

EMERGENT EVOLUTIONARY DYNAMICS OF SELF-REPRODUCING CELLULAR AUTOMATA

Christopher Salzberg

In partial fulfillment of requirements for the Master of Science degree

Section Computational Science
Universiteit van Amsterdam

Supervised by Dr. Hiroki Sayama
University of Electro-Communications

October 1, 2003

CONTENTS

1. <i>Introduction</i>	5
1.1 Motivation	5
1.2 The idea of “Artificial Life”	6
1.3 Focus of this thesis	6
1.4 Outline of contents	7
2. <i>Synthetic models of evolution</i>	8
2.1 Introduction	8
2.2 First steps	8
2.2.1 Theory of Self-Reproducing Automata and the Universal Constructor	8
2.2.2 Separation for evolution	10
2.3 Cellular automata as modeling device	10
2.4 Emergence, complexity, and multilevel evolution	11
2.5 CA models for self-reproduction	13
2.5.1 Background	13
2.5.2 The SR Loop	15
2.5.3 The SDSR Loop	15
2.6 Summary	18
3. <i>The evoloop</i>	19
3.1 Introduction	19
3.2 An emergent mutation mechanism	19
3.3 States and transition rules	20
3.4 Emergent evolutionary dynamics	22
3.5 Summary	23
4. <i>New tools of analysis</i>	24
4.1 Motivation	24
4.2 The evoloop as a biological system	24
4.3 Detection and identification	25
4.3.1 Birth detection mechanism	25
4.3.2 The “scan” layer and the evoloop “footprint”	29
4.3.3 Death-detection mechanism	30
4.4 Genealogy tracing	30
4.4.1 The “trace” identifier	31

4.4.2	Multiple-parent detection	33
4.5	Summary	35
5.	<i>Morphology and dynamics: A closer look</i>	36
5.1	Introduction	36
5.2	Passive and active mutation of the evoloop	36
5.3	A classification of loops	37
5.3.1	Stationary properties	38
5.3.2	Reproductive properties and graph-based genealogy	39
5.4	The evoloop state space	40
5.5	Behavioural study of different loops	42
5.5.1	Pattern formation and morphogenesis	42
5.5.2	Population dynamics and genealogy	44
5.6	Summary	51
6.	<i>A new kind of loop</i>	52
6.1	Introduction	52
6.2	Conservation properties of the evoloop	52
6.3	Experiments	58
6.3.1	Convergence to the minimal loop	58
6.3.2	Cyclic genealogy and emergent selection	62
6.3.3	Further observations	64
6.4	Summary	65
7.	<i>Conclusions and Future Directions</i>	66
7.1	Summary of results	66
7.2	Conclusions	67
7.3	Future directions	68
7.4	Acknowledgments	69

ABSTRACT

In this thesis, a new approach to the study of evolutionary processes is presented using as medium the mathematical formulation of cellular automata (CA). Previous work in this field limited the scope of analysis to focus on model construction; our emphasis is to expand this view. We investigate a system of self-reproducing loops (“evoloops”) in which mechanisms for variation and selection are considered *emergent*. Simple, deterministic state-transition rules result in a highly robust and flexible morphology, allowing self-replicators to modify the structure of offspring through phenotypic interaction. No explicit conditions are imposed on survival: selection criteria are an emergent property of the global system and cannot be deduced from local CA rules. As a result, populations undergo self-organized evolutionary processes leading to complex and unpredictable spatial and temporal dynamics on multiple scales.

The study of these complex dynamics forms the primary focus of this thesis. Earlier investigations have largely targeted aspects of design and structural mechanics of the self-replication process. The possibility of discovering *open-ended, complexity-increasing* evolution has been implicitly precluded by inadequate methods of analysis and a limited subset of initial conditions. The work presented here aims to emphasize the wider potentialities of this system by addressing a number of open questions.

We begin with the introduction of new mechanisms for event-driven detection, identification, and genealogy tracing, each of which operate with negligible computational overhead. Given this new framework, we show that the genealogical state-space of replicators scales exponentially with the size of replicator and that populations exhibit far greater behavioural diversity than earlier estimates would suggest. To capture the full connectivity of genealogical links spanned by these populations, we introduce the concept of a “genealogy graph” as a generalization on the familiar tree hierarchy. A classification of species based on fundamental evolutionary properties arises from this graph-based idea of genealogy and leads to the discovery of a new class of species. For a subset of these species Darwinistic selection is observed to balance replication speed with new, emergent criteria, extending evolutionary time scales by orders of magnitude. Exploration of this new genealogy space with simple, small-scale experiments, demonstrates a richness of dynamics exceeding our every expectation.

1. INTRODUCTION

The complexity of the simplest known type of cell is so great that it is impossible to accept that such an object could have been thrown together in some kind of freakish, vastly improbably, event. Such an occurrence would be indistinguishable from a miracle.

Michael Denton, *Evolution: A Theory in Crisis* [10]

1.1 Motivation

One of the most fundamental problems facing modern science is to understand the emergence of *open-ended, complexity-increasing* evolution, a phenomena central to all forms of biotic life. The seemingly boundless diversity of the earthly biosphere tempts us to believe that a general theory, one which describes the emergence of complex life in a general way, must exist. Yet mechanisms governing this process elude our grasp, beyond the reach of physical laws and biological theories.

At the core of any theory of evolution is the process of *self-replication*, the procedure by which an entity (or pair of entities) creates a new entity. The study of self-replication is traditionally associated with natural biological systems of carbon-based organisms such as cells or bacteria. Prior to the 1950s, research addressing this topic was limited in scope by the range and variety of organisms offered by natural ecosystems, hence the approach taken was largely taxonomic. By its very nature, such an approach fails to describe the fundamental principles underlying the evolution of life in an abstract way; all empirical observations are inherently linked to complex biophysical systems, the components of which (molecules, atoms) are themselves not completely understood. A formal description of the problem necessitates a complete abstraction from the biological system; this was only achievable given advances in the theory and implementation of computational machines starting in the mid-twentieth century.

The first steps towards a formal understanding of self-replication were taken by von Neumann and Ulam in the 1950s. Von Neumann, a brilliant mathematician, logician and critical thinker, had as his aim to “abstract form the natural self-reproduction problem its logical form” [5]. Inspired by a suggestion by Ulam, von Neumann introduced the idea of *cellular automata* (CA) to realize an abstract machine capable of what he called *universal construction*. These first steps laid the groundwork for a new field of science, one which takes a *synthetic* approach to the study of living systems. Nearly forty years later, Langton [25] — another critical thinker and researcher in this new field — coined the term “Artificial Life” (ALife for short) to describe the study of what he called “life as it could be”.

1.2 The idea of “Artificial Life”

Only when we are able to view life-as-we-know-it in the larger context of life-as-it-could-be will we really understand the nature of the beast. Artificial Life is a relatively new field employing a synthetic approach to the study of life-as-it-could-be. It views life as a property of the organization of matter, rather than a property of the matter which is so organized.

Christopher Langton, Proceedings [24].

The field Artificial Life can be traced back to the work of von Neumann and Ulam in the 1950s, but only truly came together in the late 1980s. In the intervening decades, milestones such as Conway’s “Game of Life” [13, 14] in the 1970s and Langton’s Self-Reproducing (SR) Loop [22] in the 1980s set the stage for a new inter-disciplinary field of research, an emerging discipline inspired by biology yet rooted in computer science. Ever-increasing processor speeds and the advent of the personal computer served as a catalyst and driving force for computational simulation and experimentation. Growing interest in this field in the early 1990s culminated in a number of evolutionary systems exhibiting a variety of life-like phenomena. Most notable and by far the most popular among these systems was Ray’s Tierra Virtual Life Simulator [30], a system of digital organisms (computer programs) living and reproducing inside the memory of a computer. Tierra was truly a breakthrough for ALife, demonstrating that biological phenomena such as competition, emergence of parasites, and punctuated equilibrium could be synthesized in a digital computer using a flexible, adaptable structure and minimal specification.

Even in these early stages of its development, the spirit of Artificial Life research differed from related fields such as Artificial Intelligence (AI) and Cybernetics. Rather than obtain desired behaviour using a “top-down” model (such as with e.g. expert systems), ALife systems implement a “bottom-up” approach in which collective behaviour *emerges* from local interactions [1], and in which general properties (diversity, self-organization, complexity) are prioritized over narrow goals (speed, accuracy, efficiency). Describing this new field, Ikegami [17] states: “To be able to describe a system not from without, but from within the system itself, characterizes research with artificial life.” It is this *synthetic* approach — central to research in Artificial Life — which forms the basis for the work presented in this thesis.

1.3 Focus of this thesis

The work presented in this thesis focuses on a particular model of self-replication and evolution constructed by Sayama [35, 37] and based on an earlier model by Langton [22]. This model is called the “evoloop”, a self-reproducing cellular automata with a small state-space and relatively simple rule set. The remarkable feature of the evoloop is that it naturally generates both variation and selection leading to evolution, hence exhibiting the first example of an *emergent evolutionary process* in a (deterministic) cellular automata space. Previous work on the evoloop by Sayama — as with most other research in Artificial Life — focused on model construction rather than detailed analysis. Though promising as an abstract model for evolution, the result of this focus led to a somewhat disappointing conclusion: on small-scale periodic domains, loops would consistently gravitate towards the smallest possible size, at which point evolution would essentially stop.

The work presented here aims to emphasize the wider potentialities of this system by answering the following questions:

1. Is it computationally intractable to identify self-replicators and their genealogical history *exactly* given a sufficiently well-defined detection scheme?
2. If we have such a scheme, how big is the genetic landscape of these replicators? And what is the topology of this genealogy-space?
3. Does selection optimize on rate of replication alone, or are there additional emergent properties which may override this criteria?

Note that, although these questions will be answered in this thesis for the particular case of the evoloop CA, they are broad in scope. As such, we aim to arrive at conclusions whose relevance extends beyond the simple system under investigation; the evoloop, in this context, acts as a prototypical case study through which a much larger set of self-replicating systems may be characterized and understood.

1.4 Outline of contents

Chapter 2 briefly reviews relevant evolutionary systems in the field of Artificial Life, focusing on models which use the concept of synthesis as a means to derive general conclusions about the evolutionary dynamics of self-replicators. We briefly discuss von Neumann's Theory of Self-Reproducing Automata[49] and the invention of Cellular Automata (CA), followed by an introduction to the two main directions of research which use CA as modeling device. Two relevant CA models are discussed: Conway's Game of Life[13, 14] and spatially embedded hypercyclic catalytic sets exhibiting spiral wave formation[4, 34, 42]. We then review models of self-replication in CA, starting from Langton's Self-Reproducing (SR) Loop[22] followed by the Structurally Dissolvable Self-Reproducing Loop (SDSR)[36].

In Chapter 3, we introduce the evoloop CA by Sayama[35, 37]. We relate the self-replication mechanism in this system to the concept of *separation* as discussed by e.g. von Neumann[49] and later by Ikegami[18]. A brief outline of states and transition rules for this CA is presented, followed by some brief observations regarding global evolutionary trends as observed previously by Sayama[35, 37].

Chapter 4 introduces new tools of analysis to better describe evolutionary dynamics of the evoloop CA. It is in this chapter that we provide a tentative answer to question 1 as posed above in the form of new, computationally inexpensive analysis routines. This is followed by an in-depth investigation using these new tools in Chapter 5, in which we address the question of genetic landscape and describe the topology of genealogy-space. In Chapter 6, we briefly discuss a new class of evoloop species in response to the third question stated above. In the case of these new species, selection is not exclusively dependent on the reproduction-rate (size of loop).

We conclude the thesis in Chapter 7 with a discussion of current results and an outline of future goals.

2. SYNTHETIC MODELS OF EVOLUTION

There exists a critical size below which the process of synthesis is degenerative, but above which the phenomenon of synthesis, if properly arranged, can become explosive, in other words, where the synthesis of automata can proceed in such a manner that each automaton will produce other automata which are more complex and have higher potentialities than itself.

John von Neumann, *Theory of Self-Reproducing Automata*[49]

2.1 Introduction

In this chapter we give a brief outline of some relevant models of evolution in the field of Artificial Life. We focus on models which use the concept of *synthesis* as a means to derive general conclusions about evolutionary dynamics in systems of self-replicators. We begin with the first work on this topic, von Neumann's Universal Constructor, the design of which coincides with the invention of cellular automata as a medium for self-replication. We then briefly introduce two directions of research which use cellular automata as modeling device. In Section 2.4, we introduce the topics of emergence, complexity and multi-level evolution. We discuss Conway's Game of Life and the formation of spiral waves in hypercyclic catalytic sets as examples of these concepts. We then focus our attention on CA models for self-reproduction and describe the most well-known CA in this class: the self-reproducing (SR) loop by Langton[22]. We conclude with a description of the Structurally Dissolvable Self-Reproducing Loop[36], a modified SR Loop and precursor to the evolooop[35, 37].

2.2 First steps

2.2.1 Theory of Self-Reproducing Automata and the Universal Constructor

Von Neumann's seminal work on self-reproducing cellular automata is widely acknowledged as having set the stage for future research in the field of Artificial Life[28]. As stated by Burks[5]:

Von Neumann was interested in the general question: What kind of logical organization is *sufficient* for an automaton to be able to reproduce itself? The question is not precise and admits to trivial versions as well as interesting ones. Von Neumann had the familiar natural phenomenon of self-reproduction in mind when he posed it, but he was not trying to simulate the self-reproduction of a natural system at the levels of genetics and biochemistry. He wished to abstract from the natural self-reproduction problem its logical form.

Von Neumann's work on self-reproduction resulted in a 29-state cellular automata with a 5-cell (von Neumann) neighbourhood capable of *universal construction*: the functionality to construct any other machine. These ideas are illustrated in Figure 2.1. Though this demand for universal construction was later found to be unnecessarily restrictive [22], von Neumann's overall attempt to differentiate the logical form of evolution and self-replication from the realm of biology was a bold and important step.

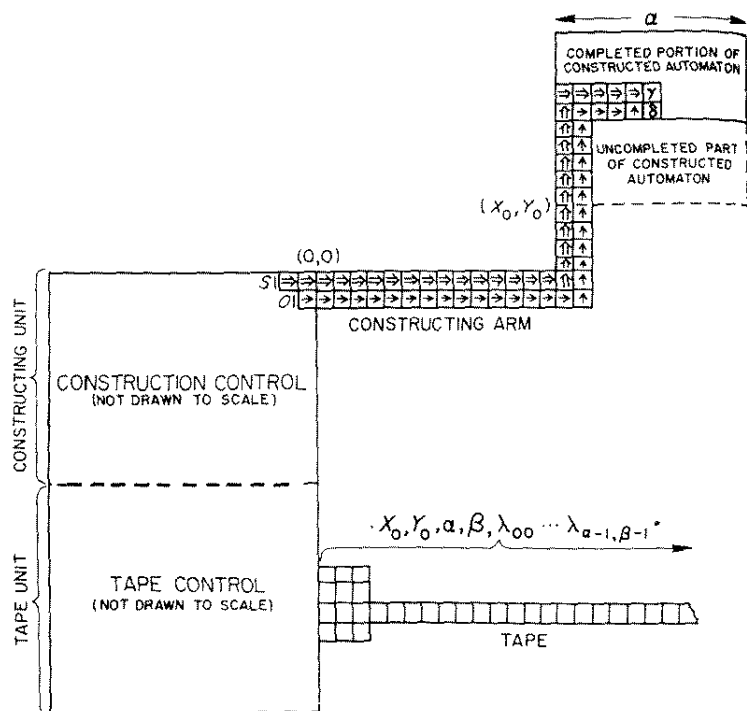


Fig. 2.1: von Neumann's Universal Constructor

The basic operation of the Universal Constructor is loosely based on Turing's idea of computation in which a machine (computer) reads a tape and executes instructions (writes to the tape) based on a prescribed code[46]. The critical difference in von Neumann's model is that his machine is capable of *construction*: its instruction tape describes how to make a *new* automaton of arbitrary type. That this constructor is embedded in a CA space means that challenges in construction linked to kinetic properties of physical space are circumvented. Hence CA transformation rules disregard conservation of material, energy and momentum, and may arbitrarily appear and disappear. In his more abstract work, however, von Neumann was also researching a theory of self-reproduction in a kinetic space[48, 49].

One of the truly remarkable achievements of von Neumann's Universal Constructor is that, as a model for self-replication, it predicted the separation between genetic information and translators/transcribers prior to the discovery of double helix structure of DNA[35]. Packard states: "At the time, the construction was surprising, since many had argued its impossibility, e.g., on the grounds that such an entity would need to contain a description of itself, and that description

would also need to contain a description, *ad infinitum*.” [28]. This separation is now widely understood as central to the self-reproduction paradigm, as emphasized for example by the later work of Ikegami[18] with his work on machines and tapes (see Section 5.2 for details). Von Neumann’s model for self-replication represents the first attempt to *reduce* the complexity inherent in physical biological processes, something very different from *mimicking* biology at the level of empirical observation. This is a distinction which is often overlooked, even by researchers in the field of Artificial Life, yet its importance cannot be underemphasized. Von Neumann’s ultimate goal was to find a minimal description of the logical interactions governing the emergence of complex forms of life, a search which was ended prematurely by his early death in 1957.

2.2.2 Separation for evolution

The idea of *separation for evolution* — first predicted by von Neumann and shortly thereafter discovered in the division between genetic information (RNA/DNA) and translators/transcribers — has remained a central paradigm in the study of self-replication. In both biological self-reproduction and von Neumann’s artificial reproduction, construction is *universal*: the RNA/DNA’s copy process synthesizes a complementary chain according to the information contained in the double helix, just as the UC constructs a new machine according to code in the instruction tape. On this topic, Ikegami[18] states: “It is remarkable that for this universality to exist in both both biology and cellular automaton, the separation of program (i.e., machine) and data (i.e., tape) is prerequisite.”

Although the requirement for universality has since been discarded, the idea of separation has remained central to the self-reproduction paradigm. Later work by e.g. Langton[22] on self-replicating loops (to be discussed in Section 2.5.2) implements simple self-replication and discards universality, demonstrating that nontrivial self-reproduction is achievable via self-description (gene sequence). Yet this type of self-replicator — as well as von Neumann’s original UC — suffer from lack of structural mutability: they cannot *evolve*. This is the contradiction of an evolving self-replicator: a structure which reads its description, and replicates an object which is *different*. Despite its importance as a prerequisite for complexity-increasing evolution, Von Neumann never addressed this contradiction.

The role played by separation is critical to both self-replication and to evolution. Ikegami[18] makes the analogy that “separation is to evolution as language is to creativity”. In order to create a copy which is in some way different, a replicator must have a framework (description tape) within which modifications can be made (either randomly or deterministically); this is analogous to making novel statements within the limitations of a given language. In evolutionary terms, this process of modified self-reproduction is what we call *mutation*¹. We present an example of this process in Chapter 3 when we introduce an evolving self-replicator, the evolloop[35, 37].

2.3 Cellular automata as modeling device

The original formulation for cellular automata (CA) by von Neumann and Ulam in the early fifties quickly developed into a powerful modeling device for a wide variety of applications. Broadly speaking, models using CA as medium can be divided into two groups. The first set of models

¹ The topic of mutation in terms of separation is discussed in more detail in Section 5.2.

focus on the emergence of unexpected nonlinear behaviour at a global scale from simple rules at the local scale; this will be the focus of this thesis, and is further discussed in the next section. A quite distinct perspective is emphasized by e.g. Toffoli[45], who applies cellular automata directly to the modeling of physical laws. In this case the CA formulation is used as a *computational device*, offering an inherent potential for parallel-processing[20]. The best example of such physical modeling comes in the form of lattice gas automata (LGA), an approximation of fluid flow on a multi-dimensional lattice of grid points.

One of the most difficult aspects of working with CA as a modeling device is that there is no simple rule for how a given physical system (e.g. in the form of a differential equation) maps to a CA, or even whether a suitable CA exists. This poses a considerable challenge for those researchers who wish to use CA to model an explicitly defined set of equations. For the purposes of this thesis, however, we emphasize the other side of the coin: that many CA models exhibit highly complex behaviour which appears to have *no continuous analog* in the form of a dynamical system. These CA fit in the fourth qualitative behavioural class (**c4**) as defined by Wolfram[50], a class characterized by complex, localized propagating structures. CA in this class are typically *irreducible*, hence there is no “short-cut” to predicting their behaviour; many are moreover capable of *universal computation*, the ability to implement any computable function[51]. Ilachinsky emphasizes: “cellular automata represent not just a formalism for describing a certain particular class of behaviors, but a much more general “template” for original and heretofore untapped ways of looking at a large class of unsolved or only poorly understood *fundamental* problems.”[20] This makes CA a very attractive medium for modeling the complexity and mystery of life, and goes a long way to explaining the prominent role CA play in the field of Artificial Life.

2.4 Emergence, complexity, and multilevel evolution

In describing the research methodology of Artificial Life, Packard et. al. [28] express the view that “the spirit of most artificial life work is to look for the complexity in the emergent global behavior of the system, rather than to program the complexity directly into the elements.” Some go as far as to claim emergence to be the “hallmark of life”, yet the term itself is notoriously ill-defined. What is this “emergent” behaviour, and what do we mean by “complexity”? Furthermore, how can we relate the concept of emergence to Darwinistic natural selection?

Crutchfield[8] attempts to answer the first of these questions by defining emergence as “a process that leads to the appearance of structure not directly described by the defining constraints and instantaneous forces that control a system”. Packard et. al.[28] describe it as “resultant aggregate behaviour of complex systems.” The idea that a system may exhibit global structure which is distinct from the local behaviour of its component parts strongly relates to the processing and storage of *information*. The question of how meso-scale structures — structure not explicitly defined in the dynamics of component parts — emerge from only simple, local interactions is very much unanswered, largely due to the lack of a coherent theory describing the physics of information[20].

Emergence is observed in a wide variety of phenomena, from spiral waves produced in oscillating chemical reactions, to power laws and $1/f$ noise in critical phenomena[3], to structures appearing in biological morphogenesis[47]. Given the localized nature of interactions in their basic formulation, cellular automata demonstrate perhaps the most direct (and elegant) representation of such systems. Ilachinsky[20] makes the observation that cellular automata “teach us that not all phenomena

appearing on a given level or scale of observation have explanations that can be derived solely from the dynamics operating on that same level.” This idea of *multi-level* dynamics is important, all the more so for systems which *replicate* and *evolve*. Indeed, Hogeweg emphasizes that “the ‘emergent properties’ of a (one-level) coevolutionary system may generate a new level of selection”[15]. CA represent in many ways the ideal medium to analyze such new levels of selection.

The concepts of emergence and complexity were well known to mathematician John Conway when he began devising a set of rules for the binary CA which would make him famous. The Game of Life[13, 14], a binary (2-state) cellular automata with Moore neighbourhood (9 neighbours), is decidedly the most popular and well-known CA, exhibiting complex dynamics from an elegantly simple rule set:

- A dead (quiescent) site will turn into a living (active) site if and only if it is surrounded by exactly three living sites.
- A living site will remain alive if it is surrounded by two or three other living sites. Otherwise the site will die.

From the above specification, a wide variety of structures are found to emerge in the CA space. The most well-known of these is the “glider”, a structure which — through a sequence of four transitions, moves diagonally in the CA space. This is shown in Fig. 2.2. Following its widespread rise in popularity in the 1970s, many other structures were found to emerge from the Game of Life rules, such as eaters, shuttles, puffer trains, and most notably a complex structure called a *glider gun*. This last structure is a carefully constructed assemblage of signal streams made of gliders which together generate a kind of periodic signal. The importance of the glider gun is that it plays the pivotal role in transforming the Game of Life into a *universal computer*, complete with memory and CPU. The fact that such a simple CA was found capable of universal computation remains as astounding today as it was when it was first discovered.

A quite distinct example of emergence, realized in cellular automata, was simulated first by Boerlijst et. al.[4] and later generalized by Stassinopoulos et. al.[42] using a CA based on a model for self-replicating molecules by Eigen and Schuster [12]. The basic idea of this model focuses on an *autocatalytic* set of chemicals “in which each member is the product of at least one reaction catalyzed by at least one other member”[42]. In such a set, groups of chemicals which have *catalytic closure* — in which each species lends catalytic support to the next in the cycle — are able to co-exist at the exclusion of species which are not part of the cycle. This group of species is called a Hypercycle[12, 26].

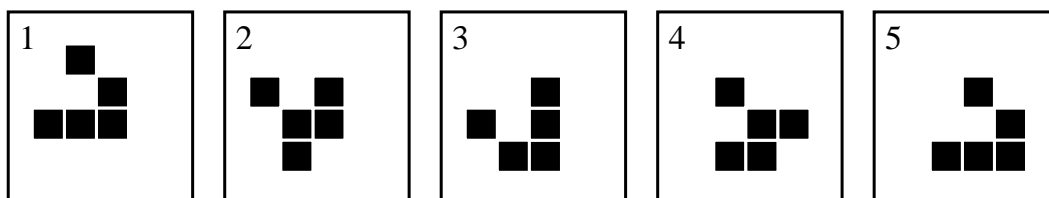


Fig. 2.2: Motion of a “glider” in Conway’s Game of Life

The CA model analyzed in [4, 34, 42] embeds the concept of a hypercycle into a spatially extended domain. As with the Game of Life rules, the rules for this model are simple and localized. What is remarkable about experimental results of Boerlijst and Hogeweg is that the emergent dynamics of the system change the selection process through spatial structuring[4]. The addition of parasites to the molecular system is used to demonstrate *positive selection* for an altruistic property due to self-structuring of molecules into spiral waves:

Because a spiral consists of more than one species there appears to be inter-specific “group selection” on the individual level, i.e. there is selection for “helping” (giving catalytic support to) another molecule species.[4]

Later results by Savill et. al. [34] refer to evolution being *enslaved* by self-reinforcing spatial patterns (spiral waves). That such complex dynamics can emerge from such simple models is an indication of the power of CA as a tool in scientific research. Self-structuring as a modifier for selection will in particular play an important role in the results we present in Chapters 5 and 6.

2.5 CA models for self-reproduction

The cornerstone of a “minimal living system” is self-replication[1], a process which has been synthesized in a wide variety of CA in recent decades[6, 7, 16, 21, 22, 27, 36, 37, 38, 40]. In this section we briefly review CA models, then focus on two which use a loop as structure for reproduction: Langton’s Self-Reproducing (SR) Loop[22] and a modified version called the Structurally Dissolvable Self-Reproducing (SDSR) Loop[36] by Sayama. These models form the basis for the evoloop CA as discussed in the following chapter.

2.5.1 Background

The field of Artificial Life has recently witnessed a moderate resurgence in theoretical studies on self-replicating and evolving structures[16, 21, 40, 41, 43]. Among them are the efforts of template-based self-replication[16, 41]. They aim at simulating the emergence of living (i.e. self-replicating and possibly evolving) structures from a soup of non-living components under virtual physical laws, which are of more physical realism than those used in traditional pure abstract models of self-replication such as cellular automata and rewriting systems. Template-based replication imitates several important aspects of RNA/DNA molecules, which is directly relevant to the understanding of the origin of life and necessary of sufficient conditions for it to occur.

Besides the physical plausibility of transition rules, one may notice that there is another key difference between those template-based models and other traditional models of self-replication delegated by von Neumann’s universal constructor[48, 49]: the dependence of self-replication mechanisms on the structure of self-replicators. In template-based self-replication models, as well as some of the CA-based ones like [21], almost any kind of structure can replicate itself, since virtual physics laws are solely responsible for replication processes in those models. On the other hand, in self-replication models based on universal construction capabilities, the actual shape of self-replicating structures is crucial in the execution of self-replication tasks. This difference can be manifested when one tries to see if randomly created structures can self-replicate in these models;

they can indeed in the former, but not in the latter, because the latter class of models generally requires carefully designed complex structures that could virtually never arise in a probabilistic arrangement. This problem has its roots in the same issue as discussed in the context of the origins of cells in real biology.

To bridge the gap between these extremes in complexity has been a major challenge in self-replication studies. The Self-Reproducing Loop by Langton[22], one of the most famous and well-studied models for self-replication in CA, and its descendant models, reviewed in [39], were mostly intended to explore possibilities in this gap. Langton's goal in constructing the SR Loop was to reduce von Neumann's goal of *universal construction* to *simple self-reproduction*. In doing so, Langton's idea was to make the replication process more simple and robust, hence more closely resembling replication in natural biological systems. The structure which Langton chose to realize his idea was a *loop* consisting of a gene sequence and protective sheath, analogous to Turing's idea of an instruction tape and tape-reading machine. The importance of this separation was that it supplied the basic framework for constructing nontrivial self-replicating structures in a CA space, a framework which has been re-implemented in many subsequent models.

Since Langton's original model, many other artificial replicators have been constructed. Byl[6] simplified Langton's original CA to a smaller spatial structure using only 6 states. Chou and Reggia[7] studied the emergence of self-replicating loop structures where loop structure emerges from a random initial state. Morita and Imai[27] modified the basic loop structure into a self-replicating "worm", where phenotypic structure is used to encode part of the genotype; Sayama[38] later augmented this basic model by the inclusion of gene transmission to increase structural complexity, and Suzuki et. al.[44] has recently incorporated these ideas into a model for interaction-based evolution. In 1998, Sayama[35] introduced structural dissolution[36] (next section) and evolvability[37] (next chapter) to transition rules of the SR Loop, creating the SDSR Loop and evolloop, respectively. Along a parallel line of research, Hutton[16] recently designed an artificial chemistry in a CA space (Squirm3).

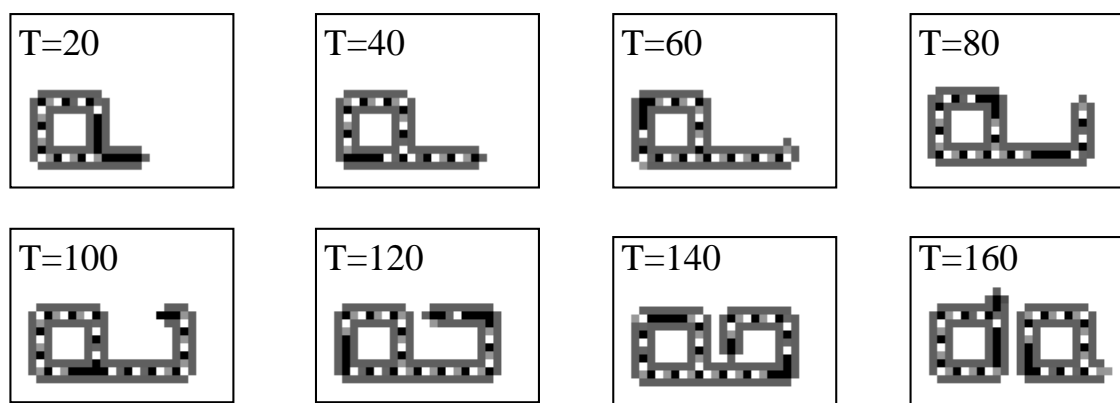


Fig. 2.3: Replication of the SR Loop

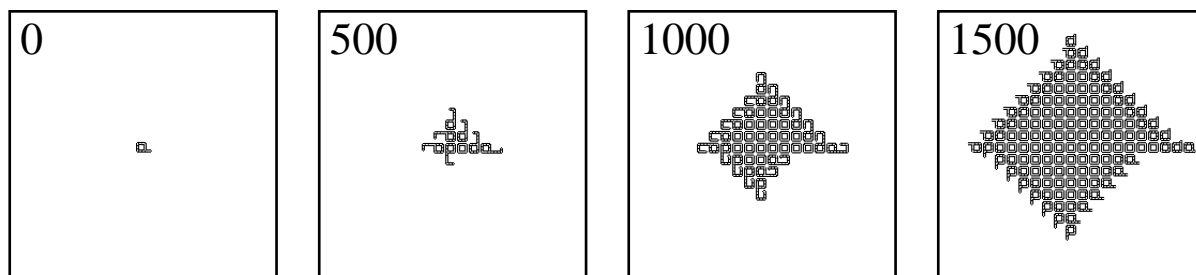


Fig. 2.4: Pattern formation of the SR Loop: Death by *functional failure*.

2.5.2 The SR Loop

The SR Loop[22] consists of a simple 8-state CA with a von Neumann neighbourhood. Transition rules for this CA obey strong rotational symmetry: rotation operations involve no replacement of states. The structure of the SR loop consists of an external sheath (state '2') surrounding a moving sequence of “gene” states ('4' and '7') which flow counter-clockwise within the sheath. In addition to these there is a quiescent state ('0'), core state ('1') through which the genes flow, and a number of states playing multiple roles in the replication process: indicators/messengers ('3', '5', '6'), a bonder ('3'), a sprout capper ('3'), a sprout generator ('3'), a sprout finisher ('6'), sprout guides ('5', '6'), and an umbilical cord dissolver ('5'). The overall result of this construction is a structure which nontrivially replicates itself in the CA space, as shown in Figure 2.3.

Figure 2.4 shows the evolution of a colony of SR Loops self-reproducing into an open domain. The large-scale structure recursively generated via the reproduction process is a diamond-shape formation with a *front* of live loops and *core* of loop “corpses”. The corpses consist of loops which — through a collision during the replication process — have lost their gene sequence and hence have been rendered unable to reproduce. This type of death, depicted in Figure 2.5, is referred to as *functional failure*[35, 36, 37].

Although significant in changing the course of research on self-replication in CA, Langton’s SR Loop has very limited functionality and predictable dynamics. There are two factors accounting for these deficiencies: nondestructive spatial occupation (death by functional failure) and lack of adaptability in replicator structure. The first of these factors is clearly evident in Figure 2.4, in which loop fronts may only move outwards into open space and hence never retrace the steps of dead loops; this is a result of the method of death in the SR Loop rule set (Fig. 2.5). The second factor is more subtle, relating the self-replicator as an *evolving* structure; its importance will become clearer in the next section with the introduction of the SDSR Loop.

2.5.3 The SDSR Loop

The first deficiency of the SR Loop mentioned above relates to its mechanism for death, which limits replication in periodic domains. Sayama[36] remarked on this problem and introduced a new state to overcome it. This state he called the *dissolving state* in reference to the role it plays in the replication process: faced with configurations which are not integral to the self-reproduction

process² the dissolving state appears, spreading through contiguous loop structures and dissolving them away into the background (quiescent) state. This process is depicted in Figure 2.6. In the fourth frame of this series ($T = 9$), the dissolving state appears in the top left corner in the tube of the loop. In the SR Loop, under the same configuration, a sheath ('2') state appears, resulting the destruction of only gene states as depicted in the last eight frames of Fig. 2.5. Behaviour of the dissolving state is quite distinct, resulting in death not by *functional failure* but by *structural dissolution*.

Structural dissolution introduces a drastic change to the dynamics of loop populations, as demonstrated in Figure 2.7. By clearing away dead loops, the dissolving state potentially allows for reproduction to occur over infinite time scales given a sufficiently large periodic domain. The dynamics of the spread of this new state also produce a level of complexity in the motion of reproducing loops in the CA space.

Discussing the significance of structural dissolution, Sayama remarks on a number of unique phenomena generated by the SDSR Loop which are not present in the SR Loop[35]. Here we emphasize some of these phenomena:

1. Partial structural dissolution allows loops to overcome difficulties during the self-reproduction process.

² Here we refer to configurations which are not integral to the basic process of self-replication in Langton's SR Loop.

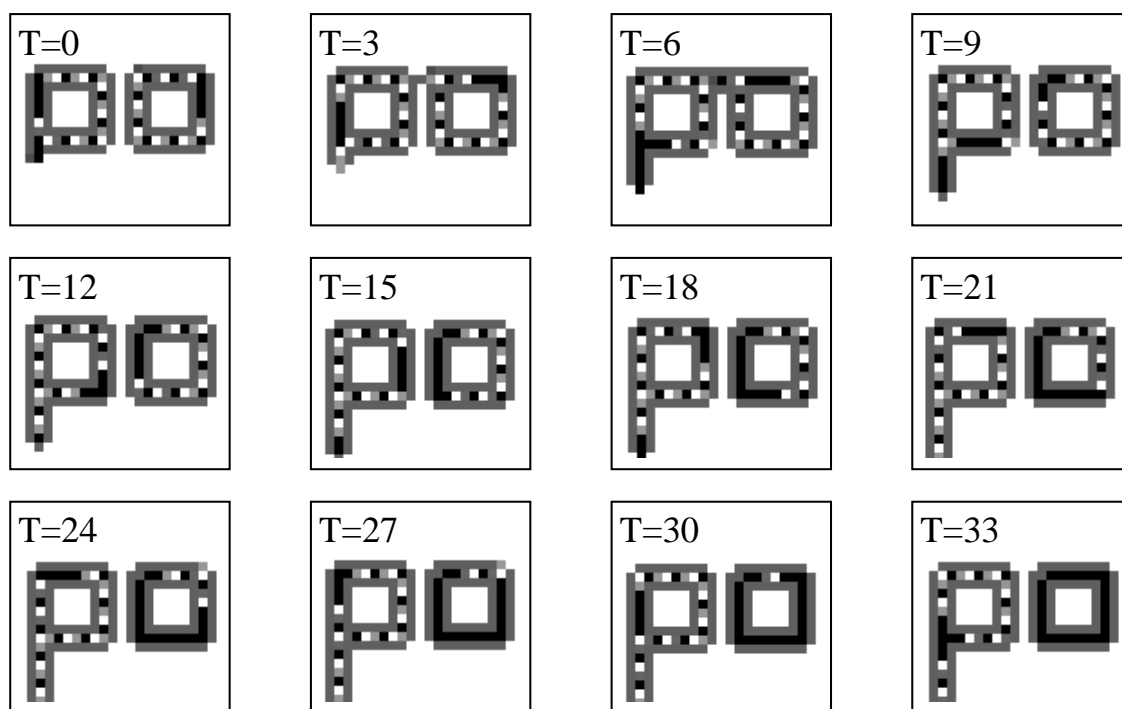


Fig. 2.5: Death by functional failure.

2. Loop populations may reproduce continuously in a finite periodic space (given a sufficiently large grid).
3. Competitive exclusion acts between loops of different types living in the same finite space.

The first and second phenomena result in a tendency of SDSR Loops to form “merged” loops through the direct interaction of phenotypes. Unlike the SR Loop, which is limited to continuous self-reproduction only in infinite domains, populations of SDSR Loops evolve in periodic domains over potentially infinite time spans. Given these conditions, SDSR Loops occasionally form bonded structures through rare interaction events. These structures do not have construction arms necessary for self-reproduction, hence they act as static corpses in the CA space. The appearance of these structures, however, are a hint that — given the space afforded by structural dissolution — the self-reproduction process has the potential for *adaptation*.

The last phenomena mentioned above introduces an interesting behaviour of SDSR Loops: when two loops of different sizes are introduced into the same periodic domain, offspring compete for

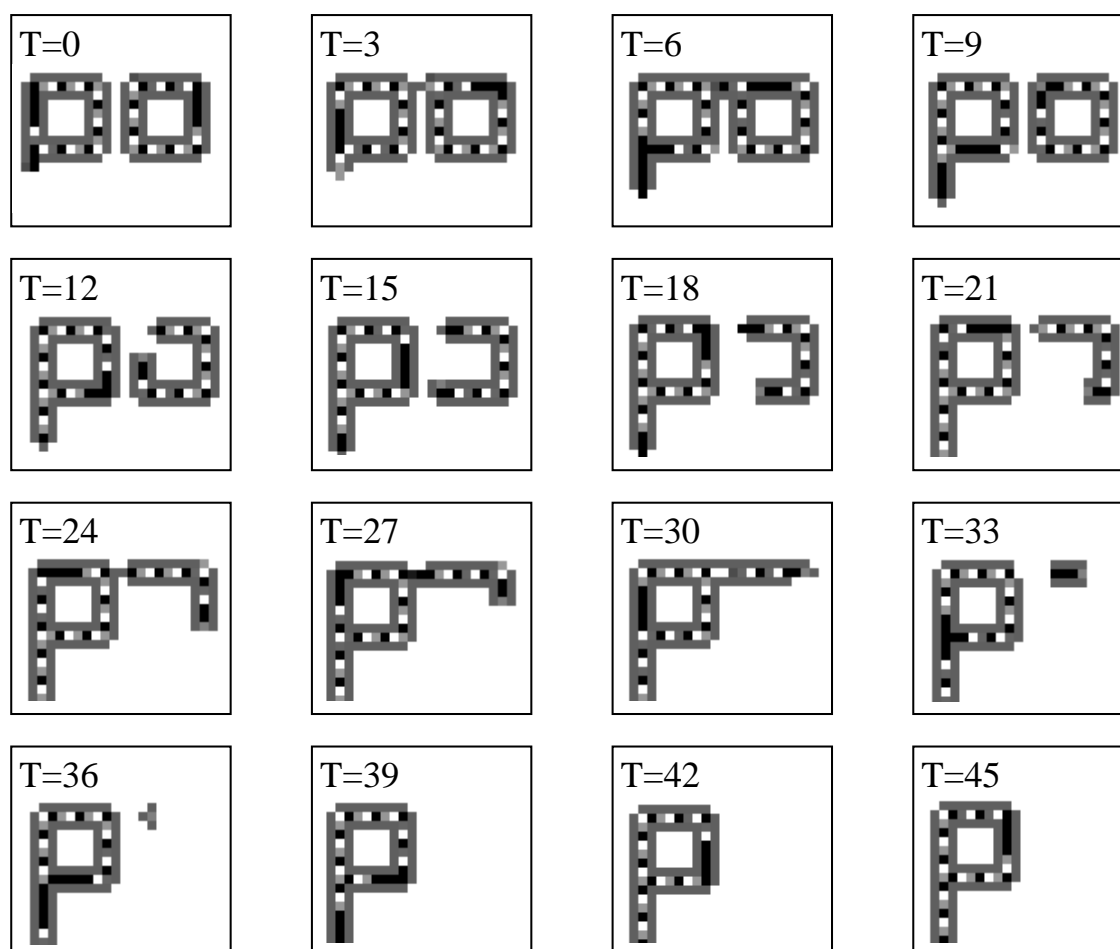


Fig. 2.6: Death by dissolution.

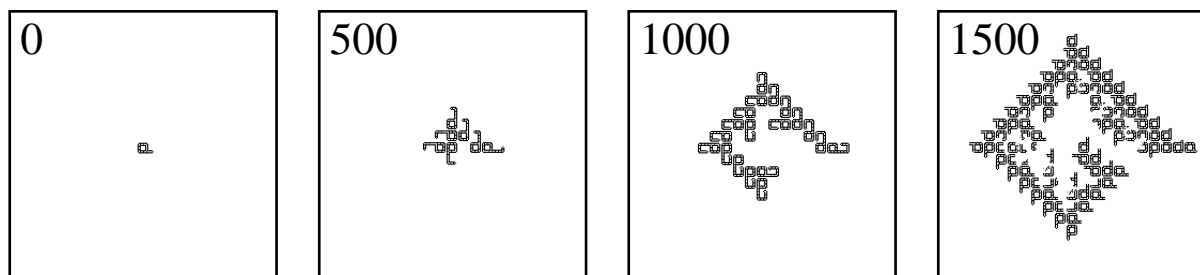


Fig. 2.7: Pattern formation of the SDSR Loop: Death by *structural dissolution*.

space through competition (collisions) enabled by the new dissolving state. Generally, this competition was found to favour smaller-sized loops due to their higher reproductive rate and structural robustness. In evolutionary terms, the new dissolving state provides the necessary environment for an emergent, artificial *selection* to act on differential fitness between distinct species of loops.

Given the results obtained above, the SDSR Loop exhibits the potential to satisfy the three conditions necessary for evolution to occur: replication, variation (mutation — though it creates only sterile structures in this model) and differential fitness (competition)[9]. Yet structural variability observed in this CA, encountered only in static, non-replicating structures, is clearly insufficient for evolutionary processes to emerge. In the next chapter we discuss the natural predecessor to the SDSR Loop, the evolloop, with CA rules which emphasize structural adaptability leading to variation in populations. It is this variation, coupled with continuous self-reproduction in periodic domains, which generates an evolutionary process, one whose selection criteria are emergent and, as we will show in Chapter 6, not yet fully understood.

2.6 Summary

In this chapter we discussed synthetic models of evolution relevant to the results to be presented later in this thesis. We began with an introduction to von Neumann’s Theory of Self-Reproducing Automata and emphasized the importance of *separation* for evolution, We then briefly outlined two lines of research with cellular automata, one which emphasizes their usefulness as a *computational device*, the other which stresses the power inherent in their *complex emergent dynamics*; the research in this thesis falls in the latter category. From this starting point we then went on to discuss the topics of complexity, emergence and multilevel evolution using a number of different model examples. We described Conway’s Game of Life and spatially embedded hypercyclic catalytic sets studied by [4, 34, 42]. We then reviewed models of self-replication in CA, starting from Langton’s Self-Reproducing (SR) Loop[22] followed by the Structurally Dissolvable Self-Reproducing Loop (SDSR)[36].

3. THE EVOLOOP

... *If we could populate a large area with multiple copies of such reproducing colonies, and introduce variation into at least the portion of the description that codes for the extra machinery, we would have all of the raw material necessary for natural selection to operate among variants and hence we would have a sufficient basis for the process of evolution.*

Christopher Langton, Studying artificial life with cellular automata [23].

3.1 Introduction

Models and evolutionary systems described in the previous chapter introduce the medium of cellular automata (CA) as a powerful tool for the synthesis of emergent phenomena. The realization of nontrivial self-replication (SR Loop) demonstrates that simple reproduction in CA is achievable using a minimal specification. Introduction of structural dissolution (SDSR Loop) grants the simple self-replicator the capacity for continuous reproduction in periodic domains, extending its applicability and greatly increasing the complexity of its dynamics.

Yet neither the SR nor SDSR Loop have the functionality to *evolve*. The first *evolutionary process* in a CA space — viewed here as a process in which self-replicators vary towards fitter individuals through natural selection — is realized by the “evoloop” [35, 37], a descendant of the SR and SDSR Loops. This self-replicator can be seen as the first CA to naturally generate both variation and selection leading to evolution; hence evolution in this system is *emergent*.

The emergent nature of evolutionary dynamics in the evoloop CA differentiates it from the great majority of existing evolutionary systems, both in Biology and in Artificial Life. In this chapter we begin by briefly discussing mutation mechanisms in artificial evolutionary systems. This leads us to the evoloop, a model which embeds the concept of emergent evolution into a deterministic CA space. We discuss states and transition rules for this CA, and conclude with a brief discussion on results of earlier experiments and how they relate to the work presented in this thesis.

3.2 An emergent mutation mechanism

In most existing artificial evolutionary systems [29, 30, 52], mutation is induced by pseudo-random noise in the form of bit flips. Although the results of such mutations are inherently unpredictable, the transition probabilities between different species are *predefined* by the mechanism and rate of mutation; hence the topology of transition space is prescribed in the specification of the system itself. Generally this means that gene sequences which have small differences will mutate into each other with high frequency, regardless, for instance, of gene (or program code) positioning.

While this approach has led to a wide variety of successful ALife models, the mechanism governing mutation in these systems is not emergent: for a given replicator, we know beforehand which other replicators are likely to appear through mutation.

The CA replicator discussed in the next sections utilizes a fundamentally different mutation mechanism. Given that the self-reproducing loops discussed so far (SR, SDSR) are deterministic and that gene sequences are structurally embedded directly in a CA space, a simple random bit-flip algorithm would be complicated to implement. A novel approach is introduced instead. Sayama[35] states: “the introduction of variation was achieved in implicit ways — through enhancing the adaptability of the self-reproductive mechanism of the SDSR loop and slightly modifying its initial structure.” More precisely, it is through the structural interaction of phenotypes (collisions) that gene sequences are modified; inclusion of this adaptive replication mechanism is what allows for evolution to emerge in a simple deterministic CA space. In the next section we discuss the states and transition rules of this self-reproducing loop with adaptive dynamics.

3.3 States and transition rules

As noted above, the evoloop CA has the same 9 states and von Neumann neighbourhood as the SDSR Loop. Hence in terms of number of CA rules it is no more complicated, with a total of $9^5 = 59,049$ rules. The structure of the evoloop also conforms to the basic SR/SDSR framework: an inner and outer sheath of square or rectangular shape filled by a moving sequence of signal states. The mechanism for self-replication is likewise analogous to earlier loop models: an external arm extends, rotates, and retracts over 90 degree counterclockwise intervals to form a child loop. Coordination of the duplication process is controlled via a genotype encoded in a gene sequence within the sheath of the loop. The difference between the evoloop CA and the SR/SDSR loops relates to changes introduced to the state-transition rules. We will only briefly discuss relevant properties of this new rule set in this thesis; for a more detailed description we refer to [35, 37].

Cellular automata states for the evoloop can be grouped according to their function. These are presented in Figure 3.1. Background (0), core (1) and sheath (2) states are the fundamental elements of this CA and play the role of a passive environment. Signal states (3 through 7) govern the dynamics of the self-replication process, often playing multiple context-dependent roles. States 4 and 7 are of particular significance; these are the *gene* states which decide the identity of a loop’s offspring. The relative placement of genes and their distribution within the core decide the basic properties of the loop. The dissolving state (8), adopted from the SDSR Loop, extinguishes any sheath structure which neighbors it in the CA domain, leading to the creation of free space into which other loops may reproduce. This ninth state expands the limited applicability of the SR Loop on finite periodic domains, granting to SDSR the capacity for continuous self-reproduction. The evoloop inherits this same quality.

State-transition rules of the evoloop are developed from those of the basic SR/SDSR rule set with adaptability in mind [35]. As such, the number and variety of situations in which structures of this CA can retain regular operations (discussed in earlier work [37]) is substantially higher. The most obvious morphological difference between the evoloop and SR/SDSR loops is exhibited in the length of its umbilical cord, which is substantially longer in the evoloop due to the explicit arm growth by gene 7’s in its body. This more adaptable form of arm extension allows the evoloop to readily change the shape of offspring loops given obstacles or interactions with other loops; hence

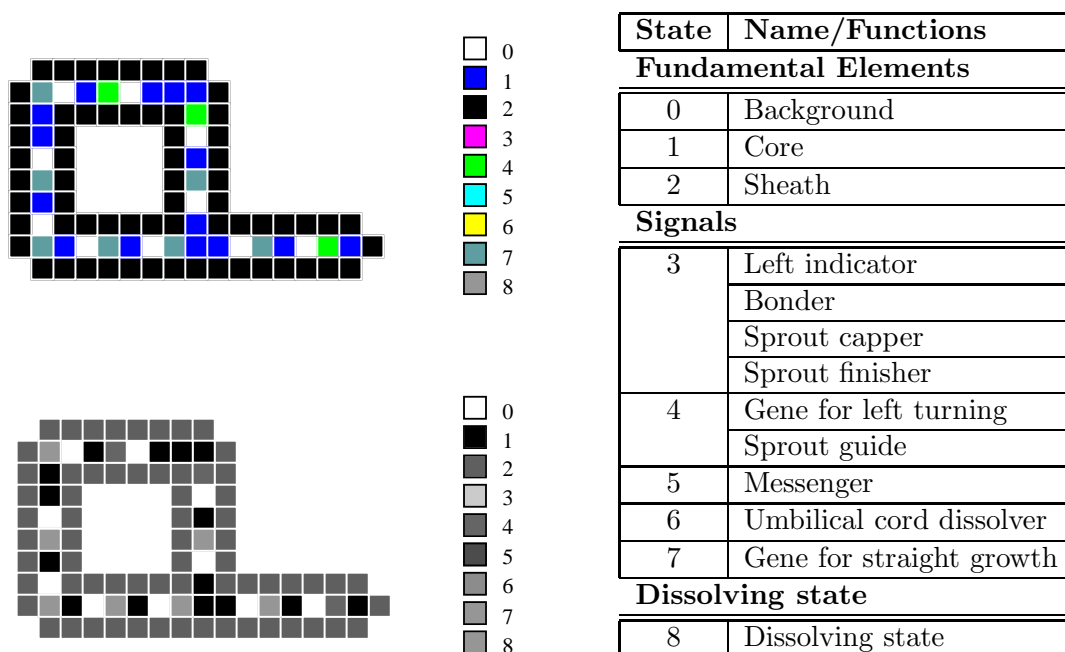


Fig. 3.1: Evoloop states and functions.

a variety of loops are often observed to co-exist in a single space. Some examples of evoloops of different sizes are shown in Figure 3.2.

Figure 3.3 outlines in detail the self-replication process of the evoloop. The sequence of genes (4 and 7), background states (0) and signal states (2) which rotate in the parent loop also propagate outwards to form the structure of the offspring. The first rotation constructs the umbilical cord which separates the parent from the child; each subsequent rotation constructs a side of the offspring loop. Divisions between segments of the offspring structure (limbs) are signaled by the turning genes (4), which alter the orientation of the tube. Typically, mutations are induced when an extending arm encounters another loop, at which point a propagating growth gene is “lost” and the length of a limb is shorter than it otherwise would be. However there are also loops whose gene sequence generates offspring which are different than itself. The distinction between these two types of mutation are explored more thoroughly in Section 5.2.

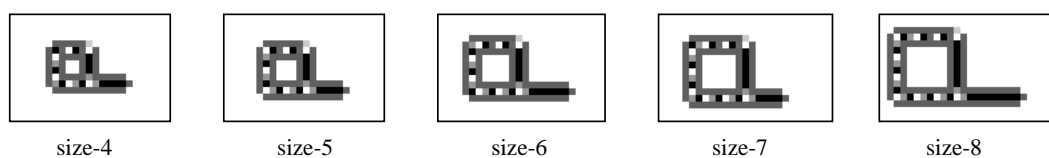


Fig. 3.2: Evoloops of different sizes.

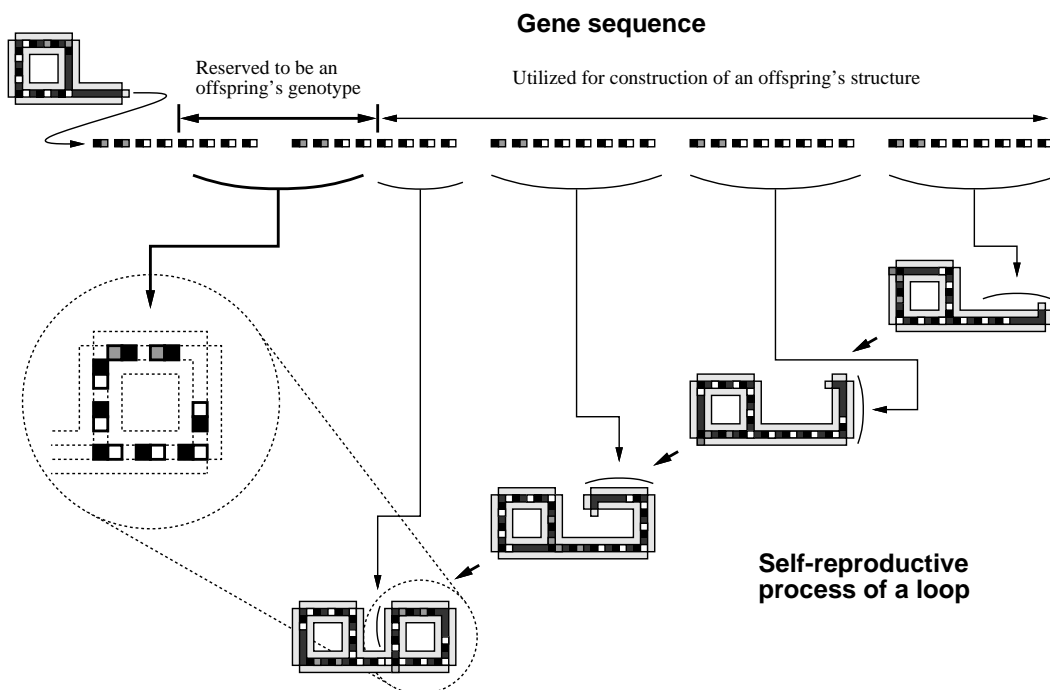


Fig. 3.3: Manner of utilization of a gene sequence in a loop (courtesy H. Sayama).

3.4 Emergent evolutionary dynamics

The surprising result of the replication process described above is that — with no explicit mechanism introduced to promote evolution — loops in the CA space alter their shape and gene sequence through direct interaction of their phenotypes (outer sheath), inducing a selection process based mainly on a loop's size (length of inner sheath). [37] On small-scale periodic domains (200^2 up to 2000^2 grid points) the observation was made [35] that this selection process favours loops of a smaller size. An example of an emergent evolutionary process in this system is shown in Figure 3.4, in which a population of size-8 loops is observed to evolve over time into loops of size 4 (smallest possible size, given this initial configuration). These results constitute the first evidence that self-replicating cellular automata have the potential to generate emergent evolutionary processes through variation and natural selection. The significance of this result is all the more powerful when one considers the simplicity of the evoloop: with 9 states and a 5-cell neighbourhood, the size of the rule set for this CA is $9^5 = 59049$ or roughly sixty thousand, many orders of magnitude less than other self-replicator CA's [5, 7].

The emergent evolution observed in [35, 37] was observed to consistently favour smaller-sized loops. From these results the conclusion was reached that “smaller individuals were naturally selected thanks to their quicker self-reproductive ability, and the whole population gradually evolved towards the smallest ones.” [35] In this thesis we demonstrate that this result only holds in the case of a limited subset of evoloop species, and that there is a large class of species for which selection does not necessarily minimize loop size. To arrive at this result, we introduce in the next chapter a

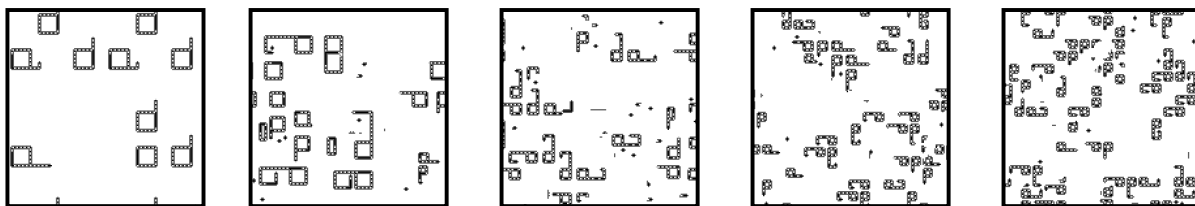


Fig. 3.4: Evolution over time of evoloop colonies to smaller size.

detailed identification scheme, with which a better understanding of loop taxonomy is achievable. In Chapter 5, we show that this new level of detail leads us to a clearer understanding of complex evolutionary dynamics in populations of evolving loops. Building on these results, in Chapter 6 we analyze a new class of loop species which demonstrate complex evolutionary dynamics distinct from the species analyzed in previous work. We show that this class of loops includes species for which selection acts not only on reproduction rate, but also on emergent dynamical properties which are not yet fully understood.

3.5 Summary

The model of study for this thesis, the evoloop[35, 37], was introduced in this chapter. We began by drawing a connection to earlier models of self-replication in CA discussed in the previous chapter: the SR Loop[22] and SDSR Loop[36]. The distinction was emphasized between models which introduce random but predefined mutation operators, such as Tierra[30] and its descendent models, and those which allow the system to generate *emergent* mutation events, such as the evoloop.

A cursory overview of states and transition rules was then included. The emphasis in this overview was on aspects of CA mechanics relevant to this study, in particular the role of gene and dissolver states as well as basic operations governing the self-replication process. Results of previous experiments — which observed size-based differential fitness and evolution to smaller-sized loops — were briefly summarized and related to the work to be presented in later chapters of this thesis.

4. NEW TOOLS OF ANALYSIS

4.1 *Motivation*

The question of how to describe emergence and in particular how to understand the flow of information between micro- and meso-scale structures in complex adaptive systems is a pivotal one[2, 4, 7, 8, 18, 32, 33, 34, 44]. In considering traditional theories of phase transitions and critical phenomena, Crutchfield[8] states: “What these theories provide is a set of coarse tools that describe large-scale statistical properties. What they lack are the additional, more detailed probes that would reveal, for example, the architecture of information processing embedded in those states”. It is to the design of these probes that we focus our attention in this chapter.

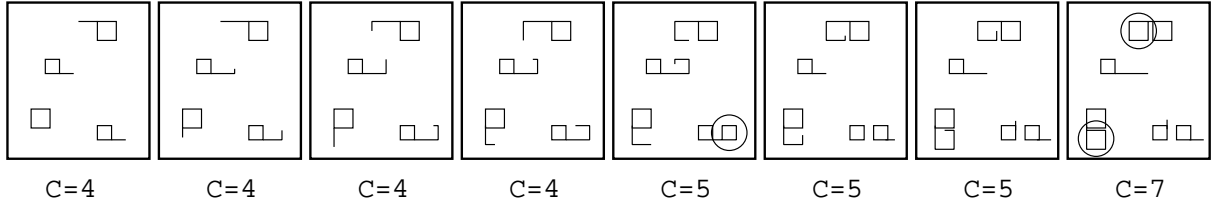
4.2 *The evoloop as a biological system*

Evolutionary dynamics in the evoloop CA are emergent, hence no criteria for basic evolutionary mechanisms (birth, death, mutation, variation) are predefined; these ideas are left up to the observer to impose. This makes the evoloop in some ways more like a real biological system than existing artificial models. The distinction is that dynamics of the evoloop occur in a deterministic CA space where a complete historic record is available. Distinguishing relevant information from this potentially huge data set presents an interesting challenge for analysis. On this topic, Sayama[37] remarks: “This means that observation of such a world needs a considerable amount of computation ... in order to detect any meaningful information from the configuration of the whole CA space.”

In the following sections we show that this statement need not be true, by showing how the computational overhead required for analysis can be minimized to levels well below the time spent on CA update rules. We begin by discussing a detection mechanism for *birth* which is vastly more efficient than previous counting routines for this model. Using this mechanism, we show that a fully complete and unambiguous identification scheme can be implemented with negligible overhead in computation time. This is followed by the introduction of a new CA layer exclusively for analysis which is used to track the “footprint” of live loops. By adding a condition for *death* of a loop, we show our new method replaces the “time-driven” loop-counting routine implemented in previous work by an “event-driven” mechanism which is highly efficient and fully descriptive.

In the second half of this chapter the second (analysis) CA layer is augmented by a genealogy-tracing mechanism. A “trace” is introduced as a way to track the identity of a parent loop during replication. Using this mechanism, birth events are described by the time of birth, location of birth, parent species, and child species. In later chapters this complete description allows us to fully track genealogical evolution of loop populations, leading us to the discovery of a new and very different class of loop species.

Time-driven detection



Event-driven detection

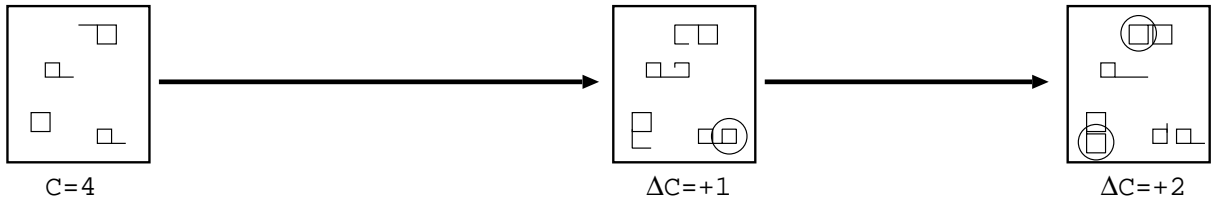


Fig. 4.1: Time driven and event-driven detection for the evoloop CA

4.3 Detection and identification

The method introduced in this section for loop counting requires three basic components for implementation:

1. A mechanism for identifying the *birth* of a new replicator.
2. A storage space to maintain a record of all live replicators.
3. A mechanism for detecting the *death* of live replicators.

Note that the above steps are general in scope, and together constitute what we mean by an *event-driven* detection scheme. This scheme should be contrasted with what we refer to as a *time-driven* detection scheme, in which replicators are (re)-counted at each iteration. The scheme we introduce here differs fundamentally from previous methods of loop-counting in that only *differences* in population are detected (see Fig. 4.1). This in turn results in a vastly more efficient computational method of analysis for the evoloop.

4.3.1 Birth detection mechanism

In structural terms, the evoloop can be defined by the configuration of CA states forming its basic shape. As described earlier in Chapter 3, these comprise an outer sheath of state '2' cells surrounding a gene sequence of state '0', '1', '4' and '7' cells. The sheath and gene sequence together form the *loop* structure and extensible *arm*. During replication this arm extends outwards to become the structure of the offspring, then retracts towards the parent loop. It is this retraction mechanism which forms the basis for the birth-detection mechanism which is introduced here.

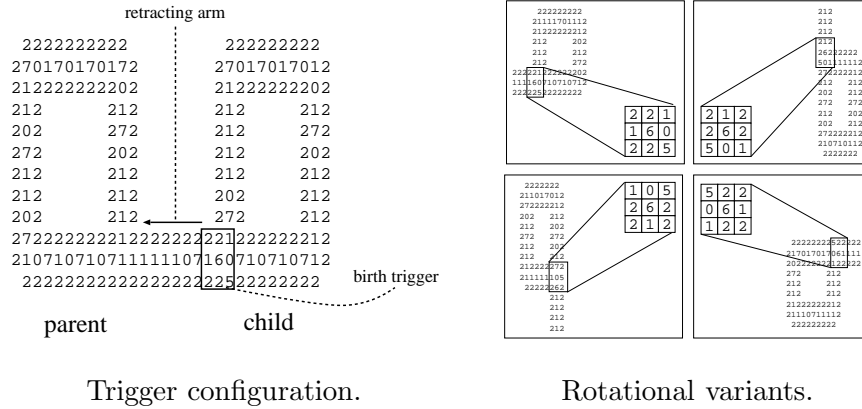


Fig. 4.2: Trigger for birth detection and rotationally symmetric variants.

Trigger configuration

To begin, we require arbitrary but well-defined criteria to pinpoint time-of-birth in a precise and unambiguous manner. For this purpose we use the *umbilical cord dissolver* (state '6' of the evolop CA), named for the role it plays in dissolving the extended arm following loop replication. The choice of this state for signaling birth events is a logical one: appearance of this state was constructed to coincide with retraction of the loop arm during duplication, hence its appearance is a *necessary* condition for new loops to appear in the CA space[37]. This criteria can in fact be further simplified by noting that a state '6' cell always emerges from a state '0' cell during arm retraction, hence we need only monitor the '0' \rightarrow '6' transition.

Appearance of a state '6' cell alone is not a *sufficient* condition to guarantee that loop replication has truly occurred. This guarantee requires further conditions on our detection mechanism, for which we use the Moore neighbourhood states surrounding the state '6' cell. The configuration of states which appears at the moment of replication is illustrated in Figure 4.2, as well as its symmetrical rotations. These configurations define our birth "trigger" function. Appearance of a state '6' cell in one of these configurations leads to the *birth* of a loop.

Loop identification

To complete our detection mechanism, we require a routine to assess the structural integrity of the child loop at the moment of birth. We would also like to identify this new loop and classify it according to its structural makeup. These two steps are carried out simultaneously during a single gene sequence "traversal" using a routine which records gene states inside the sheath of the newborn loop. As these states are read from the CA space, a consistency-check is performed according to criteria for what we define as a *loop*. In Figure 4.3, the basic mechanisms of this routine are illustrated: *detection* (by the birth-trigger configuration) and *identification* (as discussed below).

For present purposes, we will only loosely describe what we mean by a *loop*; in the next chapter we define this structure more rigorously. Intuitively, a loop is defined by a rectangular sequence of cells containing states '0', '1', '4' and '7', with '4's and '7's always followed clockwise by a state '0' cell, directly enclosing an inner layer of state '2' cells. The sequence of state '0', '1', '4' and '7' cells

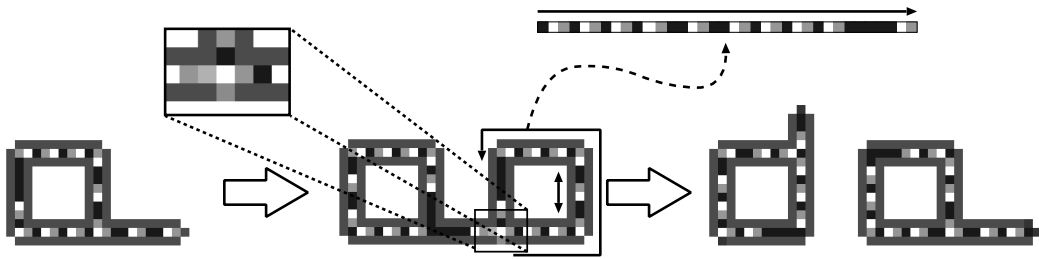


Fig. 4.3: Loop detection and identification.

in a loop together constitute what we refer to as a *genotype*. The length and width of the inner sheath we call a *phenotype*. Note that the outer sheath of state '2' cells surrounds the loop but is not necessary for (and hence not included in) its basic definition.

Once birth has been signaled, our routine for birth-detection must traverse the loop gene sequence to check for structural integrity. It is possible that, during replication, the child loop has collided with a neighbouring structure and no longer satisfies our structural criteria for a loop. Various other anomalies are also possible. Three examples of trigger-signaled non-birth events are shown in Figure 4.4, indicating that the appearance of the configuration depicted in Figure 4.2 may incorrectly signal a non-loop structure.

To filter out non-loop structures such as those in Figure 4.4, we introduce the following routine for identification:

1. Upon detection of the one of the trigger-configurations in Figure 4.2:
 - (a) Initiate the traversal position $p = p_i$ to the state '0' cell adjacent to the umbilical cord dissolver.
 - (b) Initiate the direction of traversal d to point along the direction to the state '0' cell from the neighbouring state '6' cell.
 - (c) Initiate the current limb to zero: $k = 0$.
 - (d) Initiate the gene sequence string g to NULL, and phenotype $p = (l, w)$ to $(0, 0)$.
2. Begin loop traversal:
 - (a) Read the CA state at position p :

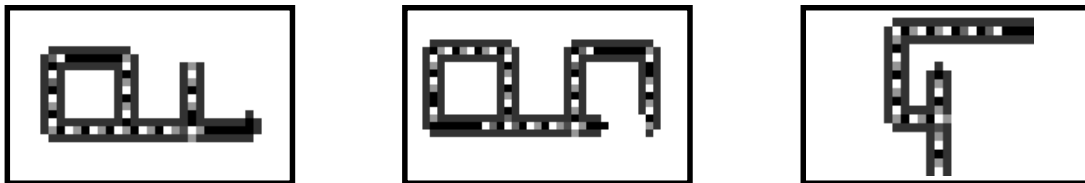


Fig. 4.4: Anomalies of birth-detection

-
- i. If the state at position p is a sheath state ('2') or left indicator ('3'):
 - A. Move one step back (along $-d$).
 - B. Rotate the direction d clockwise 90 degrees
 - C. Increment the current limb ($k++$). If $k = 4$, check that $p = p_i$. If not, structure is not a loop. If yes goto 3.
 - ii. Otherwise:
 - A. If the state at p is '0', '1', '4' or '7' append it at the end of g , increment the position along the direction d , and increment l if $k = 0$, w if $k = 1$.
 - B. Otherwise exit unsuccessfully.
 - (b) Move p one step along the direction d .
 - (c) Goto 2b.
3. Check g for sequence consistency:
- (a) Every '4' and '7' should be preceded by a '0', and '0's should only appear in the sequence following a '1' state and followed by either a state '4' or '7'.¹

In the actual implementation of the above algorithm, additional checks are performed to eliminate runaway processes caused by various anomalous structures with unusual configurations. These are omitted for brevity here, however the basic nature of the algorithm is represented in the outline above.

Having successfully completed a gene sequence traversal, the birth-detection algorithm returns with a genotype g and phenotype p . The pair $\theta = (g, p)$ we call an *identifier*. Loop identifiers uniquely distinguish different loop *species* in the CA space according to their configuration at birth. For simplicity, we map this identifier to an index k . The number k we use in the following section to track live loops in the CA space.

Labeling scheme

For convenience, we would like to map the gene sequence g returned by the traversal algorithm to a compact number. By the criteria given in the sequence consistency check (3 above), a "raw" gene sequence will contain any number of '1' states interspersed by '071' and '041' triplets. As an example, we take the sequence

07107107107110710410410711111071

This gene sequence is contained in a 7×7 loop (see Figure 4.5). To efficiently translate the sequence of '0', '1', '4' and '7' states to a unique label we first group '071' and '041' triplets and replace them by symbols. We use the symbol **T** for the turning triplet '041', **G** for the straight growth triplet '071', and **C** for free core states '1'. Hence we apply the map: $071 \rightarrow G$; $041 \rightarrow T$; $1 \rightarrow C$, resulting in the reduced representation **GGGGCGTTGCCCGG**.

To express this string more compactly, we convert it to a binary representation. We want that the first (leftmost) triplet to be represented by the lowest order bits; in this way, we can easily

¹ By these conditions, '7's and '4's must always appear in '071' and '041' triplets.

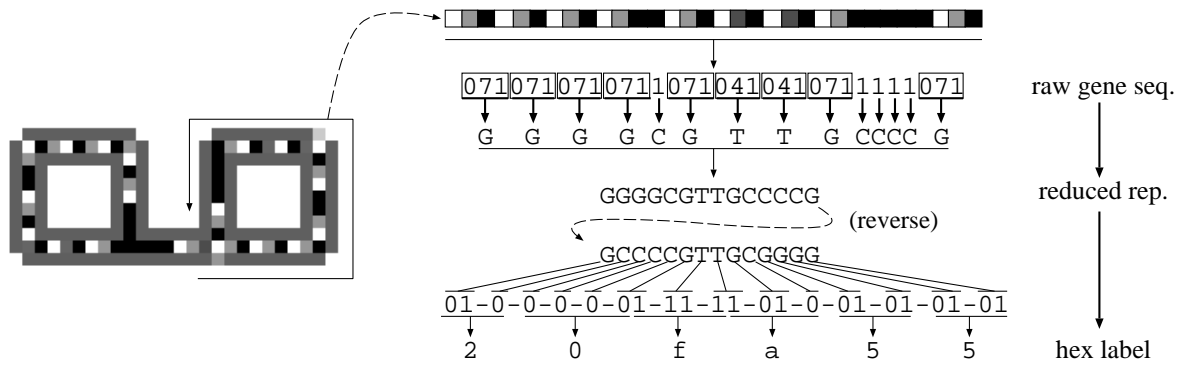


Fig. 4.5: Labeling procedure for the evolop gene sequence.

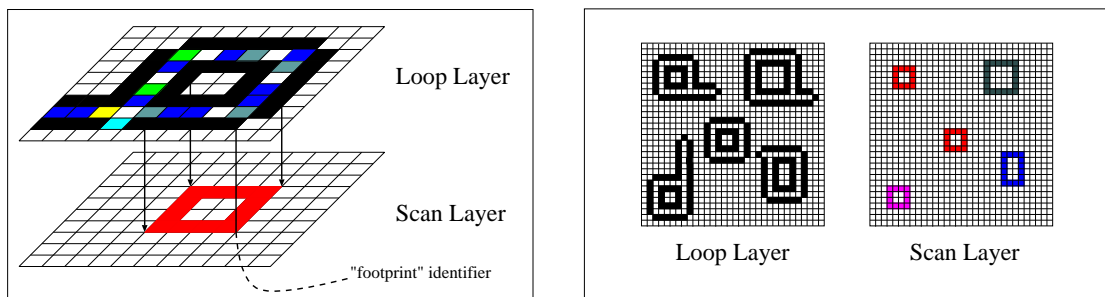
convert back moving from lower to higher order bits in the bit sequence. Hence we traverse from right to left, replacing Cs by “0”, Ts by “11” and Gs by “01”. This sequence thus becomes:

01-0-0-0-0-01-11-11-01-0-01-01-01-01

This we convert to hexadecimal: 20fa55. To convert back, we start from the lowest order bits and stop when we have reached the end of the sequence (set by the phenotype). Thus this loop type is identified uniquely, with a suffix of its phenotype (length and width), as 20fa55/7×7.

4.3.2 The “scan” layer and the evolop “footprint”

Having detected and identified a new loop, we require a method to maintain a record of the newborn’s presence and position in the CA space. This is the “storage space” mentioned at the beginning of Section 4.3. For this purpose we attach a new CA layer which we call the “scan layer”, as illustrated in the two frames of Figure 4.6. This layer serves as a tool of analysis and in no way interacts with the CA layer on which loops evolve (which we henceforth refer to as the “loop layer”).



“Footprint” of newborn loop.

Loops identified in scan layer.

Fig. 4.6: Loop layer and scan layer. Colours in scan layer represent distinct types of loop.

When a new loop is detected and identified, the index k assigned to its identifier is used to mark its presence in the scan layer, directly beneath the inner sheath structure in the loop layer; this is shown in the second frame of Figure 4.6. Note that different *types* of loop are distinguished on the basis of their configuration at birth (genotype and phenotype) and hence are assigned different indices in the scan layer; this feature results in loop footprints marked by different numbers, mapped to the colour of scan-layer cells in Figure 4.6. When a new loop is born and entered into the scan layer, a counter representing the number of that type of loop is incremented; criteria for decrementing this counter are described in the next section.

Though the idea of a scan layer was first introduced in earlier work on the evolloop CA, the implementation described here differs fundamentally from the original version. The basic difference lies in the *event-driven* nature of our method of identification; loops are identified and marked only when the trigger-configuration of Figure 4.2 is detected. The method described here is moreover *completely descriptive*, in the sense that its identifier provides sufficient information to completely reconstruct the loop at birth. This level of detail provides a powerful and necessary tool for analysis.

4.3.3 Death-detection mechanism

To complete our detection and identification scheme, we require criteria for loop *death* in order to remove footprints from the scan layer and decrement the loop-type counter. Given the nature of the scan-layer footprint described above, death-detection is made very simple. By imposing criteria for death of a loop in an analogous manner to the birth configuration described in Section 4.3.1, we can also significantly minimize computation time for this mechanism.

Figure 4.7 illustrates the process of death-detection. Since our goal is to monitor the structural integrity of loops in the CA space, our criteria for death is simple: if one of the inner sheath states (state '2' cells) is dissolved away (i.e. undergoes transition to the dissolver state '8') and there is a footprint beneath it then we consider the loop to be no longer alive and remove its footprint from the scan layer. Since this check only requires that we monitor the transition '2' \rightarrow '8', negligible computation time is required for this task.

With birth-detection, monitoring of live loops, and death-detection implemented, we have all the necessary components for a powerful analysis tool. Note that, given the nature of this method, no additional analysis routines are required during cell update; we only require monitoring for the transitions:

1. '0' \rightarrow '6' : Appearance of umbilical cord dissolver.
2. '2' \rightarrow '8' : Dissolution of inner sheath state.

Using carefully constructed `if/then` constructs, these type of checks require virtually no extra computation time. Since both transitions 1 and 2 are relatively infrequent events, these check are rarely performed compared to normal cell updates.

4.4 Genealogy tracing

The method of detection and identification described in the previous section is effective in monitoring the appearance and disappearance of loops and tracking populations of different types (species).

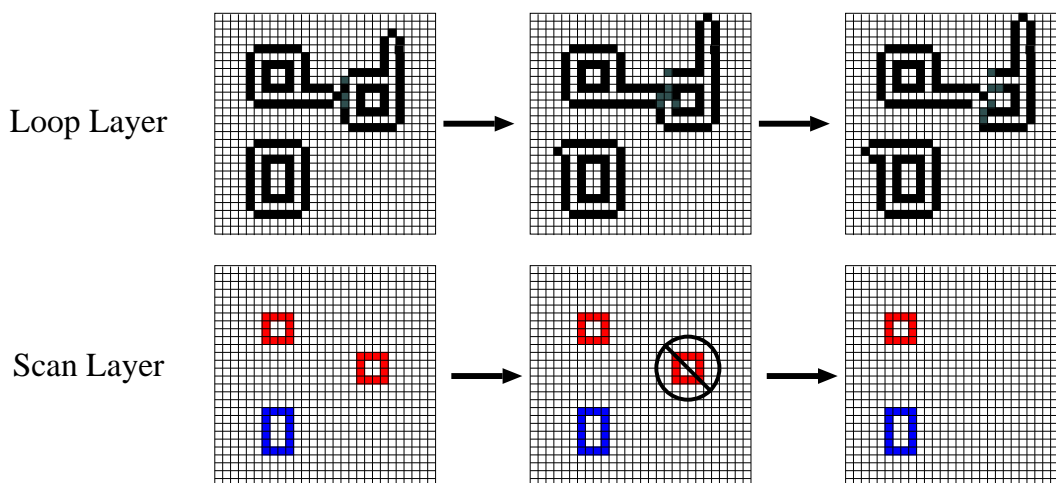


Fig. 4.7: Collision leading to death-detection. In middle frame inner sheath sheath state dissolves away (transition '2' \rightarrow '8'), triggering a death of loop-type “red” and leading to the removal of its footprint in the scan layer.

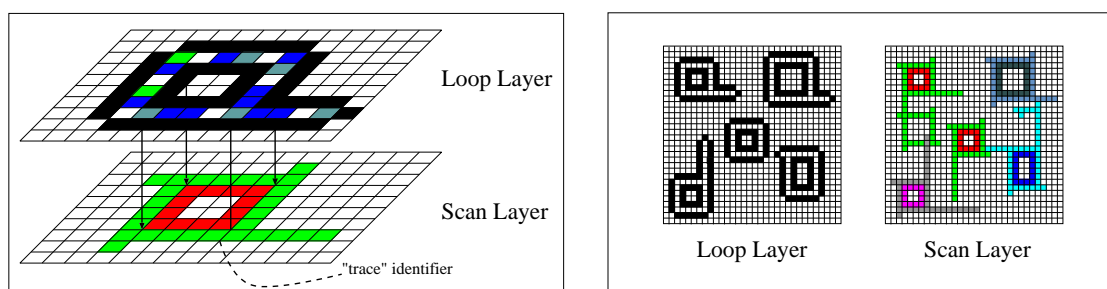
However no mechanism has yet been introduced to track evolutionary *links* between these species, links which are critical to evolutionary dynamics of the evolooop CA. The importance of this fact can best be understood by remembering that evolution of the evolooop is *emergent*. As such, there is no mutation routine imposed on gene sequences, no random noise introduced during replication, no “predictable” modifications of any kind. In terms of its behaviour, this distinguishing feature of the evolooop makes it more like a real biological system than any artificial model. Hence our analysis routines must “catch” the evolutionary steps as they occur. To this end, we devote a section to augmenting the detection/identification method introduced in the previous section with a genealogy-tracing mechanism.

4.4.1 The “trace” identifier

The challenge we face in tracing genealogy is to find a simple mechanism through which parent and child species may be easily and efficiently identified during replication events. As we have already added an extra CA layer for identification, a natural choice would be to use this same layer for genealogy analysis. Hence we would like to find a mechanism which — in a straightforward manner — can detect the identity of a parent species at the moment of birth (as defined by the trigger-configuration of Figure 4.2).

In relation to the above requirements, there are some subtle caveats to be considered:

1. More than one parent may combine to form a child loop (sexual reproduction).
2. The parent(s) may no longer be alive at the moment of birth of a child. In this case, footprint(s) will no longer be present in the scan layer. This may happen if the parent structure(s) is/are dissolving away as the arm(s) closes to form a new loop.
3. Parent arm(s) may follow a long and complex path(s) before closing to form child loop.



“Trace” of parent loop.

Loop identities traced in scan layer.

Fig. 4.8: Genealogy identification by “trace”.

The above subtleties make genealogy detection an inherently tricky business for the evolloop CA. As stated above, this is again a consequence of the fact that evolution in this model is emergent and hence not strictly “defined” in any prescribed way. To accommodate this unpredictability we need a method of analysis which is as flexible as possible, while simple enough to implement with minimal computation overhead.

The term we use for the method we introduce here is that of a loop “trace”. The idea is quite straightforward: as loops extend arms in the CA space, they copy forward in the scan layer a number which represents the identifier for their species. This “trace” copy mechanism can be implemented in much the same way as detection mechanisms of Section 4.3, by defining certain transition events which occur during arm extension. Using this idea, the identifier of a parent loop can then simply be “read” from the trace identifier in the scan layer when a new loop is born.

In Figure 4.8, the new trace identifier is introduced to the scan layer beneath core/signal states of the gene sequence in the loop layer. In the implementation used for experiments, footprints and traces are distinguished by assigning them positive and negative values, respectively. Note that, unlike the loop footprint, a trace need not be removed following loop death as it plays no role in loop counting. Moreover, a loop’s trace identifier will continue to be copied forward by its extended arm even after its inner sheath has been dissolved away; as such, subtleties 2 and 3 (above) are immediately resolved by this mechanism.

To implement the idea of a trace, we need to identify the state transitions which occur as a loop arm is extending outwards and use these transitions to copy the trace identifier in the scan layer. The four basic transition sequences involved in this forward-copying are depicted in Figure 4.9: arm extension (forming umbilical cord and offspring structure, respectively), left turning, and tip bonding. For arm extension and tip bonding, the important transition takes place in the middle and bottom frames of Figure 4.9, at which instant the trace identifier is copied forward as the arm is extended. Left turning requires that we copy the trace sideways (at an angle of 90 degrees counter-clockwise with the arm orientation).

The transitions we need to monitor for these sequences are:

1. '0' \rightarrow '3': Extension of sheath during arm extension (umbilical cord formation).
2. '0' \rightarrow '2': Extension of sheath during arm extension (offspring structure formation).

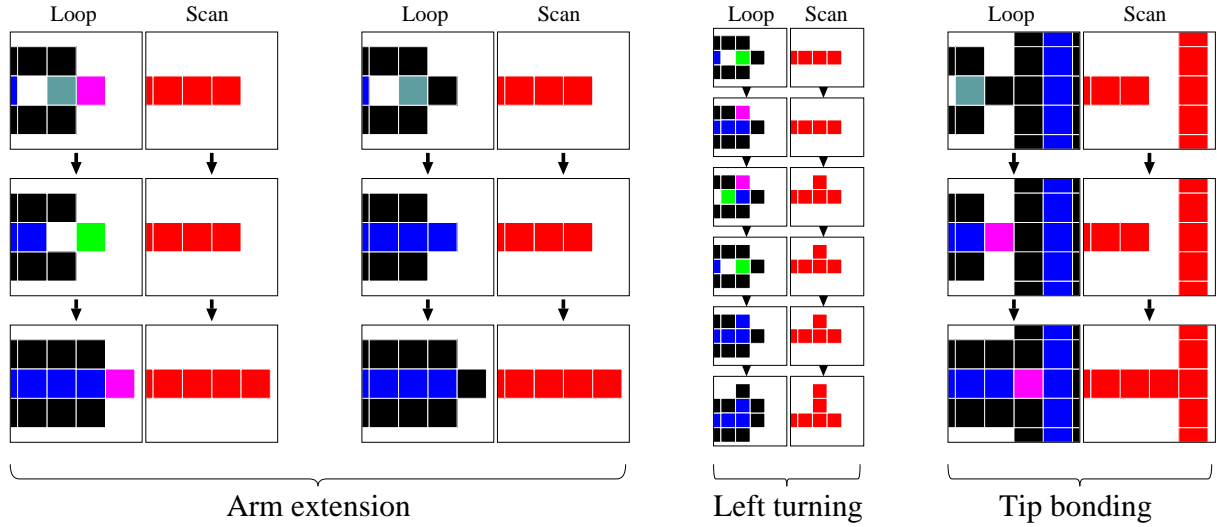


Fig. 4.9: Transitions which copy trace identifier forward in scan layer.

3. '2' \rightarrow '3': Left turning / Tip bonding.

Note that the effect of this trace copy procedure is to create rectangular traces immediately surrounding the footprint identifier, with single-cell trace sprouts extending from each corner. These sprouts are important as they initiate trace copying in next-generation replication events of the newborn offspring. This action of trace copying during a replication event is depicted in Figure 4.10. In this case we see that the offspring loop is structurally different from the parent, hence it has a different identifier and corresponding footprint/trace. Upon birth (trigger-configuration of Fig. 4.2), this new loop is identified and its parent species picked up from the trace identifier in the scan layer. Following this identification procedure, the parent trace is overwritten by the offspring trace and CA updates continue.

4.4.2 Multiple-parent detection

The method of genealogy tracing described thus far has been demonstrated to overcome the last two difficulties highlighted at the beginning of Section 4.4.1: namely, it identifies parents which have died prior to the moment of replication (no footprint identifier), and it can follow the possibly complex growth of loop arms during the formation of new loops. The first caveat, that child loops may have multiple parents, is also handled implicitly by this method. This feature is illustrated in Figure 4.11.

That multiple-parent replication (generally at most two-parent) occurs in the evolloop CA is due to the adaptability built into its state-transition rules. As with single-parent (asexual) reproduction, evolution in this process is emergent, hence 2-parent replication events cannot be predicted in any prescribed way. Detection and genealogy tracing must therefore be flexible to accommodate the variety of configurations which may emerge during the formation of new loops.

When two loop arms connect to form a new loop, trace identifiers written to the scan layer are left intact up to the moment of birth. Thus when parent-identification is performed, multiple

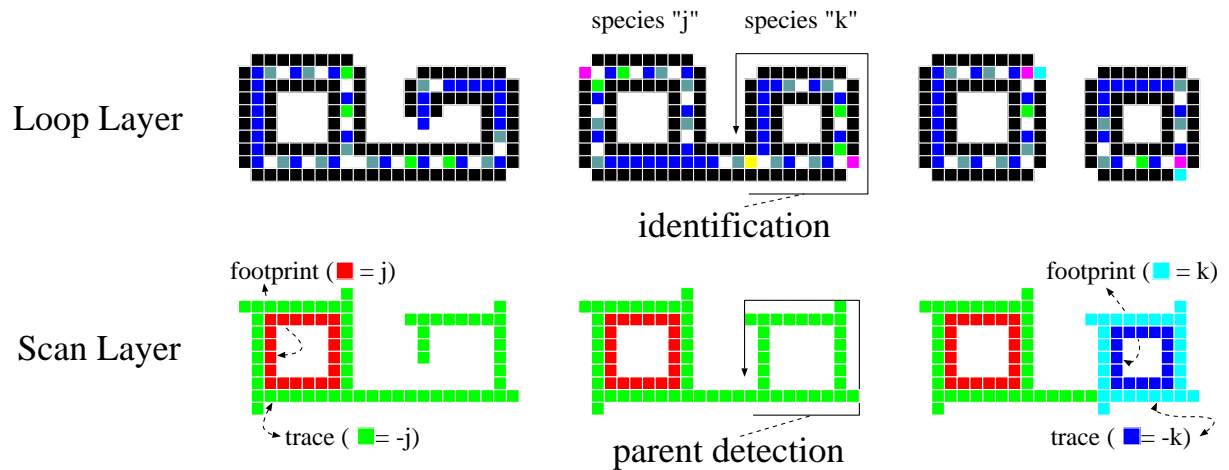


Fig. 4.10: Replication event leading to new species “k” from parent species “j”. Identification of newborn loop species and parent tracing is performed in middle frame. Footprint and trace in scan layer are represented by green and red cells for left loop, dark blue and light blue for right loop.

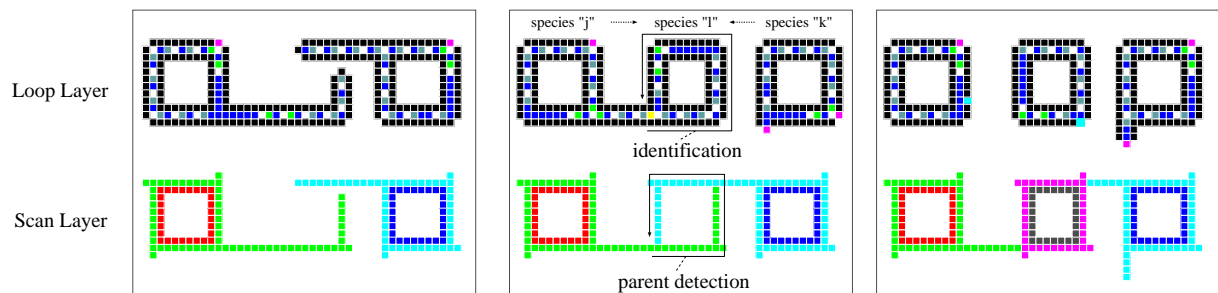


Fig. 4.11: Two-parent birth leading to new species “l” from parent species “j” (left) and “k” (right). Identification of newborn loop species and parent tracing is performed in middle frame. Footprint and trace in scan layer are represented by green and red cells for left loop, dark blue and light blue for right loop, and pink and grey for newborn loop.

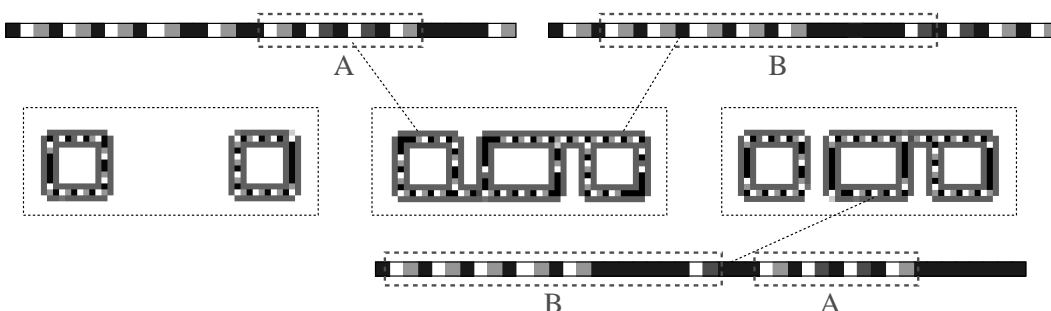


Fig. 4.12: Two-parent replication with crossover. Segments A from left loop and B from right loop are combined in the gene sequence of the child loop.

different trace identifiers are found along the path of the new loop structure. This is depicted in the second frame of Figure 4.11. By collecting all identifiers encountered along the path of the new loop, multiple parents can easily be identified, thus two-parent births are distinguished from single-parent births.

It is important to note that multiple-parent births in the evolop CA are rare events, occurring at rates of roughly 10% versus single-parent births. Moreover it should be stressed that even in the case of a multiple-parent birth, generally the gene sequence propagating into the child loop will come *exclusively from one parent*; this parent, which we call the *principle* parent, is identified with the trace identifier at the closure of the loop. In the case of the two-parent birth of Figure 4.11, for instance the structure (phenotype) is produced through the interaction of two loops, yet the gene sequence comes exclusively from one parent (in this case species “j”); thus species “j” is the principle parent of species “l” in this birth event. Actual genotype crossover events are highly infrequent due to the fact that timing for such events is crucial and rarely accommodated by the dynamics of this CA. One example of a true crossover event is depicted in Figure 4.12.

Note that a weakness of the current implementation is that, in the case of two parents of the *same species* combining to form a new child species, the current method of genealogy tracing will incorrectly record a single-parent birth. Small modifications would overcome this weakness.

4.5 Summary

New tools of analysis necessary for the results presented later in this thesis were described in this chapter. We began with a re-introduction to the evolop as a biological system rather than an artificial model, stressing the importance of sophisticated methods for analysis. The necessary components for an efficient detection and identification scheme were outlined: a definition for *birth* and *death*, and a memory of the existence and position of live replicators. An implementation of these ideas in terms of simple transition rules in the evolop CA was discussed. This included a definition of the *birth trigger configuration*, *death configuration*, and *scan layer*. Of particular importance for later results, a *labeling scheme* was introduced to describe self-replicators with necessary and sufficient detail for full state-space reconstruction. A method for genealogy tracing was also described, implementable with minimal computational overhead. Multiple-parent detection, in the case where parents were different, was shown to be detected by this method.

5. MORPHOLOGY AND DYNAMICS: A CLOSER LOOK

5.1 Introduction

The aim of Artificial Life is to abstract biological phenomena from the realm of natural ecosystems. The synthesis of life-like systems on a computer or any artificial media allows us to overcome the very limited set of initial conditions provided to us by complex biophysical systems. Yet a persistent difficulty in advancing research in this area is the lack of a formal, quantitative description for the phenomena we seek to reproduce, without which there is no concrete basis for comparison. Packard[28] states: “In general, measurable characterization of phenomena is a prerequisite to quantitative comparison, and much progress is needed in order to achieve this for many target phenomena.” It is this characterization which ultimately we aim to quantify for the evoloop CA.

In this chapter we perform an investigation on evolutionary dynamics of the evoloop CA using the tools of analysis introduced in the previous chapter. We begin by reviewing the basic structure of the evoloop gene sequence with more detail than in earlier chapters, highlighting key features overlooked in previous analysis of this model[35, 37]. We apply the concept of *passive* and *active* mutation as introduced by Ikegami[18, 19], leading us to a new classification scheme for evoloop species. We present an upper bound on the number of evoloop species for loops of a given size and use this bound as an estimate on the size of the complete state-space. We conclude with a set of small-scale experiments illustrating complex pattern formation and diverse population dynamics for three different loop species.

5.2 Passive and active mutation of the evoloop

A distinction between different types of mutation is introduced by Ikegami[18, 19] in the context of an evolutionary system consisting of machines and tapes. These types of mutation are called *passive* and *active*. In this section we will borrow these terms to describe mutation processes in the evoloop CA, though their precise meaning will differ slightly. We begin by clarifying this difference.

In Ikegami’s machine-tape model, tapes consisting of 7-bit binary sequences are read by machines which consist of a head, a tail, and a transition table, producing a next generation machine-tape pair. The term *passive* is used to refer to mutations introduced from external random noise, while *active* mutations are caused by the rewriting process intrinsic to the machine itself. In Figure 5.1, machine **A** reads tape \mathbf{x} and produces an exact copy of both machine and tape. When external noise is introduced, machine **A** reads the same tape but produces a different machine **B** and tape \mathbf{y} ; this is an instance of passive mutation. If we replace tape \mathbf{x} by a different tape \mathbf{x}' , a different offspring machine **C** and tape \mathbf{z} are produced; this is an instance of active mutation.

In contrast to Ikegami’s machine-tape system, the evoloop groups machine and tape together as one. In this case, we again refer to passive mutation as mutation which is caused by external

noise; this may include factors out of the control of the self-replicator itself, such as interaction with the environment. Active mutation refers to mutation which is *programmed* into the pair of sheath structure (machine) and gene sequence (tape) of the loop itself. In Figure 5.1, loop **A** produces an exact copy of itself. In the presence of an obstacle (a “remnant”), the self-replication process is disrupted and loop **A** reproduces loop **B**; this is a case of passive mutation. If we modify the gene sequence of loop **A** by removing a 7 gene, we are left with a different species **A'**. This loop undergoes active mutation, producing a different offspring loop **C**.

Note that there is no “random” noise in the evoloop CA as it is a purely deterministic system. As such, the definition for passive mutation given above is slightly misleading in this context. In the following sections, we will use passive mutation to refer to mutations incurred through *external* interactions (collisions) of reproducing loops. For clarity, our definition for active mutation will refer to alteration in gene sequence incurred during only the *first* replication following birth of a new loop. All other mutations (for instance interaction with offspring loops) we refer to as passive.

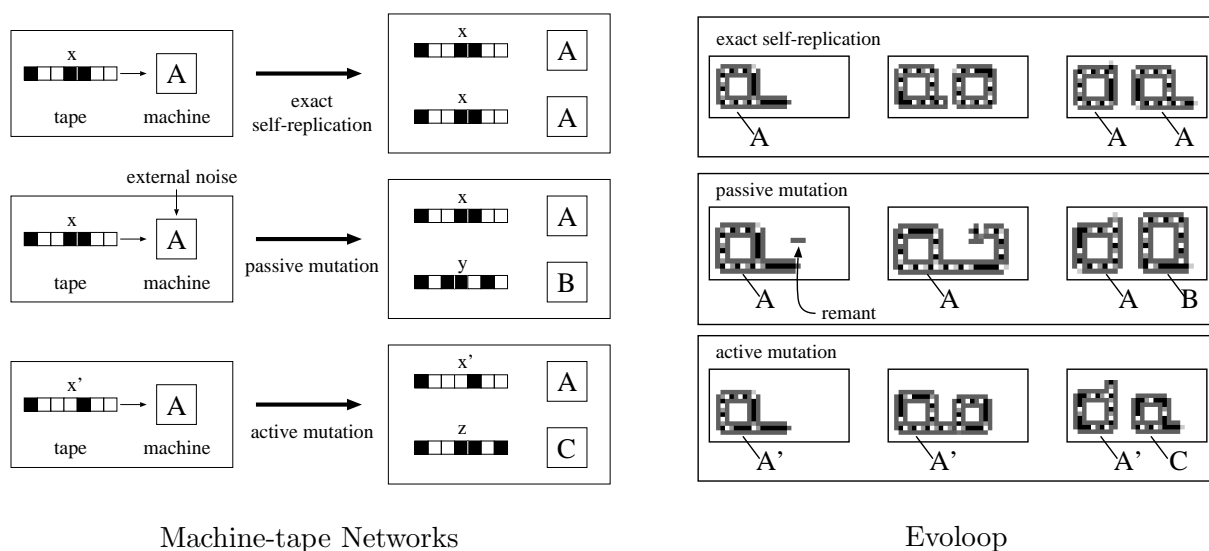


Fig. 5.1: Passive and active mutation for a machine-tape system.

5.3 A classification of loops

The tools of analysis presented in the previous chapter provide the necessary mechanisms to monitor the emergent evolutionary processes acting on populations of loops in the evoloop CA. In this chapter we introduce a framework of definitions to better describe and classify loops according to static and dynamic properties. We do so by dividing evoloop properties into two categories: *stationary* and *reproductive*. Stationary properties describe a loop on the basis of its configuration in the cellular automata space. Reproductive properties group loops of different stationary types by the existence and identity of their first-born offspring. Together, these properties provide an indication of a loop’s local intrinsic behaviour. We use these properties to categorize loops as well as to estimate the size of state-space through which these replicators evolve.

5.3.1 Stationary properties

Stationary properties define what we mean by a *species*. In previous work on the evolloop [37] this definition was limited to square-shaped loops and made reference mainly to the size of the internal sheath. Here we narrow this definition and clearly distinguish different loops based on their exact configuration at birth, subdividing this description into a *genotype* and *phenotype*. For this we make reference to the more general definitions of GTYPE and PTYPE as introduced by Langton [25], while retaining the original terminology. In this context, a loop’s genotype describes the configuration of gene sequence states — previously mostly inferred on the basis of its size — inside the outer sheath. The phenotype is defined by the size and shape (both length and width) of its inner sheath. Together, these two descriptions allow for unique and unambiguous identification of loops in the CA space. Moreover, they do so in such a way that the loop’s configuration at birth can be completely reconstructed.

Stationary classification is assigned at the moment when a loop is born. This moment of birth is defined at the instant where a state 6 (umbilical cord dissolver) appears in the space with the configuration illustrated in Figure 4.2 or any rotational variants. To avoid ambiguity, we impose the following definitions:

A **loop** is defined by a rectangular sequence of states containing 0’s, 1’s, 4’s and 7’s, with 4’s and 7’s always followed clockwise by a 0, directly enclosing an inner layer of 2’s. The sequence of states 0, 1, 4 and 7 in a loop together constitutes a **genotype**. The length and width of the inner sheath is called a **phenotype**. An outer sheath of 2’s surrounds the loop but is not included in its definition.

A **species** is defined by the combination of a **genotype** and **phenotype**. A unique **identifier** can be assigned to each species and is used to mark every state which is part of the loop as well as any adjoining sequence of genes (arm).

Appearance of a 6 state (umbilical cord dissolver) with surrounding states 22105221 (the configuration depicted in Figure 4.2) leads to the **birth** of a loop. Beginning from the 0 state, if the adjoining gene sequence and sheath satisfy the definition for a loop, the configuration of states is distinguished and marked with the appropriate species identifier.

Dissolution of an inner sheath state leads to the **death** of a loop. When death occurs, the configuration of states previously identifying the loop are de-referenced.

Upon birth, loops are assigned to a species based on their genotype and phenotype. Measurements for loop size are made along the inner sheath, with orientation defined such that length is parallel and width perpendicular to the arm of the parent loop. The genotype is determined by tracing the sequence of states 0, 1, 4, and 7 counter-clockwise along the inside of the loop, beginning at the corner of the sequence adjacent to the umbilical cord dissolver (state 6). This is the “raw” gene sequence. Sequences are labeled concisely using the method presented earlier in Section 4.3.1.

5.3.2 Reproductive properties and graph-based genealogy

Reproductive properties describe species according to their first-born offspring. Using the terminology discussed in Section 5.2, these properties describe species on the basis of whether or not they reproduce, and — if they do — whether or not they undergo *active* mutation. Hence three types of loop species are distinguished, as illustrated in Figure 5.2. A species which copies itself exactly we classify as *stable*. There are other species which do not self-reproduce at all or self-destruct during the duplication process; these species we classify as *terminal*. Finally, there are loops which evolve directly into a different species via active mutation. These species we classify as *transitional*.

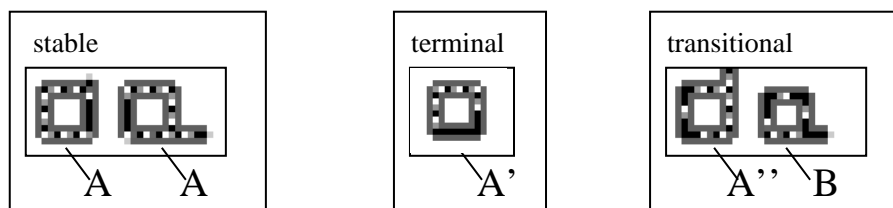


Fig. 5.2: Classification of loops by reproductive properties.

Based on this categorization, we introduce a directed graph to describe species and genealogical links, with nodes (vertices) representing individual species and links (edges) representing evolutionary paths between them. In analogy to the usual genealogy tree representations in biology, we term this structure a *genealogy graph*. An important type of genealogical link is distinguished: this is what we have termed a *free-space* link. Free-space links describe the next step in a genealogy graph which a species will follow in the absence of any external environmental interaction. Every species either has just one such free-space link, or does not have it at all, which is determined solely by its stationary properties (genotype and phenotype) independent of external factors. The traversal of links other than this one (passive mutation) represent evolutionary jumps induced by interaction with self-generated or external environments (other loops, static obstacles or stationary self-modification). A free-space link to another species corresponds to active mutation, and a free-space self-link corresponds to exact self-replication. The classification introduced in Figure 5.2 can easily be thought of in terms of free-space links in a graph-space representation; this is illustrated in Figure 5.3.

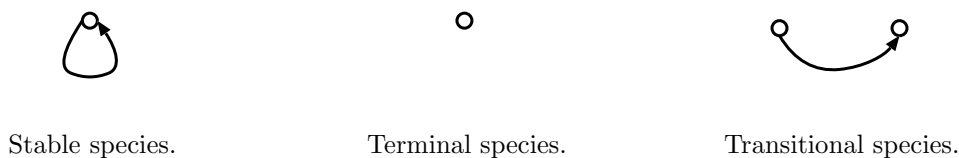


Fig. 5.3: Classification of loops in graph-space.

The fact that evoloop species exhibit strongly graph-based genealogy has been observed experimentally in a variety of experiments; some of the strongest evidence to this effect is demonstrated in results of Section 6.3. Given this fact, species are found to fit naturally into the classification scheme described above. In practice, stable species are most commonly encountered and have been

observed to be exclusively square-shaped. Terminal species are either incapable of self-reproduction (no 4's in their gene sequence) or are self-destructive. Transitional species can be either square or rectangular. Square transitional loops are commonly found to either shrink or grow, depending on the number of 7's in their gene sequence; in the absence of external influences, these form *chains* in graph-space. Rectangular transitional loops normally evolve directly into square ones, which themselves are either stable or transitional.

Using the genealogy graph representation, we can compile time-dependent statistics on transition frequencies between nodes (loop species) in the CA space. Deviation from the exclusive traversal of free-space links is an indication of the relative level of mutability of a species due to interaction with the environment. The frequency that species traverse these “non-free-space” links depends both on stationary properties and on the nature of their environment. For stable species, the existence and traversal frequency of these links plays a critical role in the evolution of populations. A stable species with many high-frequency links to other stable species, for instance, will encounter great difficulty in maintaining survival due to the high likelihood of invasion. On the other hand, a species with low fitness but many incoming links will frequently recur as long as its parent species are present in sufficiently large numbers.

The idea of graph-based genealogy bears close resemblance to the quasispecies model of Eigen, McCaskill and Schuster[11]. Under certain conditions, this model is found to characterize a chemical system as a single, *master sequence* and a nonsymmetric cloud of mutants. Such a characterization is valid on the condition that “significant cross-catalysis of neighboring sequence polymerization occurs, i.e., if mutations happen much more frequently than it is commonly assumed in population genetics”[11]. In this context, selection favours a dominant *ensemble* — consisting of this master sequence with its frequent mutants — rather than a single species. In relation to the points discussed above, there are close similarities to the graph-based genealogy of evoloop populations. In particular, statements relating fitness and graph-based connectivity are relevant and will be discussed in more detail in Chapter 6.

5.4 The evoloop state space

A striking consequence of the description scheme of Section 4.3.1 is that, contrary to earlier analysis [35, 37], heredity in the evoloop system spans a potentially huge state-space. To determine a bound on this size, we narrow our focus and look at stable species as defined in the previous section. As will be demonstrated in the coming sections, these are loops which dominate evoloop populations. We wish to arrive at an estimate on the size of state-space spanned by this type of loop, in order to better understand the evolution through genealogy graph space.

Figure 5.4 highlights relevant details of a stable size-6 evoloop. A stable loop of size n has $4(n + 1)$ states in its gene sequence, with $(n+1)$ along each limb (side). Each 7 gene in this sequence extends a limb by one cell, and the two 4 gene combine to extend it one more cell. Hence, for the child loop to be the same size as the parent, the parent loop must have the same number of G (growth) triplets ($G = '071'$) as its size (n). This adds up to a total length of $3n$ cells in the sequence (3 cells per triplet). Of the remaining $4(n + 1) - 3n = n + 4$ states, 6 are allocated to the two T (turning) triplets ($T = '041'$). Hence there are a total of $(n - 2)$ free C (core) states ($C = '1'$). It is the placement of these free Cs in the sequence which determine the species to which a loop belongs.

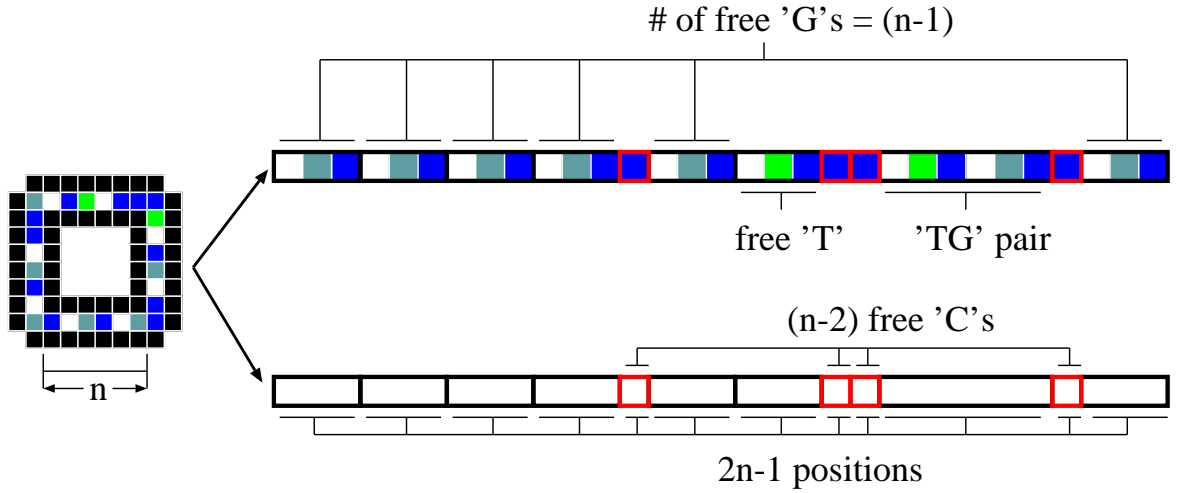


Fig. 5.4: Placement of genes in size-6 stable loop.

Additional dynamical constraints impose conditions for successful reproduction. The first of these is that the T triplets must have no intervening G triplets between them. The second is that the trailing T triplet must be immediately followed by a G triplet. Both of these conditions are a result of the exact mechanics of the state-transition rules for the evoloop, which for brevity we will not describe in detail here.

We would like to calculate the number of possible arrangements of free 1 states between gene triplets. We can freely shift the $(n-2)$ C states in the sequence, which additionally consists of $(n-1)$ free G triplets, one free T triplet, and one TG coupled triplet. Hence there are $(n-2) + (n-1) + 1 + 1 = (2n-1)$ positions available (see Figure 5.4). Due to periodicity, first and last positions are the same, hence there are strictly $(2n-1) - 1 = (2n-2)$ possible “slots” for Cs. This results in a total of $\binom{2n-2}{n-2}$ possible arrangements. If we tabulate these values for different sizes of loop, we find:

Loop Size	# of Species	Loop Size	# of Species	Loop Size	# of Species
4	15	9	11,440	14	9,657,700
5	56	10	43,758	15	37,442,160
6	210	11	167,960	16	145,422,675
7	792	12	646,646	17	565,722,720
8	3,003	13	2,496,144	18	2,203,961,430

Note that these numbers strictly represent only an upper bound on the number of species; it is possible that additional emergent properties of the state transition rules may disrupt the self-reproduction process due to gene placement. One such exception has been noted for a subclass of species which, following successful self-replication, self-destruct. An example of one such loop of size 5 is shown in Figure 5.5. According to the definition given in Section 5.3.2, this loop still satisfies the requirement to be a stable species, yet it is not very successful as a self-replicator.

Despite the exception of Figure 5.5, our calculation of the size of the evoloop state-space represents an accurate estimate as verified by experiment. What is surprising about these results is the



Fig. 5.5: A size-5 stable loop which only replicates once before self-destruction.

sheer size of this space. Previous analysis of the evolloop [35, 37] used a classification scheme which did not include spacing of C's in the sequence. Rather, relative placement of the TT pair between G genes was used to distinguish different species. This classification was thus limited to a total of s different species for a loop of size s . In contrast, the classification scheme we have introduced shows far greater potential diversity. A size-13 loop, for example, may be a member of one of over 2 million stable species of this size versus only 13 different species in the earlier classification. Indeed, there is a combinatorial explosion in the number of possible genotypes as the loop size grows, allowing an increasingly vast number of different gene permutations.

In the next section we show that this state-space of evolloop replicators is not only *huge*, but also *diverse*. Given different gene sequence permutations, we demonstrate widely varying behavioural characteristics: different levels of mutability, genealogical connectivity, and self-generated environments.

5.5 Behavioural study of different loops

The focus of this section is a case study of three size-6 loops: the 5f541/6×6 (GCCCCGGGTTGG), 13e95/6×6 (GGGCGTTGCGCC) and 41f55/6×6 (GGGGTTGCCCG). According to earlier analysis methods based on size[37] these loops would be classified as the same species. The identification scheme described in Section 4.3.1 however distinguishes them by the spacing and sequence of genes in their gene sequence; hence we now can observe *exclusively genotype-based* characteristics of different types of loop. The differences we observe makes clear the importance of the framework for analysis which we have introduced.

5.5.1 Pattern formation and morphogenesis

Figure 5.6 shows the time-evolution of these three species over 4000 iterations as they replicate into open space. Pattern formation is strikingly different for each loop due to self-generated environments and space occupation. These self-generated environments have profound effects on evolutionary dynamics, yet their emergence in the space are *not predictable from local analysis of behaviours alone*. It is only through phenotypic interaction that the necessary conditions for their emergence are generated.

The first case of Figure 5.6 depicts the evolution of species 5f541/6×6, a stable species with gene sequence GCCCCGGGTTGG. Interactions between loops in the evolution of this species lead directly to the appearance of the dissolver state, which in turn clears away parent loops. Hence loops behind the advancing front are destroyed, resulting in a moving line of loops of fixed-thickness; this is shown in the dynamics of Figure 5.7.

It is interesting to note the conditions in which this species was discovered. An experiment was performed in [2] with a dynamic environment consisting of a “persistent dissolver”, another kind

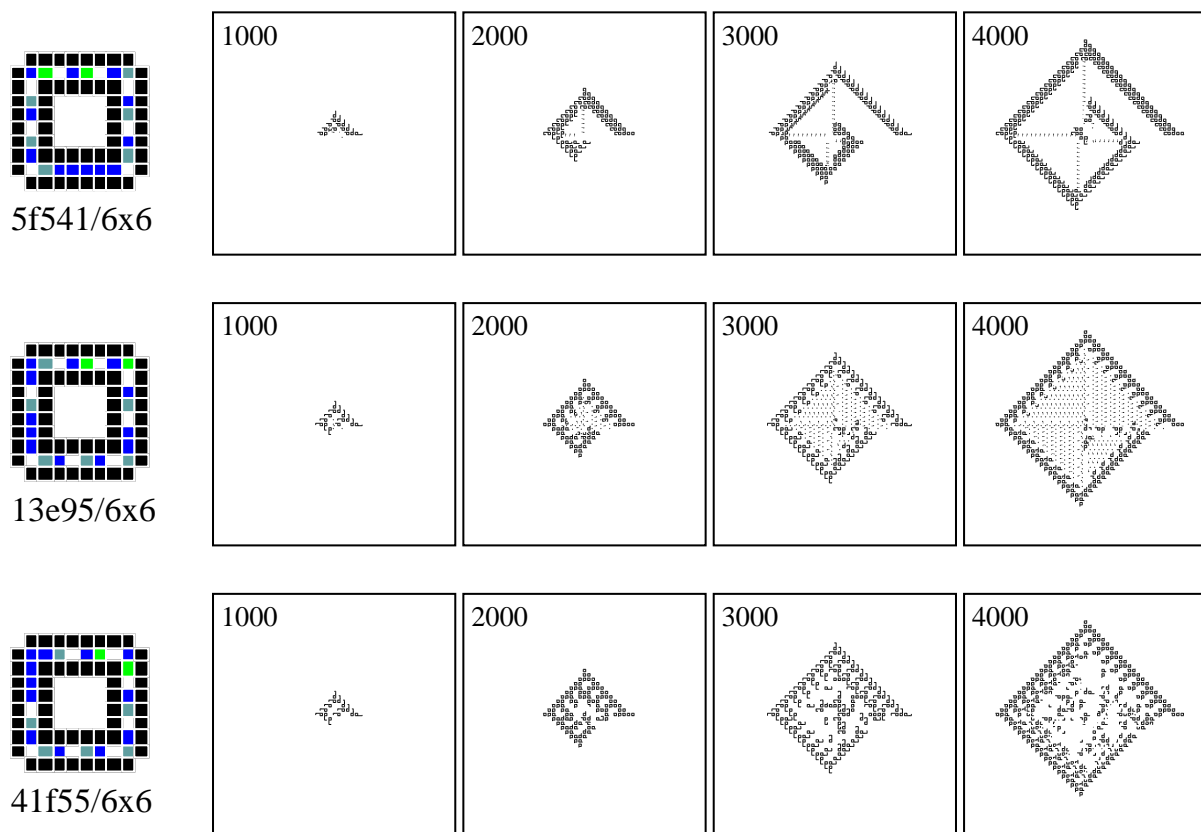


Fig. 5.6: Pattern formation from different loops.

of dissolver state with extended lifetime. The basic behavior of the persistent dissolver is a wave-like propagation over susceptible populations, similar to those reported in spatial predator-prey or host-pathogen systems. The effect of this new environment is twofold: to open free space and to dynamically partition the periodic domain[2]. In this experiment, a catastrophic initial surge of this dissolver state destroyed all but *one* species, this being the 5f541/6 \times 6. The fact that this species leaves *no trail* of live loops means that it may be better able to escape the spread of a pathogen than other species such as those depicted in Figure 5.6. As we will show in subsequent sections, however, survival of this species over the longer term is made infeasible by its high mutation rate.

The second species of Figure 5.6 demonstrates a clear example of a self-generated environment through the creation of what we call *remnants*. A remnant is the term we use to describe a contiguous structure composed of only sheath and core states, disconnected from any living loop and without the gene 4 state necessary for self-replication. These structures have neither the capacity for self-reproduction nor a configuration leading to self-destruction¹.

The creation of remnants can produce a drastic and immediately perceivable effect on the evolution of loop populations, as demonstrated in the second row of Figure 5.6. Figure 5.8 shows an example of this process, where the dissolving state — through a complicated series of transitions

¹ By this loose definition non-reproducing loops that remain sterile in the space are also a kind of remnant.

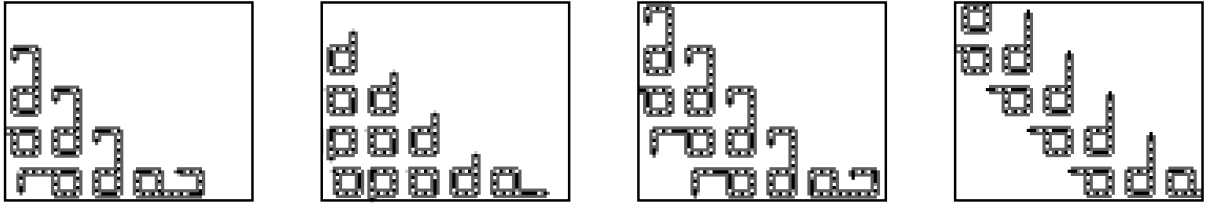


Fig. 5.7: Moving front of 5f541/6×6 loops.

— does not manage to completely extinguish the center (parent) loop. Each subsequent child behind the loop front undergoes an identical or similar set of transitions, recursively generating a trail of remnants. This strongly resembles the behavior of the SR Loop, which leaves a region of “corpses” behind the moving front of self-replicating loops (see Fig. 2.4). Langton in fact discusses the subject of *recursively generated objects* at length. Using his definitions for GTYPE and PTYPE (generalizations of the terms “genotype” and “phenotype”), the loops and remnants in the middle row of Figure 5.6 would in fact be grouped together as one. As such they confirm Langton’s statement that the extended phenotype “will emerge under the action of the (genotype), developing through time via a process akin to morphogenesis” [25].

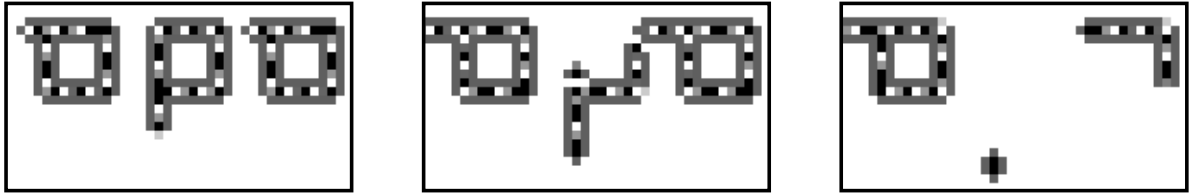


Fig. 5.8: Remnant generated from interaction of 13e95/6×6 loops.

5.5.2 Population dynamics and genealogy

Having explored pattern formation and phenotypical diversity of different loops, we now analyze population dynamics and genealogy exploration in a periodic domain in order to arrive at a qualitative understanding of the relative fitness of these very different loops. For these observations we introduce a single ancestor species into a periodic grid of size 500×500. In Figure 5.9 we focus our attention on a subset of all species in the space, plotting counts only for species whose population exceeds 25 at any time during the first 200K iterations.

Evolution of each of these species demonstrates quite distinct dynamics. In the case of the 5f541/6×6, we observe rapid convergence to two minimal-sized attractor species, the closely related 11f5/4×4 (GGTTGCCG) and the 9f5/4×4 (GGTTGCGC). Along the path to reach these attractor species, the following genotypes occur with populations of 25 or more:

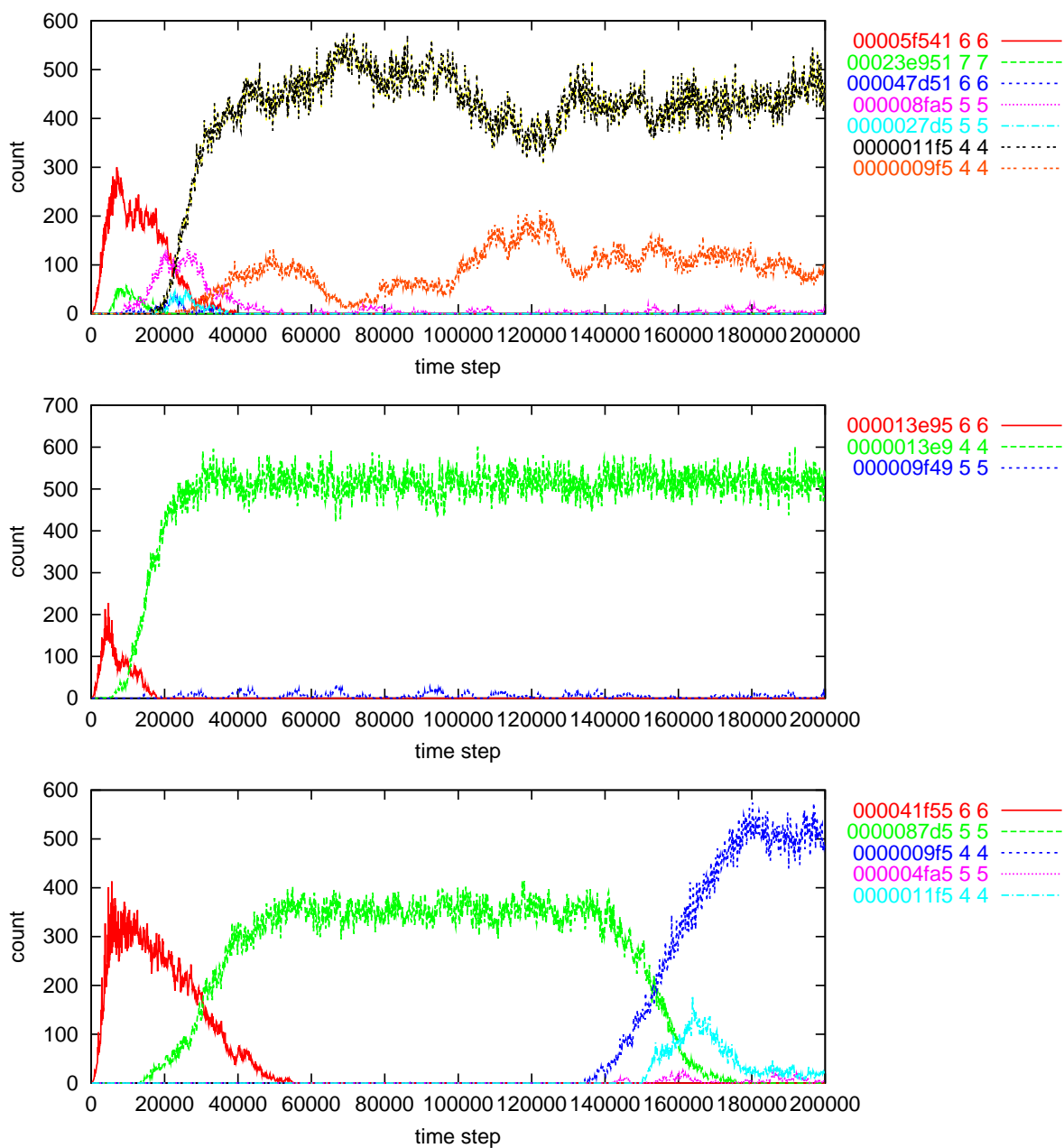


Fig. 5.9: Species with counts which exceed 25 individuals on a 500×500 grid. Top: 5f541/6 \times 6 (GCCCCGGTTGG), middle: 13e95/6 \times 6 (GGGCGTTGCGCC), bottom: 41f55/6 \times 6 (GGGGTTGCCCG).

Hex label	Gene sequence
5f541 6×6	GCCCCGGGTTGG
23e951 7×7	GCCGGGGCGTTGCCG
47d51 6×6	GCCGGGTTGCCG
8fa5 5×5	GGCGTTGCCG
27d5 5×5	GGGTTGCCGC
11f5 4×4	GGTTGCCG
9f5 4×4	GGTTGCC

Note that the cyclic permutation which is shown everywhere in this thesis describes the way genes are arranged *at birth*, which for stable species is always the same in both parent and child; hence if any of these sequences is cyclically permuted, the child species will always match the form given above. For clarity we have shifted all sequences so that TT pairs are aligned. It is interesting to note that the progression of sequences for this run demonstrate evolutionary favoritism towards species whose TT pair are shifted to the left (tail) of this sequence. The question of how gene positioning affects the evolutionary dynamics of different species is not yet well understood in this model. A more thorough investigation of conserved subsequences (e.g. the TT pair), described in the next chapter and also in [31], takes a first step in this direction.

The appearance of a larger species (23e951/7×7) in significant numbers is somewhat surprising given the assumption of size-based fitness. Using the tools introduced in Chapter 4, we can trace the path which evolution follows to reach this species. The path which accounts for the first appearance of this species² evolves through gene sequences corresponding to three transitional species: the 15057d/6×8, 2be95/7×6, and fa545/7×7. This path is illustrated in Figure 5.10 in terms of four single-parent replication events (mutations). We observe that the first and third transitions involve the injection of extra C states into the gene sequence, always coming at the end of the sequence as the loop is closed. C-injection is the most commonly observed way in which evolops modify gene sequences during replication events, two-parent crossover being much more rare. In the second transition, a consecutive sequence of genes from the parent is passed on to the larger, child species; two G-genes are lost in this process. The fourth event is a simple permutation in which gene sequence length is preserved. The variation produced through these four steps and three transitional species demonstrates the potential of evolop species to explore the large genealogical state-space described in Section 5.4. Reduced fitness (i.e. slower reproductive rate), however, prevents this new and more complex species from surviving in competition with smaller ones.

The analysis method of Figure 5.10 is insightful but has the inherent limitation of being unable to capture the dynamics of genealogical connectivity evolving collectively through an entire population. To express this global genealogy, we use a visualization method introduced by the author and his collaborators in earlier work. Using this approach, species are plotted in a 2D space whose x -axis represents a hash value for the gene sequence and whose y -axis represents its size. Genealogical links between species over a specified time window \mathbf{T} are plotted as lines between these points, with line thickness governed by traversal frequency over the link. The prominence of species in the time window is quantified using a quantity we call the *production*[32, 33] $\text{Prod}(\mathbf{T})$. If we denote the number of incoming, outgoing and buckle link traversals to a given node in the time window \mathbf{T} as

² Keep in mind that graph-based genealogy produces alternative pathways as well, see Figure 5.11.

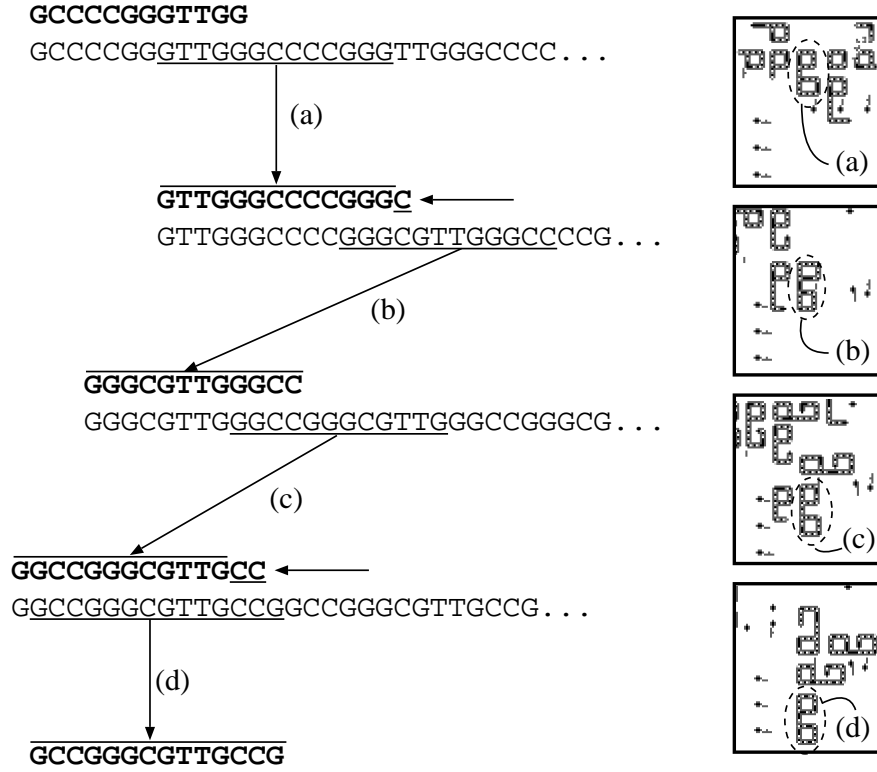


Fig. 5.10: Transitions from species 5f541/6×6 to species 23e951/7×7 via three transitional species: 15056d/6×8, 2be95/7×6 and fa545/7×7. Arrows to underlined Cs represent injected core states.

$I(\mathbf{T})$, $O(\mathbf{T})$, and $B(\mathbf{T})$, respectively, the production is derived as:

$$\text{Prod}(\mathbf{T}) = O(\mathbf{T}) + B(\mathbf{T}) - I(\mathbf{T}) \quad (5.1)$$

According to this definition, replicator species which frequently construct other species as well as replicators of their own species will have a high production; those which are frequently constructed by other species but fail to self-replicate will have negative production. The balance in equation (5.1) is hence between the capability of a species to *produce* versus the tendency to *be produced*. Ranking nodes in this way quantifies the role species play in the evolutionary process. For more details on this approach to genealogy visualization we refer to [32, 33].

Graph-based visualization for the 5f541/6×6 over the first 25K iterations is shown in Figure 5.11. Species are plotted according to their production: blue circles for positive values and red triangles for negative values; sizes are scaled according to the magnitude $|\text{Prod}|$. What is immediately evident from this plot is the complexity of genealogical connectivity emerging from dynamics in the CA space. During this initial period of instability the system explores a wide range of species in graph-space, the most prominent of which are indicated in Figure 5.11. A number of links connect the larger 23e951/7×7, to the 5f541/6×6, including *but not limited to* the path first traversed as described earlier in Figure 5.10. Our ultimate aim of open-ended evolution would imply continuous exploration and adaptation in this genealogy space. In this case, however, appearance of

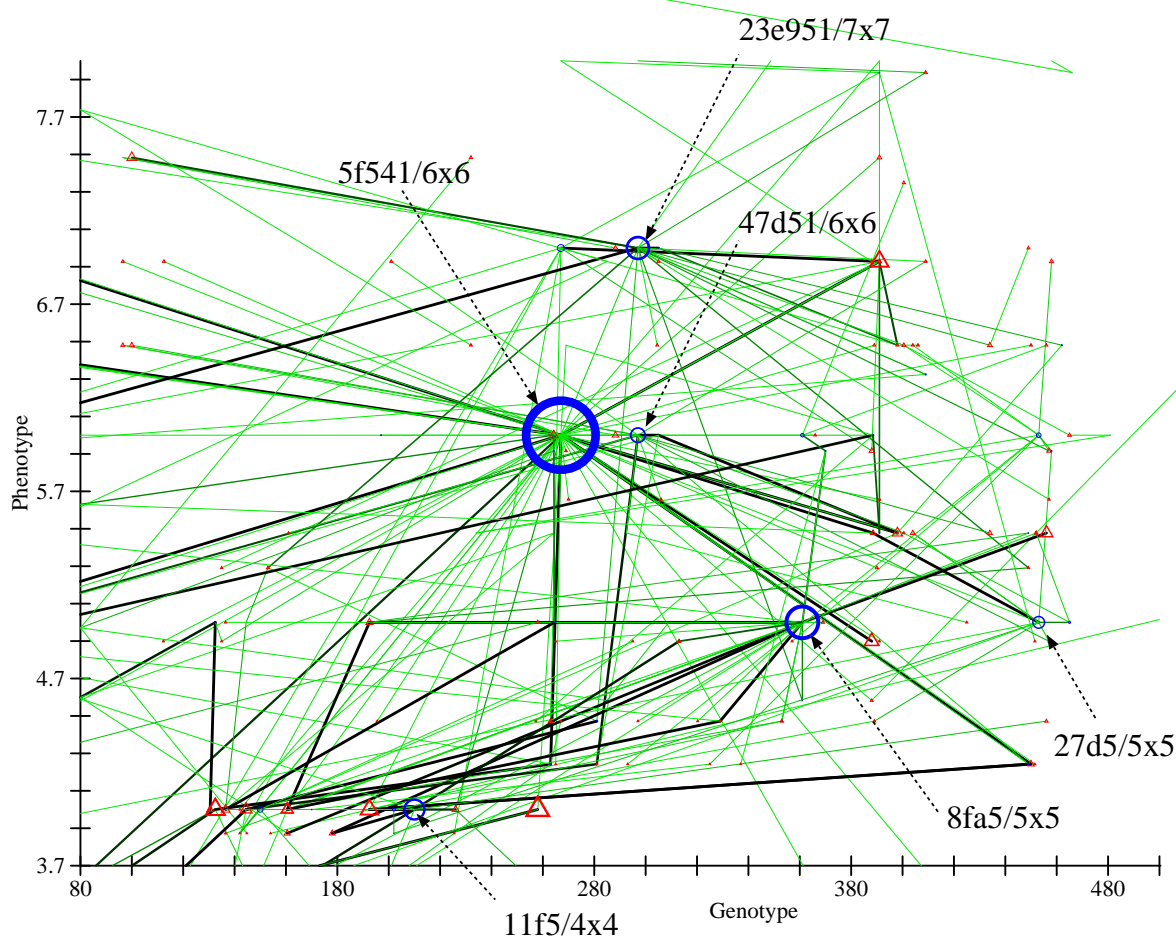


Fig. 5.11: Graph-based genealogy of the 5f541/6×6 over the time window (0, 25K). Blue circles represent positive production, red triangles represent negative production. Line thickness of genealogical links is proportional to transition frequency.

smaller species, most notably the 11f5/4×4, signals that evolution is optimizing to a small subset of dominant, high-fitness replicators at the exclusion of larger, more complex species such as the 23e951/7×7.

Following the first 25K iterations, the system rapidly collapses into the semi-stable state of two minimal-sized attractor species. A period of competition then begins during which time both species survive in large numbers. This co-existence can be understood from fitness arguments: since both loops have maximized their (size-based) fitness, neither has any evolutionary advantage, hence they compete on roughly equal terms. In the case of these two minimal species, evolutionary events are extremely rare due to the low frequency of evolutionary links from small species to other stable species. We may wonder what would happen if much larger species of equal fitness — species with much higher graph-based connectivity — were allowed to compete over extended time periods. Experiments addressing this topic will be discussed in Section 6.3 of the next chapter.

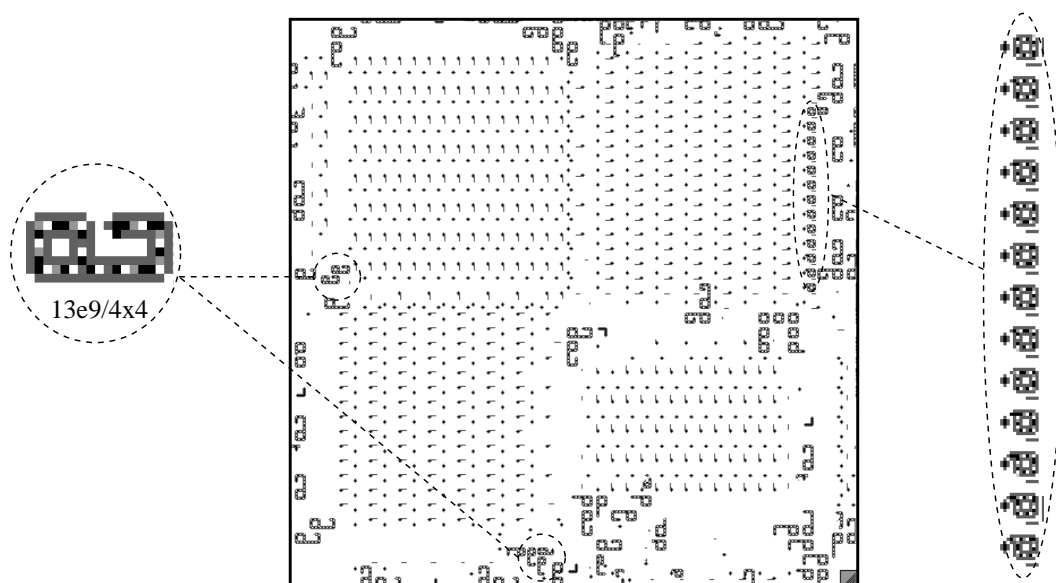


Fig. 5.12: Emergence of $13e9/4 \times 4$ after 7353 iterations. Left: clusters of $13e9/4 \times 4$. Right: recursively generated vertical front of $13e9/4 \times 4$ loops, each blocked by remnants.

The second species under investigation demonstrates quite different population dynamics, as illustrated in the middle frame of Figure 5.9. In contrast to the dynamics of the $5f541/6 \times 6$, diversity in this run is limited to only three species (with counts exceeding 25 individuals):

Hex label	Gene sequence
13e95 6×6	GGGCGTTGCGCC
13e9 4×4	GCGTTGCG
9f49 5×5	GCGCGTTGCG

In this case it is not surprising that evolution finds the minimal size-4 species very fast: the genotype of the $13e9/4 \times 4$, GCGTTGCG, is an exact subsequence of that of the $13e95/6 \times 6$, GGGCGTTGCGCC. Out of a total of 161 species encountered on this run, the $13e9/4 \times 4$ is only the fourth to appear, encountered for the first time 2458 steps into the run. The creation of remnants plays an important role in this event, partitioning the CA space and restricting loops to densely-packed clusters. A snapshot of the CA space after 7353 time steps is shown in Figure 5.12, demonstrating the spread of remnants over the entire domain. We notice at this time two separated clusters of smaller $13e9/4 \times 4$ loops emerging in the space, each having evolved through different mutation events. As evidenced in the second frame of Figure 5.9, these smaller loops rapidly rise in number to out-compete their larger ancestor. Also apparent on the right side of the grid is a vertical front of 13 instances of species $13e9/4 \times 4$, each blocked from self-replication by remnants. What is interesting about this structure is its periodic dynamics: the rotating gene sequence in each loop is out of sync from the one below it by exactly one time step. Although this structure is gradually cleared away over subsequent generations, its emergence demonstrates that self-generated environment in this CA often have dynamical properties, potentially of the kind required for complex information

processing to occur. However it has yet to be shown in a conclusive way the role, if any, these structures play in evolutionary processes of evolloop populations.

The bottom frame of Figure 5.9 follows population dynamics of the 41f55/6×6, the third species under investigation. Five species are encountered in this run with populations which exceed 24 individuals:

Hex label	Gene sequence
41f55 6×6	GGGGTTG○○○○G
87d5 5×5	GGGTTG○○○○G
9f5 4×4	GGTTG○○○
4fa5 5×5	GG○○TTG○○○
11f5 4×4	GGTTG○○○

Behaviour exhibited in this case demonstrates a clear example of weak mutability: for roughly 100K iterations, a size-5 descendent of the 41f55/6×6, the 87d5/5×5, remains the dominant species despite being of non-minimal size. The dominant status of this species is maintained over smaller, fitter species thanks to its limited graph-space connectivity. To visualize this phenomenon we use the same approach as employed for Figure 5.11, this time following the complete evolution over 200K iterations in windows of 25K iterations each. The result of this procedure is shown in Figure 5.13. The first two frames of this sequence show, among other things, a prominent link to a larger, size-7 species (20fa55/7×7). In other experiments in which the evolloop system was coupled with a dynamic environment on a large-scale domain (3000×3000)[2] it was demonstrated that there exists a collection of graph-space links forming a cyclical path in graph-space connecting these two species. The result of this cyclic connectivity is that the pair of species are genealogically *coupled* — growth of one population implies growth of the other, and vice versa. Yet Figure 5.13 shows nothing of these dynamics; the emergence of the smaller, fitter 87d5/5×5 puts to an end the possibility of more extensive graph-space exploration before it has even started.

To test whether the weak mutability of the 87d5/5×5 observed in Figure 5.13, is in any way connected to the particular circumstances of the run (i.e. grid size), two other experiments were performed on grids of 499×499 and 501×501. Surprisingly, in both cases the weak mutability is lost and evolution proceeds directly to the smallest size-4 species. Hence we observe an important point: that the periodic boundary conditions we impose play a critical role in deciding the path of evolution. This is true also of the vertical configuration of loops in Figure 5.12; this unusual formation is made possible only by the collision of two advancing loop fronts through the periodic boundaries.

The fact that periodic conditions play such an important role in evolutionary dynamics poses a considerable challenge for deriving general principles about dynamics. One possible solution to this problem is to analyze extensively the dynamics of different species on very large, nonperiodic grids — as was done in Section 5.5.1 — and characterize these species by the evolutionary events which occur *before* boundaries are encountered (i.e. simulate infinite boundaries). This type of analysis is one of our future goals, and will be further discussed in Chapter 7.

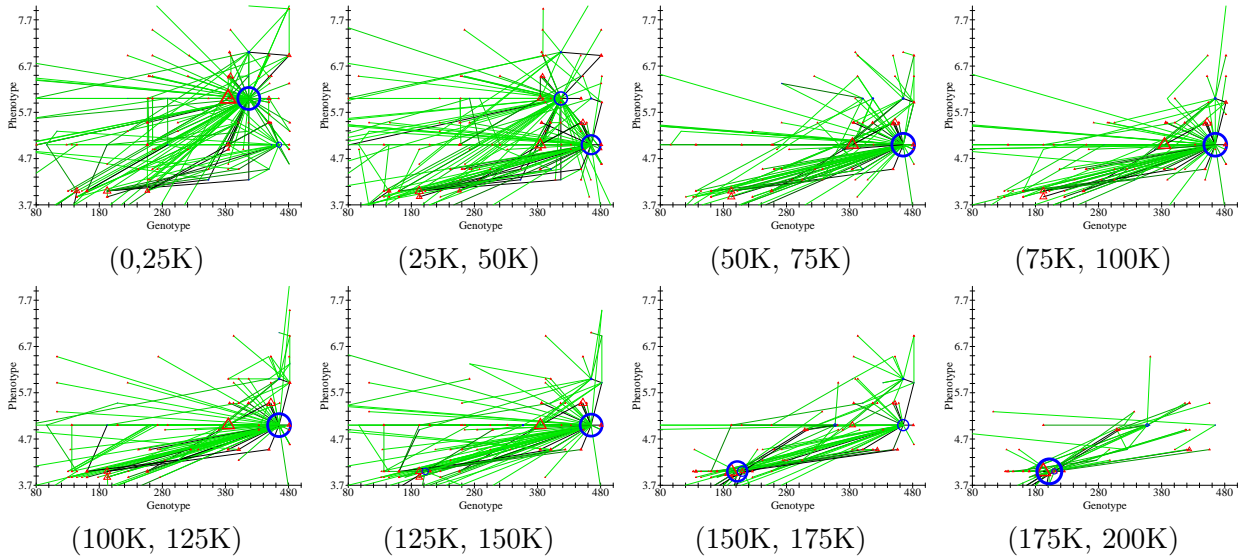


Fig. 5.13: Evolution in genealogy graph space of species 41f55/6 \times 6.

5.6 Summary

We began this chapter with a discussion of two different types of mutation, *passive* and *active*, with reference to discussions by Ikegami[18, 19]. These ideas were used to introduce a classification for evoloop species based on properties which we defined as either *stationary* or *reproductive*. Stationary properties were related to the identification of loops by genotype and phenotype as introduced earlier in Chapter 4. Reproductive properties were related to links in genealogy graph-space. From the idea of a *free-space* link, we described loop species as either *stable*, *transitional* or *terminal*. The state-space of stable species — dominant among these classes of loop — was shown to grow combinatorially with loop size, demonstrating a far more expansive genealogy state-space than predicted by earlier analyses[35, 37].

The ideas introduced in the first half of this chapter were then applied to the analysis of three different size-6 loops. Pattern formation on an infinite domain was used to demonstrate striking and *exclusively genotype-based* differences between these species. In particular, self-generated environments originating from subtleties of gene arrangement and spacing were shown to play a crucial role in evolutionary processes. A closer look at the transitions driving a single size-increasing mutation event revealed a complex series of gene re-arrangements, each of which resulted from simple collisions and arm-bonding. To visualize the collective genealogical network of the entire population, an approach introduced in [32, 33] was used. Applied to one of the trial species (5f541/6 \times 6), visualization revealed complex graph-space exploration traversing a number of different stable species. From this result it was concluded that evolutionary processes emerging in the evoloop CA are capable of finding highly diverse gene arrangements, potentially generating increases in complexity. Yet with this and other examples it was found that an overwhelming size-based fitness differential prevented long-term sustainability of these dynamics.

6. A NEW KIND OF LOOP

6.1 Introduction

Results from small-scale experiments presented in the previous chapter and from earlier analyses [35, 37] indicate a general tendency of evolloop populations to evolve towards the smallest possible loop size. Thanks to the length of their gene sequence, these minimal loops accomplish self-replication in the shortest time, allowing them to multiply rapidly and out-compete other, larger species. As Sayama concludes[37]: “smaller individuals were naturally selected thanks to their quicker self-reproductive ability, and the whole population gradually evolved towards the smallest ones.” This discovery marked a significant accomplishment: a clearly emergent selection mechanism was generated in a simple 2D CA, one bearing resemblance to real biological systems. Adami[1], for instance, remarks that: “In the absence of a mechanism that enables one string to prevent the replication of another (kill-strategy), it is conceivable that “fitness=replication-rate” is a universal feature of such low-level organisms”. The evolloop confirmed the universality of this feature.

Yet in this chapter we demonstrate results which contradict the generality of these conclusions. A larger class of evolloop species is discovered which — given an appropriate initial gene sequence configuration — maintain an arbitrary minimum size. These species are not completely new: a subset of this class was earlier discovered in [37]. As these replicators do not evolve in the expected manner (i.e. towards smaller sizes) they were dismissed and their evolutionary properties largely overlooked. In this chapter we show that species of this class evolve according to a new, emergent selection criteria, one whose properties are not yet fully understood. Based on preliminary analysis of small and medium-scale experiments we observe that these new species exhibit strong graph-based genealogy and long-term evolutionary trends. The quasispecies model of Eigen, McCaskill and Schuster [11] is introduced to explain their collective evolution as groups of species with closely linked ancestry.

6.2 Conservation properties of the evolloop

Whether the quantity be energy, momentum, or mass, conservation laws play a key role in shaping the equations of physical theories. By enforcing a strict bound on system dynamics, these laws limit state-space exploration and minimize degrees of freedom. For a sufficiently well-described system, conservation principles have the potential to reduce a seemingly complex, intractable system to a simple, computable trajectory.

In the work presented so far, however, the evolloop CA has demonstrated no obvious conserved properties. Conservation of material is clearly violated; replicators may be created and destroyed arbitrarily given sufficient space. Neither are there any clear analogs for momentum and energy conservation. If a conservation law exists for this system, it is not linked in any clear way to

traditionally fundamental physical quantities.

Yet what about the evoloop gene sequence? As an emergent property of the system, the evolution of this sequence does not at first sight to obey any simple conservational principle. If it were to do so, one would expect certain types of subsequences in the genotype to be unchanged through evolution. Experiments performed in [37], however, hold out the possibility that such subsequences exist. In these experiments, a type of loop called the *n-evoloop* was found only to evolve towards larger loops and not to smaller ones, as depicted in Figure 6.1. Sayama states:

In these results it is indicated that the evoloop actually has some evolvability. For example, in some cases the loop evolved to that of a larger species, and in other cases it generated some variants that lost their self-reproductive ability but became capable of reproducing smaller offsprings than themselves. However, self-reproductive smaller species could not emerge yet in these preliminary experiments.

Translated to the terminology introduced in Section 5.3.2, graph-space links from this species do not connect it to any smaller, *stable* species. Hence this species is trapped in a subset of genealogical graph-space, connected only to smaller species which are either transitional or terminal. Given that these smaller species cannot self-reproduce exactly, larger, stable species are the only real evolutionary competitors. The fact that evolution to smaller species has been essentially blocked points to an important property of the gene sequence, one which conserves a minimum length given a particular initial configuration.

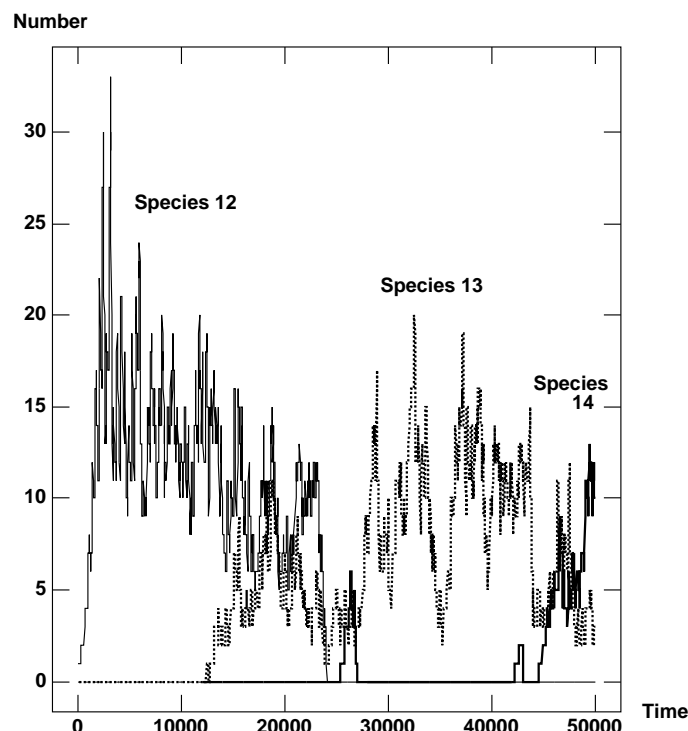


Fig. 6.1: Evolution of the n-evoloop to larger size species on 201×201 grid (courtesy H. Sayama)

To explain the results described by Figure 6.1, we need to address two points:

1. What is preventing loops from evolving to smaller sizes?
2. What evolutionary advantage do the larger loops have, such that they win out over smaller ones?

In this section we present tentative experimental evidence indicating a relationship between conserved subsequences in the evolloop gene sequence and a blocking-mechanism against evolution to small-sized loops. The second question is addressed in next section of this chapter.

We begin by restating the observations made by Sayama in [37] in terms of the gene sequence representation described in Chapter 4. An “n-evolloop” describes an evolloop with a gene sequence in which all free C states are between the second T triplet and the first G triplet. For each loop size, one such species exists, such as the size-12 “n-evolloop” depicted in Figure 6.2. The gene sequence for this species is written as `GGGGGGGGGGCCCCCCCCCTTGG`.

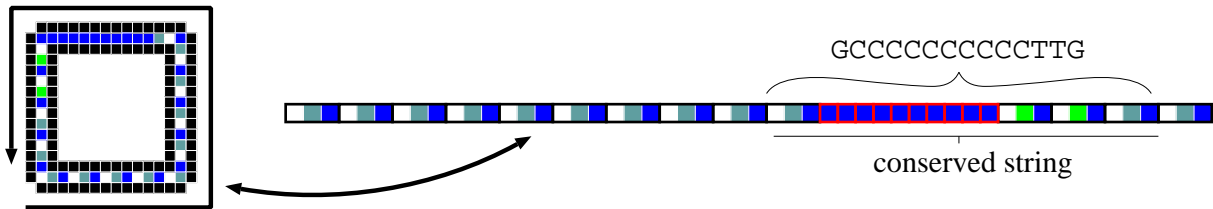


Fig. 6.2: Gene sequence of a size-12 “n-evolloop” (species 17c0055555/12×12).

Given the tools of analysis presented in Chapter 4, we can now easily generate and analyze the collection of gene sequences produced during the evolution of this species. The significant difference of this new analysis with the results of Figure 6.1 is that we can now easily identify all species exactly, analyze their gene sequences, and trace their genealogy. By doing so, we can search for any conserved subsequences which would explain the behaviour of Figure 6.1.

As a starting point, we reproduce the same experiment as performed in [37] which generated the results of Figure 6.1. This run is performed on a grid of size 201×201 beginning with a single size-12 n-evolloop (species 17c0055555/12×12). Rather than plot species by size, we plot species using the new, exact identification scheme, restricting ourselves to populations of stable species which exceed 5 individuals. While this run was earlier ended at 50K iterations, we run the simulation until 112K iterations, by which time a short-period limit cycle has been reached and all stable species are extinct. A plot of these results is given in Figure 6.1. In the first 50K iterations we observe the same dynamics as in Figure 6.1, yet a number of finer details become apparent. Notably, we see that the first appearance of a size-14 loop just after 24K iterations is not the same species as in the case of the later appearance of a size-14 loop at 42K iterations. As we run the simulation forward in time, we notice that the evolution to larger loops continues. At about 57K iterations, species 27004955555/15×15 emerges but fails to take control of the space. This species re-emerges again at 75K iterations, and again fails to survive. Just before 110K iterations, the last surviving member of species 4f800955555/14×14 perishes, and the system eventually hits a small-period limit cycle (only transitional and terminal species).

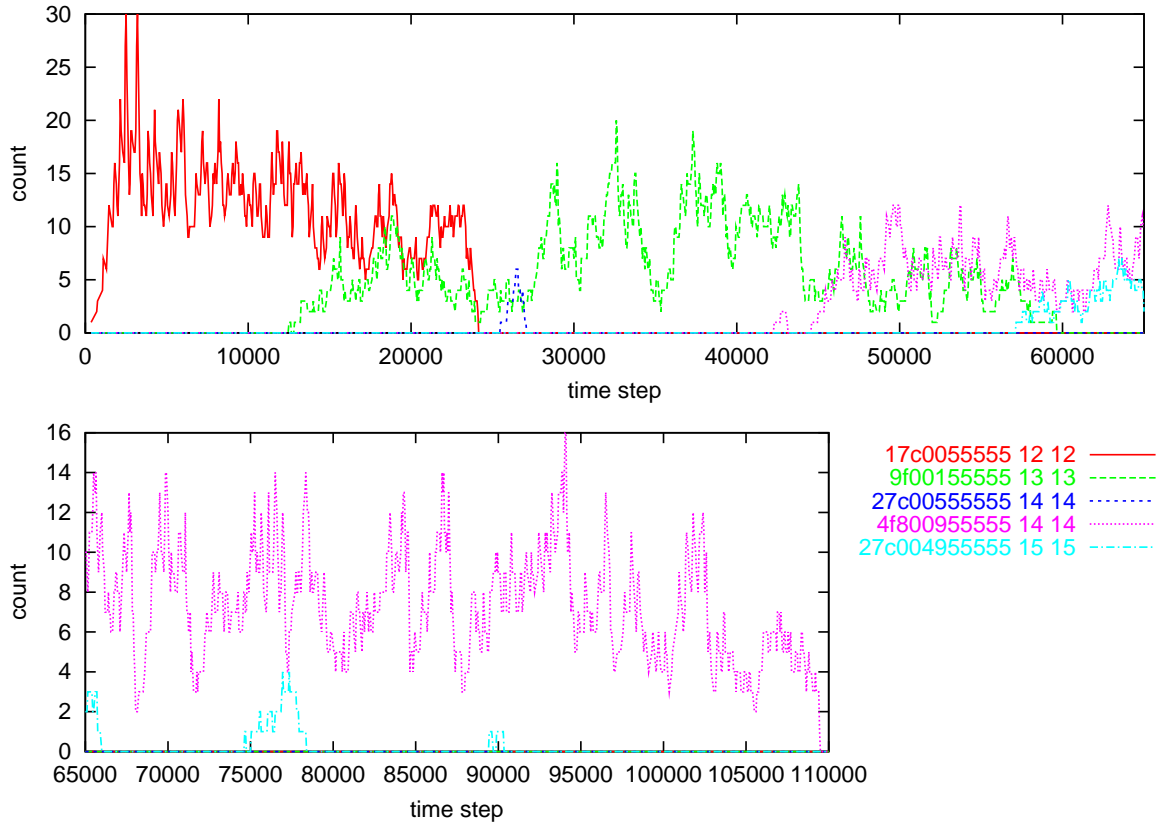


Fig. 6.3: Evolution of species 17c0055555/12 \times 12.

Our aim in this section is to analyze the subset of genealogy space explored in the run of Figures 6.1 and 6.3 in order to find any conserved subsequences which would explain the peculiar evolutionary dynamics observed. In total 224 different species are observed to appear over the course of this run, of which only 5 are stable species reaching populations of more than 5 individuals. Using the reduced representation for gene sequence configurations, these species look like:

Size	Gene sequence
12	GGGGGGGGG GCCCCCCCCCCTTG G
13	GGGGGGGGGG GCCCCCCCCCCTTG CG
14	GGGGGGGGGGG GCCCCCCCCCCTTG CGC
14	GGGGGGGGGGG GCCCCCCCCCCTTG CG
15	GGGGGGGGGGGG GCCCCCCCCCCTTG CG

We notice immediately an interesting pattern: that the evolution to stable species is modifying the sequence only in certain restricted areas. More specifically, we see that the subsequence of free C states appears to be preserved due to the presence of certain *mutation insensitive regions*[31]. In the larger set of 224 species which appear during the run, only a handful of species do not preserve this subsequence, and in doing so these species also lose the T genes (without which self-replication

is impossible). Hence we infer that the this loop *does not possess the evolutionary capability to shorten this subsequence*.

If true, how does this result generalize? Is there a general form of subsequence which cannot be broken in stable species by the emergent evolutionary processes of the evolop CA? The answer we present in this thesis is a tentative *yes*. The arrangement of genes in this subsequence, as derived from experimentation, follow the simple form:

$$G\{C\}T\{C\}TG$$

where $\{C\}$ represents any number of C states. This more general form implies that for a given loop size there are a number of different subsequences which — if injected into the ancestor loop — will be conserved through all subsequent mutations. From Section 5.4, we have derived that a loop of size n has $n - 2$ “free” Cs. Hence we can easily tabulate, for a given size of loop, which of these subsequences will fit within the loop:

minimum size	(2)	(3)	4	5	6	7	8
	GTTG	GOTTG	GCOOTTG	GCOOOTTG	GCOOOOTTG	GCOOOOOTTG	GCOOOOOOTTG
		GTOTG	GOTOTG	GCOOTOTG	GCOOOTOTG	GCOOOOTOTG	GCOOOOOTOTG
			GTCOOTG	GOTCOOTG	GCOOTCOOTG	GCOOOTCOOTG	GCOOOOTCOOTG
				GTCCOOTG	GOTCCOOTG	GCOOTCCOOTG	GCOOOTCCOOTG
					GTCCCCOTG	GOTCCCCOTG	GCOOTCCCCOTG
						GTCCCCOOTG	GOTCCCCOOTG
							GTCCCCCOOTG

In this table, a loop of size n may contain all subsequences in column n as well as all those to the left of this column. Note that the first two columns represent subsequences with less than 2 Cs, which will fit in all loops since physical constraints impose an absolute minimum loop size of 4. Thus a loop of size 4, for instance, will fit any of the 6 subsequences in the first three columns of this table.

If the conclusion we have presented is true, then in each of these cases evolution will never destroy these subsequences. While this result has not yet been analytically verified (by means of strict arguments based on CA rules), the form we have presented has been experimentally shown to resist evolutionary modification in a variety of different runs on grid sizes ranging from 200×200 to 3000×3000 , for time scales on the order of millions of iterations. In all cases we have observed, not a single stable species has been observed which does not contain the initial subsequence.

For the discussion which follows, we refer to a conserved subsequence as a *string* for short, and a string with k free Cs as a k -string. A loop with a string of k Cs we call a k -loop. We further introduce a sub-classification of stable loop species in terms of the string they conserve, which we specify by the number of Cs k in the conserved string and by the number of Cs between the T genes, which we refer to as p . Hence we arrive at the following table of loop classes and corresponding strings:

		k					
		0	1	2	3	4	...
p	0	GTTG	GTTG	GGTTG	GGGTTG	GGGGTTG	...
	1		GTCTG	GGTCTG	GGGTCTG	GGGGTCTG	...
	2			GTCTTG	GGTCTTG	GGGTCTTG	...
	3				GTCTTGG	GGTCTTGG	...
	4					GTCTTGGG	...
	⋮						...

Thus a stable loop of class $k - p$ has k ones in its conserved string, of which p are between the T genes. Using this new terminology, we note three important implications:

1. There are $k + 1$ different k -strings, with p varying between 0 and k .
2. A conserved k -string imposes a minimum loop size of $(k + 2)$. For $k > 1$, only *one* species can have this minimum loop size.
3. The state-space which a k -loop with size n will explore contains at most $\binom{2n-4-k}{n-2-k}$ stable species.

The last point is possibly the most important. It is derived from the result of Section 5.4 by noting that specifying the configuration of a given string (which includes the positions of the free T, TG pair and k Cs) removes $(k+2)$ free positions in the gene sequence, and there are k less free Cs to distribute. Hence the number of possible arrangements is reduced:

$$\binom{2n-2}{n-2} \mapsto \binom{(2n-2)-(k+2)}{(n-2)-k} = \binom{2n-4-k}{n-2-k}$$

For a loop of size $n = 4$ with the string GTTG ($k = 0$, $p = 0$), for example, there are 6 different stable species. These have the following gene sequences (cyclically permuted to align the conserved string):

(a)	GTTGCGG	(d)	GTTGCGG
(b)	GTTGCGG	(e)	GTTGCGG
(c)	GTTGCGG	(f)	GTTGCGG

Note that three of these configurations, (c), (d), and (e), were already observed in experiments of the previous chapter as species 13e9/4×4 (GCGTTGCG), 11f5/4×4 (GGTTGCCG), and 9f5/4×4 (GGTTGCGC), respectively. The small size of state-space for these short sequences highlights the importance of graph-based genealogy: from a variety of starting points with the same initial k -string (such as those of Section 6.3) evolution consistently leads to the same six attractor species.

Given the discovery of conserved subsequences, the results of Figure 6.1 are not surprising. With such a large string GCCCCCCCCCTTG ($k = 10$, $p = 0$), there is no way for loops in this experiment to evolve to smaller sizes; this string will simply not *fit* in a smaller loop! Yet if we look at the next

larger size of loop (size 13), we see that there are $\binom{(2(13)-4-10)}{(13-2-10)} = 12$ possible species. In general (as in this case), it is observed that evolution most easily accesses configurations in which free Cs are near the conserved string. This is demonstrated by the results observed for the experiment discussed above, in which dominant species of larger sizes all have additional free C states clustered near the conserved string.

Though the results discussed above may initially appear complicated, the result which they imply is in fact quite simple: if we begin an experiment with any given k -loop, evolution will never lead us to species which are smaller than a size $(k + 2)$. Thus a minimum loop-size is imposed on the global system by initial conditions. Moreover, for all $k > 1$, there is *only one such minimal species*. The question now arises: if we increase this threshold to larger sizes, will selection criteria change? As the vast majority of experiments performed in previous work [35, 37] were performed with $k = 0$ strings, this question is largely unanswered. It is clear that differences in reproduction rate will decrease in relative magnitude for larger loops. Hence it is conceivable, given a sufficiently large k , that differential fitness between loops of different sizes will lose significance, enabling other, emergent properties to play a more dominant role. In the next section we examine preliminary results from experiments which explore this possibility.

6.3 Experiments

The vast possibility space of both string types and species makes exploration of parameter space in its entirety impossible. As an alternative, we present in this section detailed results from two runs which summarize the majority of experimental results collected thus far. In the first experiment we observe evolution converging to a single minimal-sized loop along a complex trajectory in genealogy-space. In the second experiment we take the opposite perspective and observe a minimal species evolve to larger sized species, and then *back* to the minimal species, demonstrating the most evident example so far of graph-based genealogy.

6.3.1 Convergence to the minimal loop

If we assume — as was concluded in previous analyses [35, 37] — that the dominant factor contributing to the fitness of a species is the reproduction rate, then we would expect evolution to converge on the species for which size is minimized. For loops with $k > 1$, we have stated that only *one* such species exists. Hence, we might think that the system should naturally converge on this minimal species and that evolution should then stop.

Yet in making this prediction we have disregarded the emergent nature of evolution and genealogical connectivity in the evolloop system. Fitness clearly plays a prominent role in deciding the survival of one species over another, yet equally important are the collection of genealogical links through which species evolve. We have indicated in previous chapters that this genealogy space is complicated and not yet well understood. In this section we show a clear example to this effect.

For this experiment we begin with a stable size-14 species with gene sequence `GGGGCGGGGGGGG TCCCCCCCCCCTGG` on a 500×500 grid. This species has the conserved string `GTCCCCCCCCCCTG` for which ($k = 11, p = 0$), thus we expect a minimal achievable loop size of 13. The only difference between the initial size-14 species and the optimal size-13 species is a single C state near the middle

of the G triplets and one extra G triplet. The aim of our experiment is to discover whether evolution can attain this minimal size, and if it can, the path it follows in order to find it.

Population dynamics for the first 3M iterations of this run are plotted in Fig. 6.4. Surprisingly, it is only at the end of this period that the optimal size-13 species finally emerges following a series of five clearly distinguishable epochs, each characterized by a different dominant species. This protracted period of time to reach peak fitness — where we continue to assume fitness to be size-based — is orders of magnitude longer in this case than in any earlier experiments[35, 37], millions of iterations compared with tens of thousands. So too is the explosive diversity: a total of over 6500 species¹ are observed in the first 3M iterations, a further 2000 recorded before the run was stopped at 5.17M iterations. Moreover, the evolutionary path followed by the system during this period is anything but predictable, traversing both larger and smaller species before converging on the minimal size-13 loop. The sequence of species observed in this run with counts greater than 30 individuals, ordered according to first appearance in large numbers, is as follows:

Hex label		Gene sequence
5c006aaaa55	14×14	GGGGGGGGGG GTCCCCCCCCCTG G
27001aaaa955	15×15	GGGGGGGGGG GTCCCCCCCCCTG GG
11c006aaaa555	16×16	GGGGGGGGGG GTCCCCCCCCCTG GGG
87001aaaa9555	17×17	GGGGGGGGGG GTCCCCCCCCCTG GGGG
2e0034aaaa55	15×15	GGGGGGGGGG GTCCCCCCCCCTG G
5c006aa5555	14×14	GGGGGGGGGG GTCCCCCCCCCTG G
2e0034aa5555	15×15	GGGGGGGGGG GTCCCCCCCCCTG G
b800d2a95555	17×17	GGGGGGGGGG GTCCCCCCCCCTG G
b800d55555	13×13	GGGGGGGGGG GTCCCCCCCCCTG G

The first four species in the sequence above differ only by the presence of injected C genes at the head (right) of the sequence and by added G genes at the tail (left). This would seem to indicate some global pattern in the genetic adaptation process, one which favours mutations in the region directly to the right of the conserved string. The tendency of mutations to occur at the edge of the conserved string conforms with our earlier observation in Section 6.2 regarding mutability of the size-12 “n-evolop”.

It is only after roughly 550K iterations that the evolutionary trend towards larger loops is disrupted. The appearance of a new size-15 loop marks a branching point in the genealogical lineage of this population, introducing a modification to the genotype in which the central C gene is displaced two places towards the tail of the sequence. This new species maintains an evolutionary stasis for roughly 200K iterations, at which time a smaller, size-14 loop emerges to drive out its larger competitor.

The period between 750K and 1.4M iterations offers possibly the most clear example thus far of the interplay between differential fitness and genealogical connectivity in evolutionary processes of the evolop. Two related species, the 5c006aa5555/14×14 and the 2e0034aaaa55/15×15, compete fiercely over nearly 700K iterations in a period characterized by intermittent bursts of activity and brief periods of stasis. The regularity of these bursts and the similarity in their dynamics is striking, with occurrences at rough intervals of 200K iterations beginning at 950K. While the larger size-15

¹ We include here *all* classes of species: stable, transitional and terminal.

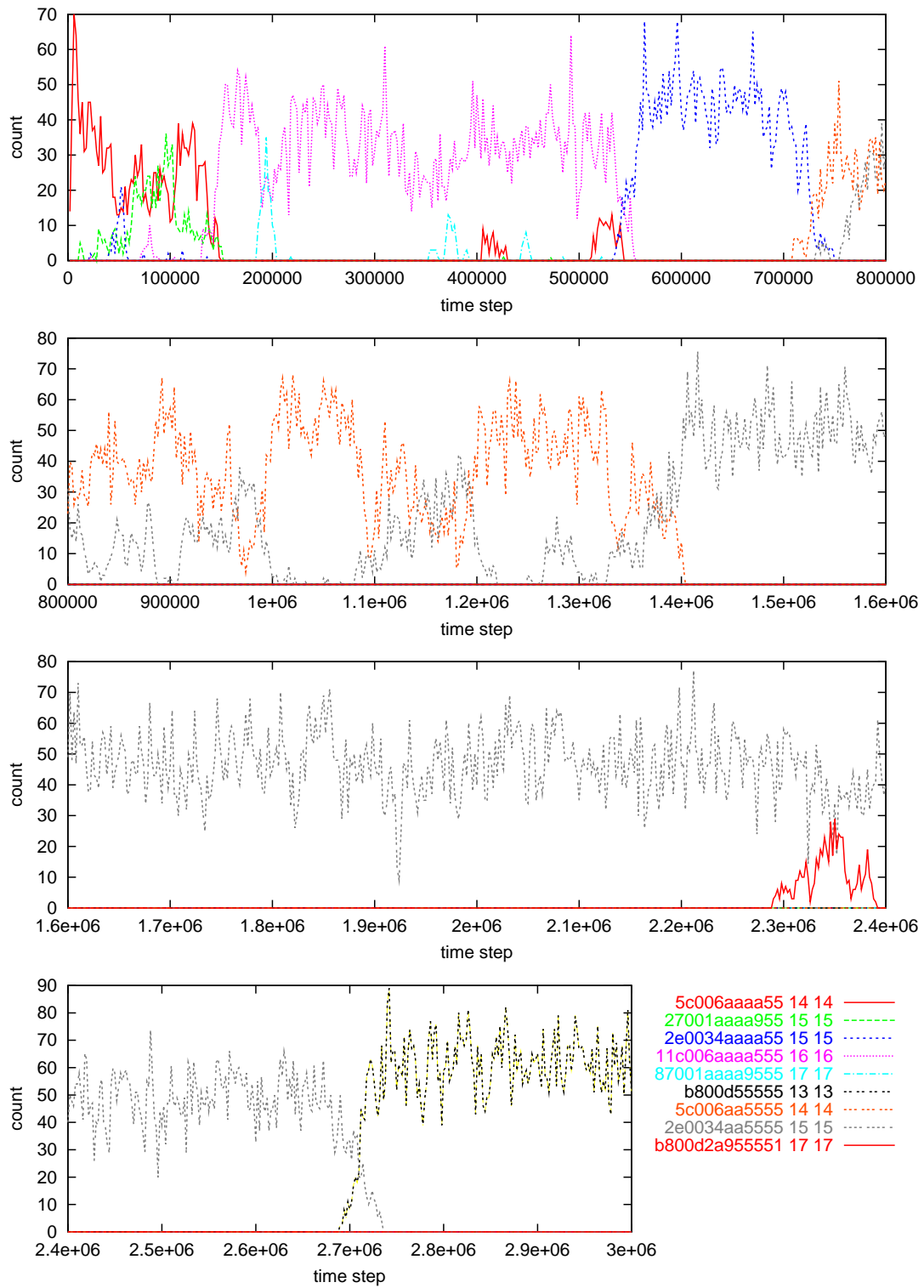


Fig. 6.4: Convergence to minimal species.

species disappears and recurs many times during this process, the smaller, fitter size-14 species goes completely extinct only once, at 1.4M iterations, after which it never returns. A snapshot of the CA space at 1.16M iterations is shown in Fig. 6.5 at a time when the size-15 species (labeled (a) in this figure) is most populous. Five instances of an even larger species, the $8e0034aa55555/17 \times 17$ (labeled (c)) are also present at this instant, though these loops were not included in the plot of Fig. 6.4 due to their low numbers. The role played this and other short-lived stable species in evolutionary processes is not yet completely clear, nor is the effect of the numerous remnants and many transitional and terminal species, which themselves form a kind of dynamic environment. We can however infer based on population dynamics that size-based fitness, which favours the smaller size-14 species, is being counter-balanced by graph-space connectivity, which flows strongly *from* the smaller species *to* the larger one.

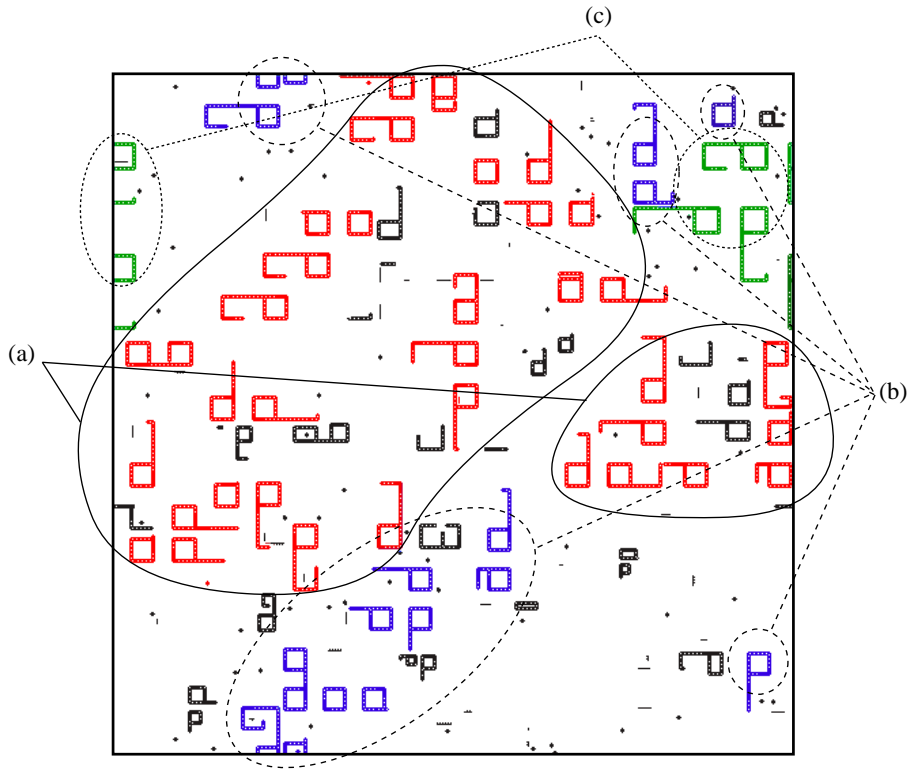


Fig. 6.5: Snapshot of CA space at 1.16M iterations. Red loops (a) = $2e0034aa55555/15 \times 15$, purple loops (b) = $5c006aa55555/14 \times 14$, green loops (c) = $8e0034aa55555/17 \times 17$. Transitional and terminal species as well as remnants are drawn in black.

As discussed earlier in Section 5.3.2, graph-based dynamics of the evoloop in many ways bear resemblance to the quasispecies model of Eigen, McCaskill and Schuster. In this model, a similar tendency for survival of genealogically well-connected species is mentioned in the context of template-based replication in a chemical system. Eigen et. al.[11] remark that “cases were investigated where two sequences have slightly different selective values, but the less efficient sequence

lives in a more efficient neighbourhood”. “Backflow” — the tendency for species recur via mutation events — is used to describe the process by which these less efficient species (sequences) are re-introduced via mutations from other, fitter ones.

The quasispecies idea describes behaviours observed in Figure 6.4 remarkably well. Yet models used in [11] to realize this concept take the form of sets of differential equations and hence differ fundamentally from states and transition rules of the evoloop CA. In particular, quasispecies models implemented as differential equations require a number of assumptions never introduced here. The mutational backflow observed in Figure 6.4, for instance, poses a serious challenge to these models. In the case of the evoloop this phenomenon emerges naturally from local CA rules. In this light, as an *emergent* model for evolution, it may be that the evoloop system better captures certain qualitative aspects of the quasispecies concept than do differential equation models presented in [11]. More thorough investigations assessing the full potential of these emergent dynamics — to be discussed in Chapter 7 — would address the implications of this possibility.

6.3.2 Cyclic genealogy and emergent selection

One of the implications of the theory of quasispecies is that selection acts not on *individual species*, but on interrelated *groups*. As stated in [11]: “the fitness of a given sequence is not only determined by its own selective value. Rather, neighboring sequences contribute too, through mutations, and their influence becomes more important as the error rate increases.” Evolutionary behaviour in experiments of the previous section reveal that qualitatively similar selection mechanisms act on evoloop populations. In this section we focus on strengthening the validity of this comparison.

The run we perform is carried out on a grid of size 1000×1000 . To this domain we introduce an ancestor species $5c001855555555/18 \times 18$ (GGGGGGGGGGGGGGCCCTCCCCCCCCCCCCCTGG), a size-18 stable loop of class 16-13 according to the definition introduced in Section 6.2. Since this is the minimal-sized species for this conserved string, we expect only stable species of larger sizes to evolve from it.

Population dynamics for the first 11.57M iterations of this run are plotted in Figure 6.6. Note that the time scale for this experiment is significantly longer than that of Figure 6.4 in part due to the larger size of domain; more time is required for a new emerging species to completely topple its evolutionary competitor. Contrary to the results of the last section, only a few epochs and three major species characterize this run. These species again differ in close proximity to the conserved string. A complete list of species whose count exceeds 30 individuals at any time during the run is given below:

Hex label	Gene sequence
5c001855555555 18×18	GGGGGGGGGGGGGGG G CCCT TCCCCCCCCCCCCCTG G
2e000c255555555 19×19	GGGGGGGGGGGGGGG G G CCCT TCCCCCCCCCCCCCTG G
27000c255555555 19×19	GGGGGGGGGGGGGGG G CCCT TCCCCCCCCCCCCCTG CG
138003095555555 20×20	GGGGGGGGGGGGGGG G G CCCT TCCCCCCCCCCCCCTG CG
11c001855555555 20×20	GGGGGGGGGGGGGGG G CCCT TCCCCCCCCCCCCCTG CG
170006125555555 20×20	GGGGGGGGGGGGGGG CG G CCCT TCCCCCCCCCCCCCTG G
9c001855555555 20×20	GGGGGGGGGGGGGGG G CCCT TCCCCCCCCCCCCCTG CG
17000615555555 19×19	GGGGGGGGGGGGGGG G CCCT TCCCCCCCCCCCCCTG G CG
b8003095555555 20×20	G GG GGGGGGGGGGGGG G G CCCT TCCCCCCCCCCCCCT

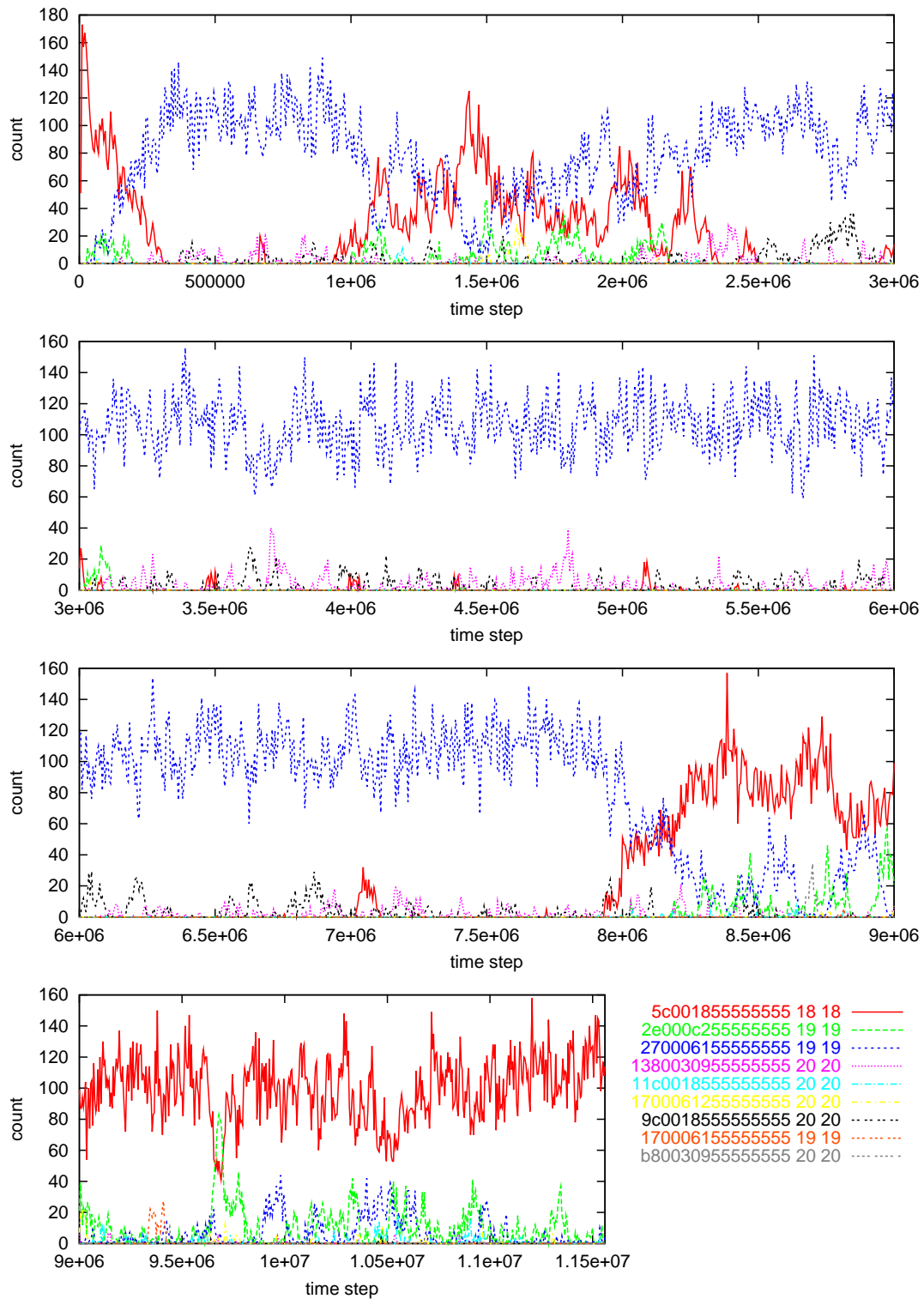


Fig. 6.6: Evolution of species 5c001855555555/18×18.

Dominant species of Figure 6.6 are the first (minimal) size-18 species and its two closest relatives, each of which differs by the addition of a **C** gene at one side of the conserved string and by an extra **G** gene; these species are the first three entries in the table above. Note that subsequent sequences exhibit similar variation. The fourth entry, for instance, is a size-20 species which has added **C** genes at *both* ends of the string. Symmetry in this dominant set of sequences is surprisingly strong. Evolutionary implications of these subtle gene re-arrangements on fitness and genealogical connectivity, however, are not yet clear.

Although the number of evolutionary epochs in this run is noticeably less than the experiment of Figure 6.4, the observed diversity is exceedingly large with a total of just under 23K species emerging at some moment during in the 11.57M iterations. Moreover, as did the two dominant species of Fig. 6.5, this system demonstrates clearly by the multiple recurrences and extinctions of the minimal size-18 loop and its two closest size-19 relatives the existence of *cyclic* chains in graph-space. In particular, during the period between 2.5M and 8M iterations the population of minimal-sized loops is suppressed by the dominant status of one of its size-19 relatives. Yet the stasis which is reached is disrupted abruptly at 8M iterations and the minimal loop returns to power. The significance of this event is that, in genealogical terms, the system has returned to its starting point, in the meanwhile having explored a relatively broad range of genealogical graph-space. Resemblances to the “nonsymmetric cloud of mutants”[11] of the quasispecies model are again evident. A brief rise in population of the other size-19 relative at roughly 9.7M iterations, moreover, hints at further unpredictable dynamics to come.

The results of this experiment clearly demonstrates that *fitness* \neq *replication-rate*, contrary to expectation[1, 35, 37]. Moreover, there is evidence that self-generated environments have the potential to *modify* differential fitness. This is evidenced by the recurrence in large numbers of the minimal size-18 loops at 8M iterations: only at this time were the conditions right for an invasion to occur, conditions created by space-management of the previously-dominant size-19 species. It is a fundamental characteristic of the evolloop system that survival is not determined by intrinsic fitness alone, but also on environmental conditions. This is much the way, as Hogeweg[15] describes, “a biotic system has to deal with itself in its most direct environment”. This result again confirms that what we perceive is *real* in this abstract model agrees with the reality biologists find in real organisms.

6.3.3 Further observations

A series of experiments were performed on species with long conserved strings, two of which were described above. Other results are still being analyzed but exhibit similar evolutionary behaviours. One of these experiments was performed with a series of ancestor species all having the number of **Cs** (k) in their conserved string but with varying p values. We would expect the minimal loop in all cases to be the same size, yet it is not obvious whether the dynamics would be the same or not. Results on a grid size of 1000×1000 demonstrated that, while behaviours were *similar*, they were not the same. In all cases, for instance, mutation was observed to act in close proximity to the conserved string. However differences in dynamics emerging from shifts of the **T** gene through the string (for different p -values) gradually propagated through the entire CA space, initiating a rapid divergence of evolutionary paths between runs with different ancestor species.

If there is a general consequence to fitness of this p -value, it is not yet clear from these

preliminary experiments. Other challenging problems, such as describing the mapping between gene arrangements and phenotypic behaviours, remain largely open and potentially unanswerable. Nonetheless, the fact that these complex and in many ways “messy” behaviours exist is both the attraction and the challenge of this model.

6.4 Summary

In this chapter we introduced a new class of stable evolloop species demonstrating markedly different evolutionary dynamics when compared with results of previous analyses. The discovery of this new class resulted from the re-examination of a very early experiment performed in [35] and by considering the full potential of the vast genealogical state-space outlined in Section 5.4. At the center of this new class of species was the idea of *mutation-insensitive regions* resulting in *conserved subsequences* in the evolloop genome. Appropriately injected into a sufficiently large phenotype, these subsequences enforce self-reproducing loops to maintain an arbitrary minimum size. Moreover, for subsequences with at least two C genes, at most *one* such minimal loop was shown to exist.

Experiments in Section 6.3 verified claims of the first section. In the first run, a loop of size 14 with a subsequence enforcing a minimum size of 13 was introduced to a 500×500 grid. Convergence to the minimal loop was achieved, yet only after 2.7M iterations and many evolutionary transitional epochs. A balance between *differential fitness* and *genealogical connectivity* was shown to play a pivotal role in this process, inviting comparison to the quasispecies model of Eigen, McCaskill and Schuster[11]. In the second experiment, this comparison was strengthened with the discovery that emergent selection mechanisms are not exclusively size-based. Gene positioning, even between seemingly fitness-neutral variants, was found to have critical implications on evolutionary process, much as in real biology.

7. CONCLUSIONS AND FUTURE DIRECTIONS

7.1 Summary of results

In this thesis we proposed a new approach to the study of artificial self-replication. The great majority of research in Artificial Life is invested in model design and construction; here we have instead focused on detailed analysis. We chose as our model the evolloop[35, 37], a simple nine-state five-neighbour cellular automaton with emergent evolutionary properties. The history of this CA traces back to the work of Langton[22] and further to the first steps of von Neumann[49]. A distinguishing feature in this line of models is the dependence of self-replication mechanisms on the structure of self-replicators. In the case of the evolloop, this structure-dependence extends to evolutionary dynamics: self-replicating loops mutate through direct interaction (collision) of sheath structures, leading to a change in the gene sequence of offspring loops.

Devising sophisticated and efficient analysis routines to detect and identify these self-replicating structures was the first goal of this thesis. We accomplished this goal by imposing strict criteria for *birth* and *death* and by introducing a concise and well-defined identification scheme. A method for efficient genealogy-tracing was also implemented. Compared with previous analysis techniques which use exclusively size-based identification, these new fine-grained tools provide a *complete* description of evolutionary dynamics; every self-replicator is identified with the resolution required for complete reconstruction of its configuration in the CA space.

Given these new fine-grained tools, we proceeded our study with a thorough investigation of mutation processes in the evolloop CA. Borrowing from ideas presented by Ikegami[18, 19], we classified mutations as either *passive* or *active* and used this terminology to discuss properties we called *stationary* and *reproductive*. This led us to classify loop species as either *stable*, *terminal* or *transitional* according to the nature of their links in an abstract genealogical graph-space, one which we termed a “genealogy graph”. An estimate on the number of stable species — those which self-replicate exactly — was shown to grow combinatorially with loop size, reaching millions for size 13 loops and billions for size 18 loops.

This vast possibility-space begged the question: is there any diversity in this huge set of evolloop species? To answer this question a behavioural study was conducted on a set of three size-6 stable species. Spatial pattern formation was shown to differ strikingly in each case due to self-generated environments and the creation of *remnants* — contiguous structures disconnected from any live loop. Population dynamics further emphasized these distinctions. Although in each case evolution proceeded to minimal size-4 loops, exploration of genealogical state-space, as illustrated using a visualization method introduced in [32, 33], was extensive in one case, limited in others. Detailed reconstruction of an evolutionary event leading to a larger loop highlighted the potential for evolloops to achieve increases in genotypical complexity through simple interactions. Yet differential fitness, strongly favouring small-sized species, overwhelmed this potential for extensive state-space

exploration, in all cases resulting in homogeneous minimal-sized populations after short transient periods.

The last chapter of this thesis was devoted to tackling this problem of *predictability*, detrimental to the sustainability of genetic diversity. Beginning with earlier experiments performed with the “n-evoloop” [35, 37], we showed that certain gene subsequences are conserved through evolution, enforcing a minimum loop size which may be arbitrarily imposed by appropriate initial conditions. This discovery revealed the existence of different classes of stable species, limited by their inability to mutate into a member of a different class. The state-space potentially spanned by members of a single class was again shown to grow combinatorially with loop size. While classes with a small minimum loop size — the subject of nearly all previous experimentation — would predictably optimize to this minimal species, it was expected that larger loops with less pronounced differences in size-based fitness might behave differently.

A set of experiments were performed to test this hypothesis. In one case, a size-14 loop was shown to search the state-space of loop species over 2.7M iterations before finding the minimal size-13 loop. During this search loop populations evolved through five transient epochs characterized by different dominant species. In a second experiment a larger yet minimal size-18 loop evolved *away* from the minimal size and *back*, after a period of 8M iterations during which time a larger size-19 loop dominated. Both experiments demonstrated conclusively a critical balance in emergent evolutionary processes between *size-based fitness* and *genealogical connectivity*. In this context, strong similarities between observed behaviours of this CA model and the quasispecies theory of Eigen, McCaskill and Schuster [11] were discussed.

7.2 Conclusions

It is a surprise, especially for those well acquainted with the capabilities and limitations of cellular automata, that the simple model employed in this thesis can produce such complex genetic diversity and nontrivial evolutionary behaviour as its emergent property. Indeed, the simplicity and scalability of the evoloop proved to be among its greatest assets, demonstrating that *complex evolutionary phenomena need not require a complex model*. Yet the detection of this complexity — traditionally assumed to emerge in an obvious manner — involved a highly specialized procedure of structure recognition and identification. Significant effort paid to the development of these “high-resolution” methods was handsomely rewarded with the discovery of complex state-space exploration through a vast and diverse graph-based genealogy. All of this was hidden in subtleties of gene arrangement and spacing, features undetectable to the casual observer. This fact manifests the importance of sophisticated observation and interpretation of model dynamics to capture the true richness of lifelike behaviour and the hierarchical complexity increase emerging at multiple scales within the model. This has long been underestimated compared to model construction in self-replication studies.

In terms of achieving *open-ended, complexity-increasing* evolution in artificial media — one of the ultimate goals of research in Artificial Life — results we have presented mark a significant step forward. The revelation of this thesis is that *things are not as simple as they seem*: the evoloops we thought we knew represent only a minute fraction of a much larger, diverse set, one which grows combinatorially with loop size. Among this multitude of species are complete subsets for which emergent selection mechanisms differ fundamentally from those investigated in previous studies.

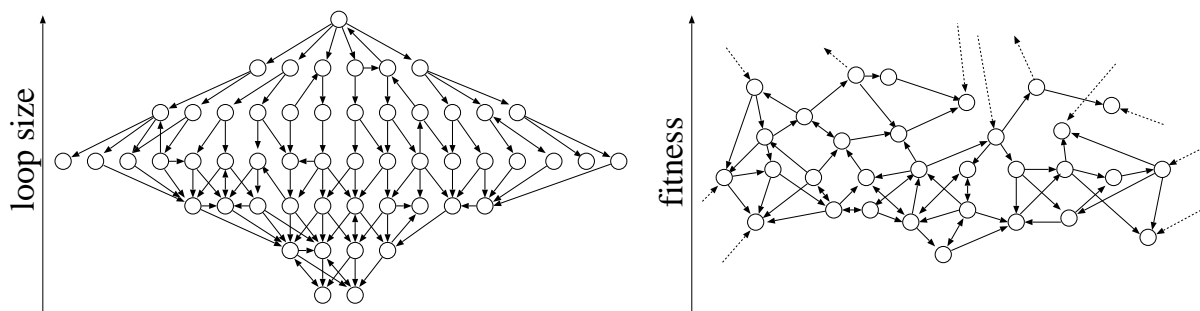


Fig. 7.1: Graph-based genealogy: optimization towards smallest size (left) and open-ended exploration (right). For species with small minimal size (left), fitness is exclusively size-based; for those with larger minimum size (right), the importance of size is diminished allowing other factors to modify the fitness landscape.

The genealogical trajectories of Fig. 7.1 emphasize this distinction. Whereas the fitness differential encountered in earlier experiments clearly favoured small-sized species, new results demonstrate unpredictable and potentially open-ended dynamics due to a “flattening” of the size-based fitness landscape. Additional, emergent selection mechanisms, previously suppressed by reproduction rate, play a relevant and critical role in this new context.

Finally, we would like to emphasize that we have not yet reached any succinct explanation or principle that governs the entire dynamics of the evolooop system, even though it is just a well-defined deterministic cellular automata whose transition rules are completely known. Each time we tried to construct some neat mathematical formulations to describe its evolution, we soon found other minor but crucial factors coming in to bring the whole system toward unpredictable directions. The messy nature of the evolooop world may well capture what biologists are struggling with in studying real life. In this sense, we believe that what we perceive is *real* in this abstract model qualitatively agrees with the reality biologists find in real organisms, and that this is the most profound proposition underlying the study of abstract lifeforms.

7.3 Future directions

A number of future project ideas branching from the results of this thesis are envisioned. Parallel to the work presented here, a large number of experiments have been performed with the persistent dissolver state discussed briefly in Section 5.5.1 and with much more detail in [2]. The effects of this new state has been found to encourage sustainable diversity leading to speciation, punctuated equilibrium, and evolutionary “bottlenecking”. Introduction of the newly discovered class of species discussed in Chapter 6 to the dynamic environment generated by this dissolver state has yet to be attempted. Long-term evolutionary dynamics of the new loops combined with partitioning properties of the persistent dissolver could produce quite interesting results. Other types of partitioning environments and nonperiodic domains may also be attempted, as well as much larger-scale experiments.

Behavioural analysis of the complete genealogical state-space of stable evolooop species is also an important topic yet to be addressed. Due to the size of this space there are serious impracticalities

involved in testing each species individually. An automated statistical approach, however, may provide insight into the evolutionary consequences of gene positioning, an issue granted only cursory analysis in this thesis.

As a final but important future goal, construction of a new model building on insight gained through the work on this project is also planned. Details of this model await further testing.

7.4 *Acknowledgments*

First and foremost I would like to sincerely thank my collaborators for all the hard work and long hours they invested in this project: my research partner Antony Antony and our supervisor Hiroki Sayama. No words suffice to express the huge contribution they have made to the results contained in this thesis. The evolloop continues to astound and amaze us.

I also owe a great deal of thanks to Roeland Merks, who brought an invaluable biological viewpoint to the project, and to Paulien Hogeweg for her enthusiasm and support. Discussions with participants and lecturers at the Santa Fe Institute's 2003 Complex Systems Summer School were influential in sharpening the focus of this thesis. The conversations I had with Takashi Ikegami and Keisuke Suzuki during my visit to Japan were very beneficial and constructive; those with Barbara Gene always brought a surprise, reminding me that Artificial Life is more than just dry science and should be fun and creative. Thanks also to Daniela Gavidia who contributed to the course project which developed into this thesis, and to the SCS for allowing me to research a topic which in many ways falls outside the scope of computational science.

Finally, thanks to my dad, who has fostered my interest in understanding how the world works for as long as I can remember; to my mom, for encouraging me to keep that essential touch of creativity in the mix; and to my brother, for his many witty yet frustratingly accurate criticisms.

Funding for this research was provided by a scholarship from the Quebec-based funding organization Fonds Québécois sur la nature et les technologies and by a grant from the International Information Science Foundation.

BIBLIOGRAPHY

- [1] C. Adami. On modeling life. *Artificial Life*, 1:429–438, 1994.
- [2] A. Antony, C. Salzberg, and H. Sayama. A closer look at the evolutionary dynamics of self-reproducing cellular automata. Accepted pending revision.
- [3] P. Bak and K. Chen. Self-organized criticality. *Physica A*, 163:403–409, 1990.
- [4] M. C. Boerlijst and P. Hogeweg. Spiral wave structure in pre-biotic evolution: Hypercycles stable against parasites. *Physica D*, 48:17–28, 1991.
- [5] A. W. Burks, editor. *Essays on Cellular Automata*. University of Illinois Press, Urbana, Illinois, 1970.
- [6] J. Byl. Self-reproduction in small cellular automata. *Physica D*, 34:295–299, 1989.
- [7] H. H. Chou and J. A. Reggia. Emergence of self-replicating structures in a cellular automata space. *Physica D*, 110:252–276, 1997.
- [8] J. P. Crutchfield. Is anything ever new?: Considering emergence. In G. Cowan, D. Pines, and D. Melzner, editors, *Santa Fe Institute Studies in the Sciences of Complexity XIX*. Addison-Wesley, Reading, MA, 1994.
- [9] D. Dennett. *Encyclopedia of Evolution*, pages E83–E92. Oxford University Press, New York, 2002. ed. Pagel, M.
- [10] M. Denton. *Evolution: A Theory in Crisis*. Adler & Adler, New York, 1986.
- [11] M. Eigen, J. McCaskill, and P. Schuster. Molecular quasi-species. *J. Phys. Chem.*, 92:6881–6891, 1988.
- [12] M. Eigen and P. Shuster. *The Hypercycle: A Principle of Natural Self-Organization*. Springer, Berlin, 1979.
- [13] M. Gardner. The fantastic combinations of john conway’s new solitaire game life. *Scientific American*, 224(2):112–117, 1970.
- [14] M. Gardner. Mathematical games: A mathematical zoo of astounding critters, imaginary and otherwise. *Scientific American*, 238(6):18–28, 1978.
- [15] P. Hogeweg. Multilevel evolution: replicators and the evolution of diversity. *Physica D*, 75:275–291, 1994.

-
- [16] T. J. Hutton. Evolvable self-replicating molecules in an artificial chemistry. *Artificial Life*, 8:341–356, 2002.
- [17] T. Ikegami. From genetic evolution to emergence of game strategies. *Physica D*, 75:310–327, 1994.
- [18] T. Ikegami. Evolvability of machines and tapes. *Artificial Life and Robotics*, 3:242–245, 1999.
- [19] T. Ikegami and T. Hashimoto. Replication and diversity in machine-tape coevolutionary systems. In C. Langton and K. Shimohara, editors, *Artificial Life V: Proceedings of the Fifth International Conference on Artificial Life*, pages 426–433. MIT Press, 1997.
- [20] A. Ilachinski. *Cellular Automata: A Discrete Universe*. World Scientific Publishing Co. Pte. Ltd., 2001.
- [21] K. Imai, T. Hori, and K. Morita. Self-reproduction in three-dimensional reversible cellular space. *Artificial Life*, 8:155–174, 2002.
- [22] C. G. Langton. Self-reproduction in cellular automata. *Physica D*, 10:135–144, 1984.
- [23] C. G. Langton. Studying artificial life with cellular automata. *Physica D*, 22:120–149, 1986.
- [24] C. G. Langton, editor. *Artificial Life: The Proceedings of an Interdisciplinary Workshop on the Synthesis and Simulations of Living Systems Held September, 1987, in Los Alamos, New Mexico*. Addison-Wesley Publishing Company, 1989.
- [25] C. G. Langton. Artificial life. In *The Philosophy of Artificial Life*, pages 39–94. Oxford University Press, 1996.
- [26] J. M. Maynard-Smith and E. Szatháry. *The Major Transitions in Evolution*. W. H. Freeman & Company Ltd., 1995.
- [27] K. Morita and K. Imai. A simple self-reproducing cellular automaton with shape-encoding mechanism. In C. G. Langton and K. Shimohara, editors, *Artificial Life V: Proceedings of the Fifth International Workshop on the Synthesis and Simulation of Living Systems*, pages 489–496, Nara, Japan, 1996. MIT Press.
- [28] N. H. Packard and M. A. Bedau. Artificial life. In *Encyclopedia of Cognitive Science*. MacMillan Reference Ltd., 2000.
- [29] A. N. Pargellis. The evolution of self-replicating computer organisms. *Physica D*, 98:111–127, 1996.
- [30] T. S. Ray. An approach to the synthesis of life. In *Artificial Life II*, volume XI of *SFI Studies on the Sciences of Complexity*, pages 371–408. Addison-Wesley Publishing Company, Redwood City, California, 1991.
- [31] C. Salzberg, A. Antony, and H. Sayama. Genetic diversification and adaptation of self-replicators discovered in simple cellular automata. Submitted.

-
- [32] C. Salzberg, A. Antony, and H. Sayama. Visualizing evolutionary dynamics of self-replicators: A graph-based approach. Submitted.
- [33] C. Salzberg, A. Antony, and H. Sayama. Visualizing evolutionary dynamics of self-replicators using graph-based genealogy. In *Proceedings of the Seventh European Conference on Artificial Life*, Dortmund, Germany, 2003. In press.
- [34] N. J. Savill, P. Rohani, and P. Hogeweg. Self-reinforcing spatial patterns enslave evolution in a host-parasitoid system. *J. Theor. Biol.*, 188:11–20, 1997.
- [35] H. Sayama. *Constructing Evolutionary Systems on a Simple Deterministic Cellular Automata Space*. PhD thesis, Department of Information Science, University of Tokyo, December 1998.
- [36] H. Sayama. Introduction of structural dissolution into langton’s self-reproducing loop. In C. Adami, R. K. Belew, H. Kitano, and C. E. Taylor, editors, *Artificial Life VI: Proceedings of the Sixth International Conference on Artificial Life*, pages 114–122, Los Angeles, California, 1998. MIT Press.
- [37] H. Sayama. A new structurally dissolvable self-reproducing loop evolving in a simple cellular automata space. *Artificial Life*, 5:343–365, 1999.
- [38] H. Sayama. Self-replicating worms that increase structural complexity through gene transmission. In M. A. Bedau, J. S. Caskill, N. H. Packard, and S. Rasmussen, editors, *Artificial Life VII : Proceedings of the Seventh International Conference on Artificial Life*. MIT Press, 2000.
- [39] M. Sipper. Fifty years of research on self-replication: An overview. *Artificial Life*, 4:237–257, 1998.
- [40] M. Sipper and J. A. Reggia. Go forth and replicate. *Scientific American*, 285(2):26–35, 2001.
- [41] A. Smith, P. Turney, and R. Ewaschuk. Self-replicating machines in continuous space with virtual physics. *Artificial Life*, 9:21–40, 2003.
- [42] D. Stassinopoulos, S. P. Colombano, J. D. Lohn, G. L. Haith, J. Scargle, and S. Liang. Spatial autocatalytic dynamics: An approach to modelling prebiotic evolution. In *Proceedings of the 1998 International Conference of Complex Systems*. Nashua, NH, 1998.
- [43] A. Stauffer and M. Sipper. An interactive self-replicator implemented in hardware. *Artificial Life*, 8:175–183, 2002.
- [44] K. Suzuki and T. Ikegami. Interaction based evolution of self-replicating loop structures. In *Proceedings of the Seventh European Conference on Artificial Life*, Dortmund, Germany, 2003. In press.
- [45] T. Toffoli and N. Margolus. *Cellular Automata Machines: A New Environment for Modeling*. MIT Press, Cambridge, Massachusetts, 1987.
- [46] A. M. Turing. On computable numbers with an application to the entscheidungsproblem. *Proc. London Math. Soc. Ser.*, 42:230–265, 1936.

-
- [47] A. M. Turing. The chemical basis of morphogenesis. *Trans. Roy. Soc., Series B*, 237(5), 1952.
- [48] J. von Neumann. The general and logical theory of automata. In L. A. Jeffress, editor, *Cerebral Mechanisms in Behavior — The Hixon Symposium*, pages 1–41. John Wiley, New York, NY, 1951.
- [49] J. von Neumann. *Theory of Self-Reproducing Automata*. University of Illinois Press, Urbana, Illinois, 1966. Edited and completed by A. W. Burks.
- [50] S. Wolfram. Origins of randomness in physical systems. *Phys. Rev. Lett.*, 55:449–452, 1985.
- [51] S. Wolfram. Undecidability and intractability in theoretical physics. *Phys. Rev. Lett.*, 55:735–738, 1985.
- [52] G. Yedid and G. Bell. Macroevolution simulated with autonomously replicating computer programs. *Nature*, 420:810–812, 2002.