster, and B O Palsson. "Comparison of network-based pathway analysis methods". In: *Trends Biotechnol* 22.8 (2004), pp. 400–405. DOI: 10.1016/j.tibtech.2004.06.010.

[PPS13] R Pascal, A Pross, and J D Sutherland. "Towards an evolutionary theory of the origin of life based on kinetics and thermodynamics". In: *Open Biol* 3.11 (2013), p. 130156. DOI: 10.1098/rsob.130156.

[PBK01] A A Pavlov, L L Brown, and J F Kasting. "UV shielding of NH3 and O2 by organic hazes in the Archean atmosphere". In: *J Geophys Res-Planet* 106.E10 (2001), pp. 23267–23287. DOI: 10.1029/2000je001448.

[PK02] A A Pavlov and J F Kasting. "Mass-independent fractionation of sulfur isotopes in Archean sediments: strong evidence for an anoxic Archean atmosphere". In: *Astrobiology* 2.1 (2002), pp. 27–41. DOI: 10.1089/153110702753621321.

[Pea93] J E Pearson. "Complex patterns in a simple system". In: *Science* 261.5118 (1993), pp. 189–192. DOI: 10.1126/science.261.5118.189.

[Per14] J A Pereira. "Transient and sustained elementary flux mode networks on a catalytic string-based chemical evolution model". In: *Biosystems* (2014). DOI: j.biosystems.2014.06.011.

[Per75] A S Perelson. "Network thermodynamics. An overview." In: *Biophys J* 15 (1975), pp. 667–685. DOI: 10.1016/S0006-3495(75)85847-4.

[Per05] J Peretó. "Controversies on the origin of life". In: *Int Microbiol* 8 (2005), pp. 23–31.

[PSB05] T Pfeiffer, O S Soyer, and S Bonhoeffer. "The evolution of connectivity in metabolic networks." In: *PLoS Biol* 3.7 (2005), e228. DOI: 10.1371/journal.pbio.0030228.

[Pie+10] G Piedrafita, F Montero, F Moran, M L Cardenas, and A Cornish Bowden. "A simple self-maintaining metabolic system: robustness, autocatalysis, bistability". In: *PLoS Comput Biol* 6.8 (2010), e1000872. DOI: 10.1371/journal.pcbi.1000872.

[PE14] M Polettini and M Esposito. "Irreversible thermodynamics of open chemical networks. I. Emergent cycles and broken conservation laws". In: *J Chem Phys* 141.2 (2014), p. 024117. DOI: 10.1063/1.4886396.

[PL06] P Pons and M Latapy. "Computing communities in large networks using random walks." In: *J Graph Algorithms Appl* 10.2 (2006), pp. 191–218. DOI: 10.7155/jgaa.00124.

[Pra03] D K Pratihar. "Evolutionary robotics—A review". In: *Sadhana* 28.6 (2003), pp. 999–1009.

[Pre+07] W H Press, S A Teukolsky, W T Vetterling, and B P Flannery. *Numerical Recipes: the Art of Scientific Computing*. 3rd. Cambridge, UK: Cambridge University Press, 2007.

[QB05] H Qian and D A Beard. "Thermodynamics of stoichiometric biochemical networks in living systems far from equilibrium". In: *Biophys Chem* 114.2 (2005), pp. 213–220. DOI: 10.1016/j.bpc.2004.12.001.

[Ras+04] S Rasmussen, L Chen, D Deamer, D C Krakauer, N H Packard, P F Stadler, and M A Bedau. "Transitions from nonliving to living matter". In: *Science* 303.5660 (2004), pp. 963–965. DOI: 10.1126/science.1093669.

[Rau+11] H Rauer, S Gebauer, P Paris, J Cabrera, M Godolt, J L Grenfell, A Belu, F Selsis, P Hedelt, and F Schreier. "Potential biosignatures in super-Earth atmospheres-I. Spectral appearance of super-Earths around M dwarfs". In: *Astron Astrophys* 529 (2011), A8. DOI: 10.1051/0004-6361/201014368.

[Rau+13] H Rauer, J Cabrera, S Gebauer, and J L Grenfell. "Detection of habitable planets and the search for life". In: *Habitability of Other Planets and Satellites*. Springer, 2013, pp. 287–310. DOI: 10.1007/978-94-007-6546-7_16.

[Rei86] W A Reiners. "Complementary models for ecosystems". In: *Am Nat* (1986), pp. 59–73. DOI: 10.1086/284467.

[Rön07] M Rönkkö. "An artificial ecosystem: Emergent dynamics and lifelike properties". In: *Artif Life* 13.2 (2007), pp. 159–187. DOI: 10.1162/artl.2007.13.2.159.

[RMBE14] K Ruiz-Mirazo, C Briones, and A de la Escosura. "Prebiotic systems chemistry: new perspectives for the origins of life". In: *Chem Rev* 114.1 (2014), pp. 285–366. DOI: 10.1021/cr2004844.

[RMPM04] K Ruiz-Mirazo, J Peretó, and A Moreno. "A universal definition of life: autonomy and open-ended evolution". In: *Origins Life Evol B* 34.3 (2004), pp. 323–346. DOI: 10.1023/B:ORIG.0000016440.53346.dc.

[Sag+93] C Sagan, W R Thompson, R Carlson, D Gurnett, and C Hord. "A search for life on Earth from the Galileo spacecraft". In: *Nature* 365.6448 (1993), pp. 715–721. DOI: 10.1038/365715a0.

[Sau94] P T Saunders. "Evolution without natural selection: further implications of the Daisyworld parable". In: *J Theor Biol* 166.4 (1994), pp. 365–373. DOI: 10.1006/jtbi.1994.1033.

[Sau+03] S M Saunders, M E Jenkin, R G Derwent, and M J Pilling. "Protocol for the development of the Master Chemical Mechanism, MCM v3 (Part A): tropospheric degradation of non-aromatic volatile organic compounds". In: *Atmos Chem Phys* 3.1 (2003), pp. 161–180. DOI: 10.5194/acp-3-161-2003.

[Sch+15] C Scharf, N Virgo, H J II Cleaves, M Aono, N Aubert-Kato, A Aydinoglu, A Barahona, L M Barge, S A Benner, M Biehl, et al. "A Strategy for Origins of Life Research". In: *Astrobiology* 15.12 (2015), pp. 1031–1042. DOI: 10.1089/ast.2015.1113.

[SLP00] C H Schilling, David L, and B Ø Palsson. "Theory for the systemic definition of metabolic pathways and their use in interpreting metabolic function from a pathway-oriented perspective". In: *J Theor Biol* 203.3 (2000), pp. 229–248. DOI: 10.1006/jtbi.2000.1073.

[Sch06] J W Schopf. "Fossil evidence of Archaean life". In: *Philos T Roy Soc B* 361.1470 (2006), pp. 869–885. DOI: 10.1098/rstb.2006.1834.

[Sch43] E Schrödinger. *What Is Life? the Physical Aspect of the Living Cell and Mind*. Dublin, 1943.

[SS93] R Schuster and S Schuster. "Refined algorithm and computer program for calculating all non–negative fluxes admissible in steady states of biochemical reaction systems with or without some flux rates fixed". In: *Comput Appl Biosci* 9.1 (1993), pp. 79–85. DOI: 10.1093/bioinformatics/9.1.79.

[SDF99] S Schuster, T Dandekar, and D A Fell. "Detection of elementary flux modes in biochemical networks: a promising tool for pathway analysis and metabolic engineering". In: *Trends Biotechnol* 17.2 (1999), pp. 53–60. DOI: 10.1016/S0167-7799(98)01290-6.

[SH94] S Schuster and C Hilgetag. "On elementary flux modes in biochemical reaction systems at steady state". In: *J Biol Syst* 2.02 (1994), pp. 165–182. DOI: 10.1142/S0218339094000131.

[Sea13] S Seager. "Exoplanet habitability". In: *Science* 340.6132 (2013), pp. 577–581. DOI: 10.1126/science.1232226.

[Seg+07] A Segura, V S Meadows, J Kasting, M Cohen, and D Crisp. "Abiotic production of O2 and O3 on high CO2 terrestrial atmospheres". In: *Astrobiology* 7.3 (2007), pp. 494–495. DOI: 10.1051/0004-6361:20066663.

[SRK11] E Simoncini, M J Russell, and A Kleidon. "Modeling free energy availability from Hadean hydrothermal systems to the first metabolism". In: *Orig Life Evol Biosph* 41.6 (2011), pp. 529–532. DOI: 10.1007/s11084-011-9251-4.

[SVK13] E Simoncini, N Virgo, and A Kleidon. "Quantifying drivers of chemical disequilibrium: theory and application to methane in the Earth's atmosphere". In: *Earth Syst Dynam* 4.2 (2013), pp. 317–331. DOI: 10.5194/esd-4-317-2013.

[Sim94] K Sims. "Evolving 3D morphology and behavior by competition". In: *Artif Life* 1.4 (1994), pp. 353–372. DOI: 10.1162/artl.1994.1.4.353.

[SM04a] E Smith and H J Morowitz. "Universality in intermediary metabolism". In: *P Natl Acad Sci USA* 101.36 (2004), pp. 13168–13173. DOI: 10.1073/pnas.0404922101.

[SM16] E Smith and H J Morowitz. *The Origin and Nature of Life on Earth: The Emergence of the Fourth Geosphere*. Cambridge, UK: Cambridge University Press, 2016. DOI: 10.1017/cbo9781316348772.

[SH10] K C Soh and V Hatzimanikatis. "Network thermodynamics in the post-genomic era." In: *Curr Opin Microbiol* 13.3 (2010), pp. 350–7. DOI: 10.1016/j.mib.2010.03.001.

[SMH12] K C Soh, L Miskovic, and V Hatzimanikatis. "From network models to network responses: integration of thermodynamic and kinetic properties of yeast genome-scale metabolic networks." In: *FEMS Yeast Res* 12.2 (Mar. 2012), pp. 129–43. DOI: 10.1111/j.1567-1364.2011.00771.x.

[SM04b] R V Solé and A Munteanu. "The large-scale organization of chemical reaction networks in astrophysics". In: *Europhys Lett* 3.2 (2004), pp. 1–7. DOI: 10.1209/epl/i2004-10241-3.

[Sol16] R Solé. "Synthetic transitions: towards a new synthesis". In: *Philos T R Soc B* 371.1701 (2016), p. 20150438. DOI: 10.1098/rstb.2015.0438.

[Spe+13] E A Sperling, C A Frieder, A V Raman, P R Girguis, L A Levin, and A H Knoll. "Oxygen, ecology, and the Cambrian radiation of animals". In: *P Natl Acad Sci* 110.33 (2013), pp. 13446–13451. DOI: 10.1073/pnas.1312778110.

[SU10] F Srienc and P Unrean. "A Statistical Thermodynamical Interpretation of Metabolism." In: *Entropy* 12.8 (2010), pp. 1921–1935. DOI: 10.3390/e12081921.

[Sta+13] N J Stanford, T Lubitz, K Smallbone, E Klipp, P Mendes, and W Liebermeister. "Systematic construction of kinetic models from genome-scale metabolic networks". In: *PLOS ONE* 8.11 (2013), e79195. DOI: 10.1371/journal.pone.0079195.

[Sto+12a] J W Stock, J L Grenfell, R Lehmann, A B C Patzer, and H Rauer. "Chemical pathway analysis of the lower Martian atmosphere: The CO2 stability problem". In: *Planet Space Sci* 68.1 (2012), pp. 18–24. DOI: 10.1016/j.pss.2011.03.002.

[Sto+12b] J W Stock, C S Boxe, R Lehmann, J L Grenfell, A B C Patzer, H Rauer, and Y L Yung. "Chemical pathway analysis of the martian atmosphere: CO2-formation pathways". In: *Icarus* 219.1 (2012), pp. 13–24. DOI: 10.1016/j.icarus.2012.02.010.

[SS02] A Strahler and A Strahler. *Physical geography*. Berlin, Heidelberg: John Wiley & Sons, 2002.

[Svi99] Y M Svirezhev. "Virtual Biospheres: Complexity versus Simplicity". In: *Ann N Y Acad Sci* 879.1 (1999), pp. 368–382. DOI: 10.1111/j.1749-6632.1999.tb10440.x.

[SB97] Y M Svirezhev and W von Bloh. "Climate, vegetation, and global carbon cycle: the simplest zero-dimensional model". In: *Ecol Model* 101.1 (1997), pp. 79–95. DOI: 10.1016/S0304-3800(97)01973-X.

[SB98] Y M Svirezhev and W von Bloh. "A zero-dimensional climate-vegetation model containing global carbon and hydrological cycle". In: *Ecol Model* 106.2 (1998), pp. 119–127. DOI: 10.1016/S0304-3800(97)00187-7.

[TS08] M Terzer and J Stelling. "Large-scale computation of elementary flux modes with bit pattern trees". In: *Bioinformatics* 24.19 (2008), pp. 2229–2235. DOI: 10.1093/bioinformatics/btn401.

[TK01] R Thomas and M Kaufman. "Multistationarity, the basis of cell differentiation and memory. I. Structural conditions of multistationarity and other nontrivial behavior". In: *Chaos* 11.1 (2001), pp. 170–179. DOI: 10.1063/1.1350439.

[Tre+07] K E Trenberth, L Smith, T Qian, A Dai, and J Fasullo. "Estimates of the global water budget and its annual cycle using observational and model data". In: *J Hydrometeorol* 8.4 (2007), pp. 758–769. DOI: 10.1175/JHM600.1.

[TWS09] C T Trinh, A Wlaschin, and F Srienc. "Elementary mode analysis: a useful metabolic pathway analysis tool for characterizing cellular metabolism". In: *Appl Microbiol Biotechnol* 81.5 (2009), pp. 813–826. DOI: 10.1007/s00253-008-1770-1.

[US11] P Unrean and F Srienc. "Metabolic networks evolve towards states of maximum entropy production". In: *Metabolic Eng* 13.6 (2011), pp. 666 –673. DOI: http://dx.doi.org/10.1016/j.ymben.2011.08.003.

[Val10] J J Vallino. "Ecosystem biogeochemistry considered as a distributed metabolic network ordered by maximum entropy production." In: *Philos T R Soc B* 365.1545 (2010), pp. 1417–27. DOI: 10.1098/rstb.2009.0272.

[VMU74] F J Varela, H R Maturana, and R Uribe. "Autopoiesis: the organization of living systems, its characterization and a model". In: *Biosystems* 5.4 (1974), pp. 187–196. DOI: 10.1016/0303-2647(74)90031-8.

[VSGCU11] R M Velasco, L Scherer García-Colín, and F J Uribe. "Entropy production: Its role in non-equilibrium thermodynamics". In: *Entropy* 13.1 (2011), pp. 82–116. DOI: 10.3390/e13010082.

[Ven+12] O Venot, E Hébrard, M Agùndez, M Dobrijevic, F Selsis, F Hersant, N Iro, and R Bounaceur. "A chemical model for the atmosphere of hot Jupiters". In: *Astron Astrophys* 546 (2012), A43. DOI: 10.1051/0004-6361/201219310.

[VI13] N Virgo and T Ikegami. "Autocatalysis Before Enzymes: The Emergence of Prebiotic Chain Reactions". In: *Adv Artif Life, ECAL*. Vol. 12. 2013, pp. 240–247. DOI: 10.7551/978-0-262-31709-2-ch036.

[VNB66] J Von Neumann and A W Burks. "Theory of self-reproducing automata". In: *IEEE Trans Neural Netw* 5.1 (1966), pp. 3–14.

[WF01] A Wagner and D A Fell. "The small world inside large metabolic networks". In: *Proc R Soc Lond B Bio* 268.1478 (2001), pp. 1803–1810. DOI: 10.1098/rspb.2001.1711.

[WD13] S I Walker and P C W Davies. "The algorithmic origins of life". In: *J R Soc Interface* 10.79 (2013), p. 20120869. DOI: 10.1098/rsif.2012.0869.

[Wan+14] L-P Wang, A Titov, R McGibbon, F Liu, V S Pande, and T J Martínez. "Discovering chemistry with an ab initio nanoreactor". In: *Nat Chem* 6.12 (2014), pp. 1044–1048. DOI: 10.1038/nchem.2099.

[Wan+16] L-P Wang, R T McGibbon, V S Pande, and T J Martínez. "Automated Discovery and Refinement of Reactive Molecular Dynamics Pathways". In: *J Chem Theory Comput* 12.2 (2016), pp. 638–649. DOI: 10.1021/acs.jctc.5b00830.

[WLP10] Y P Wang, R M Law, and B Pak. "A global model of carbon, nitrogen and phosphorus cycles for the terrestrial biosphere". In: *Biogeosciences* 7.7 (2010), pp. 2261–2282. DOI: 10.5194/bg-7-2261-2010.

[WL83] A J Watson and J E Lovelock. "Biological homeostasis of the global environment: the parable of Daisyworld". In: *Tellus B* 35.4 (1983), pp. 284–289. DOI: tellusb.v35i4.14616.

[WS98] D J Watts and S H Strogatz. "Collective dynamics of 'small-world' networks". In: *Nature* 393 (1998), pp. 440–442. DOI: 10.1038/30918.

[Wei+16] M C Weiss, F L Sousa, N Mrnjavac, S Neukirchen, M Roettger, S Nelson-Sathi, and W F Martin. "The physiology and habitat of the last universal common ancestor". In: *Nat Microbiol* 1 (2016), p. 16116. DOI: 10.1038/nmicrobiol.2016.116.

[WG02] G M Whitesides and B Grzybowski. "Self-assembly at all scales". In: *Science* 295.5564 (2002), pp. 2418–2421. DOI: 10.1126/science.1070821.

[WI99] G M Whitesides and R F Ismagilov. "Complexity in chemistry". In: *Science* 284.5411 (1999), pp. 89–92. DOI: 10.1126/science.284.5411.89.

[Wil09] T Wilhelm. "The smallest chemical reaction system with bistability". In: *BMC Syst Biol* 3.1 (2009), pp. 1–9. DOI: 10.1186/1752-0509-3-90.

[WSH97] T Wilhelm, S Schuster, and R Heinrich. "Kinetic and thermodynamic analyses of the reversible version of the smallest chemical reaction system with Hopf bifurcation". In: *Nonlinear World* 4 (1997), pp. 295–322.

[WL07a] H T P Williams and T M Lenton. "Artificial selection of simulated microbial ecosystems". In: *P Natl Acad Sci* 104.21 (2007), pp. 8918–8923. DOI: 10.1073/pnas.0610038104.

[WL07b] H T P Williams and T M Lenton. "The Flask model: emergence of nutrient-recycling microbial ecosystems and their disruption by environment-altering 'rebel' organisms". In: *Oikos* 116.7 (2007), pp. 1087–1105. DOI: 10.1111/j.0030-1299.2007.15721.x.

[Woe+04] M A Woelfle, Y Ouyang, K Phanvijhitsiri, and C H Johnson. "The adaptive value of circadian clocks: an experimental assessment in cyanobacteria". In: *Curr Biol* 14.16 (2004), pp. 1481–1486. DOI: 10.1016/j.cub.2004.08.023.

[Woo+08] A J Wood, G J Ackland, J G Dyke, H T P Williams, and T M Lenton. "Daisyworld: a review". In: *Rev Geophys* 46.1 (2008). DOI: 10.1029/2006RG000217.

[Wur82] P Wurfel. "The chemical potential of radiation". In: *J Phys C Solid State* 15.18 (1982), p. 3967. DOI: 10.1088/0022-3719/15/18/012.

[YBC11] L Yamamoto, W Banzhaf, and P Collet. "Evolving reaction-diffusion systems on GPU". In: *Lect Notes Artif Int* (2011), pp. 208–223. DOI: 10.1007/978-3-642-24769-9_16.

[YQB05] F Yang, H Qian, and D A Beard. "Ab initio prediction of thermodynamically feasible reaction directions from biochemical network stoichiometry". In: *Metabolic Eng* 7.4 (2005), pp. 251–259. DOI: 10.1016/j.ymben.2005.03.002.

[Yos10] R Yoshida. "Self-oscillating gels driven by the Belousov–Zhabotinsky reaction as novel smart materials". In: *Adv Mater* 22.31 (2010), pp. 3463–3483. DOI: 10.1002/adma.200904075.

[YN13] T J Young and K Neshatian. "A constructive artificial chemistry to explore open-ended evolution". In: *AI 2013: Advances in Artificial Intelligence*. Springer, 2013, pp. 228–233. DOI: 10.1007/978-3-319-03680-9_25.

[YD98] Y L Yung and W B DeMore. *Photochemistry of Planetary Atmospheres*. New York: Oxford University Press, 1998.

[ZN11] W Zhou and L Nakhleh. "Properties of metabolic graphs: biological organization or representation artifacts?" In: *BMC Bioinformatics* 12.1 (2011), p. 132. DOI: 10.1186/1471-2105-12-132.

CURRICULUM VITAE

# Jakob Fischer

*Artificial Life, Computing, Complex Systems*

## Education

since 05/2012 **PhD candidate at the Biosystemanalysis Group at FSU Jena.**
Schoolarship of the International Max Planck Research School for Global Biogeochemical Cycles

04/15– 04/16 **Visiting Research Student**, *Ikegamilab*, Tokyo University.
Schoolarship of the German Academic Exchange Service (DAAD)

2006–2012 **Physics (Diploma)**, *Technische Universität Dresden*.
Elective Subjects Theoretical Physics (Computational Physics, Quantum-Chaos)
Mathematics (Functional Analysis I + II, Ergodic Theory)
Thesis *Monte-Carlo-Simulations for Investigation of Olympic Gels*

2005 **Abitur (high school diploma)**, *Technisches Gymnasium*, Sigmaringen.
Mayor subject: Information Technology

## Talks, Posters

05/2015 Talk, Workshop on Artificial Life and Open-Ended Evolution at ELSI, (Tokyo)
09/2013 Talk, European Conference on Complex Systems (Barcelona)
11/2012 Poster, Workshop „Scales and Patterns in the Earth System“ (Dresden)

## Further Experiences

2016 **Machine Learning Summer School**, *Arequipa, Peru*.
2012–2015 **Various courses**, *by Graduate Academy Jena and IMPRS-gBGC (graduate school)*.
Topics include: Data Visualisation, Grant Proposal Writing, Scientific Writing, Statistics & Data Analysis, Earth Observation Techniques, Soil Biology
2013-2015 **PhD-Representative**, *IMPRS-gBGC (graduate school)*.
2013 **Teaching assistant**, *lecture on evolutionary algorithms*.
2005–2006 **Zivil-Service**, *Sozialstation St. Konrad*, Konstanz.

# DECLARATION

Declaration with respect to the doctorate procedure.

Hiermit erkläre ich ehrenwörtlich:

- Mir ist die Promotionsordnung der Fakultät für Mathematik und Informatik (vom 3. Juni 2009) bekannt.
- Die Hilfe eines Promotionsberaters bzw. einer Promotionsberaterin wurde nicht in Anspruch genommen. Dritte haben im Zusammenhang mit dem Inhalt dieser Arbeit weder unmittelbar noch mittelbar geldwerte Leistungen erhalten.
- Die Dissertation (oder Auszüge daraus) wurde noch nicht als Prüfungsarbeit für eine staatliche oder andere wissenschaftliche Prüfung eingereicht.
- Bis jetzt wurde noch keine Dissertation bei einer anderen Hochschule (oder vergleichbaren Einrichtungen auf anderen Planeten) eingereicht.

Bei der Auswahl und Auswertung des Materials sowie bei der Erstellung des Manusktripts wurde ich von folgenden Personen unterstützt: Prof. Dr. Peter Dittrich, Dr. Axel Kleidon, Prof. Takashi Ikegami, Eugenio Simoncini (PhD). Für Korrekturlesen (sprachliche und orthographische Aspekte) ausgewählter Kapitel war Stephanie Luther entgeltlich tätig.

*Jena, Juni 2017*

Jakob Fischer

# INDEX