# RBN-World

## Sub-symbolic Artificial Chemistry for Artificial Life

Adam Faulconbridge

PhD

University of York

Department of Biology

September 2011

**Abstract**

Artificial Chemistry seeks to explore how life-like systems can emerge from a pre-biotic environment. This thesis begins with the background of this research area and a re-implementation of an existing Artificial Chemistry as a case study. From this basis, ingredients and properties of Artificial Chemistries are identified. This leads to a novel form of molecular representation — sub-symbolic. A group of novel Artificial Chemistries called RBN-World is developed using Random Boolean Networks as a sub-symbolic molecular representation. It is shown that RBN-World has several properties of interest, and variants of RBN-World and elemental subsets with those properties are identified from many alternatives. This thesis concludes by comparing RBN-World to the case study and properties discussed earlier, and identifies avenues for future work.

# Contents

# List of Figures

# List of Tables

# Acknowledgements

Many thanks to all who have supported me, especially my late father Alan John Faulconbridge.

# Declaration

I declare that this thesis is entirely my own work. Where information has come from other works, this has been referenced accordingly.

Some sections of this thesis are based on work I have previously published. In particular, chapter 4 is based on work produced collaboratively by Hickinbotham, Faulconbridge, and Nellis (2010). Chapters 5 and 6 are based on work published in Faulconbridge, Stepney, Miller, and Caves (2011) and chapter 7 is based on work published in Faulconbridge, Stepney, Miller, and Caves (2010).

# Chapter 1

# Introduction

## 1.1 Aim

One of the most fundamental questions mankind has ever asked is "what is life?" Although many people would claim to know the answer, a definition that can be widely agreed upon and applied without special cases does not exist. It is assumed that one contributing factor for this is due to the common ancestry of living organisms. This prevents the separation of characteristics that are shared because they are fundamental to being alive from those characteristics that are shared by common descent.

The development of novel forms of Artificial Life may provide further information by avoiding shared descent. However, in contrast to some other works in the field of Artificial Life, the living systems must be an emergent property rather than being built-in to avoid being a trivial self-fulfilling prophecy — this is alive because it is defined as being alive. To that end, Artificial Chemistry is the selected paradigm and this work aims towards the development of prebiotic environments suitable for the emergence of artificial life-like systems.

## 1.2 Outline

In order to frame the aim described above, some background material must be discussed. This is detailed in chapter 2 and includes previous approaches to defining living systems, a summary of Artificial Life approaches and a summary of previously developed Artificial Chemistries.

In chapter 3 preliminary work is described where an established Artificial

Chemistry was re-implemented. This system is then critiqued with both outstanding issues and promising features identified.

Building on the preliminary investigation, chapter 4 discusses potential properties of Artificial Chemistries and ingredients Artificial Chemistries could be made from. Chapter 4 concludes by suggesting ingredients that might give desired properties.

The remainder of this thesis describes a novel kind of Artificial Chemistry, a sub-symbolic Artificial Chemistry named RBN-World. Chapter 5 describes the concept of a sub-symbolic Artificial Chemistry in general and chapter 6 describes RBN-World specifically. This is then followed up in chapter 7 where variations are compared in respect of some of the properties described in chapter 4. Chapter 8 then looks in detail within some of the best variations of RBN-World as identified in chapter 7.

Finally, chapter 9 summarises the outcome of the research and places sub-symbolic Artificial Chemistries, including RBN-World, within the wider context of Artificial Life.

## 1.3 Contribution

This work makes several contributions to Artificial Chemistry as well as Artificial Life. This includes the description of a novel kind of Artificial Chemistry that has interesting properties. In addition, this work contains a unified and clarified collection of definitions for properties of Artificial Chemistries as well as comparisons between them. Using this collection of properties, this work also describes a systematic approach to Artificial Chemistries by searching for the emergence of desired properties.

# Chapter 2

# Background

In order to address issues concerning Artificial Life through approaches using Artificial Chemistries, a number of background concepts need to be identified and defined. In particular, this topic has had a wide range of ideas and assumptions made that must be described so that a common reference can be established.

## 2.1 Life

Biology is the branch of natural science that studies living organisms; however, there is no strict definition of 'living'. A traditional reductionist approach is to subdivide the property of 'living' into multiple properties, and when something has all of those properties it is 'alive'. Such methods to separate 'living' from 'non-living' work for a large number of types of entities that people generally categorise into 'living' and 'non-living'. For example, a list of living things might include humans and other large mammals such as cats, dogs and elephants, as well as other vertebrates and the wider chordates including fish, birds, lizards, etc. Invertebrates such as lobsters, mosquitoes, and spiders also feature on such lists as examples of "living". Plants, fungi, corals and other static organisms can also be included, with suitable sets of properties. For example, if locomotion — movement of the complete organism from location to location under its own propulsion — is a property of 'living' then trees would not be classified as 'living'; as it would be difficult to accept a classification of trees as 'non-living', it is generally accepted that movement of some part is sufficient to qualify as 'living' rather than locomotion of the whole.

One of the traditional sets of these properties is shown in table 2.1.1 along with a modern re-interpretation (Koshland, 2002). The connection between some of the properties in those tables is obvious — "Metabolism" and "Energy", "Organization" and "Compartmentalization" for example. In some cases, multiple traditional properties have been combined into fewer modern properties; both "Growth" and "Reproduction" can be seen as "Regeneration" in some contexts, "Movement", "Response", and "Homeostasis" can be summarised as Adaptability. Modern neo-Darwinian theories suggests that the role of evolution is important, and this is reflected in the property "Program" described in Koshland (2002).

By similar steps and arguments the traditionally perceived set of properties was decided upon, and this is how 'living' has typically been defined for most of the 20$^{\text{th}}$ Century and earlier. However, as knowledge of biology has expanded, these sets of properties have come under more scrutiny. Micro-organisms have caused much of this pressure, in particular viruses. Some viruses are composed very simply — particularly compared to larger creatures — and there has been some difficulty in defining them as 'living'. Yet it can be argued that they fulfil all of the criteria and therefore either viruses are 'alive' or the properties should be revised. With the discovery of prions, which share a number of functional characteristics with viruses yet are even simpler, the issue has become even more confused.

With the discovery of DNA as the physical basis of inheritance, it might have been thought that 'living' could be defined as being DNA-based. However, there are viruses that are based on RNA rather than DNA yet have very similar characteristics to DNA-based viruses. Prions are not based on either RNA nor DNA but are protein-based instead.

Other advances in biology have lead to arguments about what constitutes an organism. It is accepted that for multi-cellular creatures such as humans, it is the combined collective that is the organism and a separated part is not. However, some multi-cellular structures can separate and re-form, for example slime-moulds such as *Dictyostelium discoideum*. Social organisms such as ants and wasps pose further problems to this issue. In those cases each separate individual typically does not reproduce nor can they survive for long when separated from the collective. Does this mean that an ants nest as a whole is an organism, rather than any one individual ant? And if so, is the ants nest 'alive' due to its own properties, or is an ants nest 'alive' because its

| Traditional | Koshland (2002) |
|---|---|
| Homeostasis | Improvisation |
| Organization | Compartmentalization |
| Metabolism | Energy |
| Growth | Regeneration |
| Movement | Seclusion |
| Response | Adaptability |
| Reproduction | Program |

Table 2.1.1: Some sets of properties that have been suggested to define living systems.

components are 'alive', or is a nest of ants not 'alive'?

As an alternative viewpoint, several 'non-living' systems have been held up as exhibiting the sets of properties though of as required for being 'alive'. Fire, for example, consumes fuel, produces waste, responds to its environment, moves, grows and can reproduce. As engineering and robotics has advanced, the distinctions between natural and artificial have become less clear-cut. Building on the impact of these advances in day-to-day life, some works of science fiction can be seen as thought experiments which test the boundary between 'living' and 'non-living' further in popular culture.

It should also be noted that 'alive' is typically regarded as a binary property — either an entity is 'alive' or it is not. However, because of the difficulties described above with the application of a binary property, there is an alternative viewpoint that proposes that the 'alive-ness' of entities is a continuum with some number of dimensions (one being the simplest). Using such a scale, a rock can be said to have almost no 'alive-ness', a fire has some 'alive-ness', a bacterium more 'alive-ness' and a sentient multi-cellular organism — such as a human — to have a very high 'alive-ness'. Furthermore, on a continuum an individual may change 'alive-ness' over time. This makes it possible for dependant infants to have a lower 'alive-ness' than their parents for example, or for 'alive-ness' to decrease over time upon physical death.

A less reductionist approach to defining 'life' is by assembling a range of examples of both 'living' and 'non-living' systems with a general consensus about which category each example belongs to. There is a problem with this as a method for defining what is and is not 'living' — all of the examples from the 'living' category are most likely related by common descent from the Last

Universal Ancestor (LUA)[a](Theobald, 2010). This means that the examples in the 'living' category are not independent; features that are shared due to common ancestry cannot be separated from features shared due to being in the 'living' category.

There are two options for sources of 'living' systems independent from LUA - either to find them in existing environments or to create suitable environments where they can form. The discovery of 'living' systems independent from LUA on Earth is highly unlikely; even extreme environments such as deep-sea vents and Antarctic desert rocks are contaminated with LUA-descended life. Discovering 'living' systems not on Earth (xenobiology) is a very challenging task given the vast distances involved, and one that does not look to be solved in the near future, despite recent advancements (Vogt et al., 2010).

This leaves the formation of new 'living' systems as a method for obtaining examples of life unrelated to LUA, and hence to better characterise 'living'. It may be the case that this is a Sisyphean task — as we have only a single example, it is unknown how common 'living' systems are even in environments with the highest frequency of formation. However, even failing to create novel 'living' systems will generate valuable information towards answering the question "what is life?".

Creating a novel living system is not an easy prospect. It presents something of a chicken-and-egg problem; there is only one type of living system that we know about to base the design of a novel one upon, but if it is based upon known living systems then it is not novel. One way to approach this problem is not to base a novel living system on existing ones, but rather to go back a stage further to the environment that life emerged from. It is accepted that in the early history of the Earth, there were no living organisms of any kind, the planet coalescing from gas and dust while also undergoing bombardment from other rocks and proto-planets. Looking at the Earth today, it is obvious that there is a wide range of living organisms, and from evidence such as the fossil record, it is clear that such diversity of living organisms has varied over time. Therefore, there must have been a transition between the life-less time of the early Earth into a state where clearly recognizable living organisms had arisen. This may have been a single event if 'alive' is viewed as a binary property, or it may have been a gradual process if 'alive-ness' is thought to be a continuous property. It is assumed that if the essential characteristics of the transition

[a]a.k.a Last Universal Common Ancestor (LUCA) or cenancestor.

can be recreated then novel living systems may emerge in the same manner.

The concept of creating novel living systems is not in itself new. The field of Artificial Life has looked to address this issue from both scientific and engineering applications. However, much Artificial Life work starts from components that already exhibit living properties (such as reproduction) rather than emerging from a previously non-living environment. The field of Artificial Chemistry arose to address this by starting from non-living components and trying to enable the emergence of living systems.

## 2.2 Introduction to Artificial Life

The topic of Artificial Life is at the same time both young and old. The term "Artificial Life" was coined by Langton in 1986 when he said:

> "Artificial Life is not life-as-it-is, but life-as-it-could-be" (Langton, 1986)

However, the concept of artificial living creatures is found throughout human history. The ancient Greeks spoke of the Colossus of Rhodes as a metal giant defending the harbour of that city from invaders by throwing boulders at attacking ships to sink them. Jewish mythology includes tales of golems; living beings crafted from inanimate materials.

The first recorded example of Artificial Life in the modern era was a mechanical duck constructed by a Frenchman Jacques de Vaucanson in 1739 (Riskin, 2003). This shocked audiences with its life-like behaviours; moving, eating, and even excreting. However, it was incapable of other behaviours expected of a duck such as flight or mating. Artificial Life work based on imitating known living systems continues to be pursued, and the current state-of-the art in this area are human androids of Nishio (2007) and other similar work. Related to this are other robotic approaches based on duplicating known living being, such as Asimo (Sakagami and Watanabe, 2002). With the rise of computing in the late $20^{\text{th}}$ century, much work was put into imitating living systems, primarily for artistic purposes in entertainment media (film, television, games, etc); an example of this approach is L-systems (Lindenmayer, 1968a,b; Prusinkiewicz et al., 1990). Although work along these approaches has produced believable simulacra, it is a common philosophical position that imitation cannot replicate the original. As a painting does not capture all

of the scene, technological improvements of photography and video have not changed this — though when experiencing these technologies for the first time, people may temporarily believe otherwise. Therefore, although work of this type must be acknowledged, it is not Artificial Life in the sense used in this thesis. If a perfect replica could be constructed then it might be classed as living — but such a perfect replica of a natural system could not be considered "artificial".

Related to the imitation approach described above, there has been a large amount of work dedicated to abstracting from known biological systems in order to create models. These approaches are typically mathematical in nature, often assisted by computational methods as they developed. Some examples of these include Turing Patterns (Turing, 1952) of cellular signalling and differentiation; ecological models such as predator-prey interactions (Lotka, 1925) and gene frequency models (Hardy, 1908); and there are many, many more. Whilst models can be an accurate representation of the biological domain being investigated, in the context of Artificial Life used in this thesis almost all models are not Artificial Life for the same reasons as the simulacra described above - they are replicates of a living system rather than living system themselves. A model that would completely replicate a real living system would be such a perfect replicate that it would not be "artificial" even if it was "life".

With the increase in computational power available to researchers, some attention turned towards virtual environments. In contrast to models, virtual environments are characterised by exploring how different situations share underlying features rather than attempting to replicate the real world within a computer. A classic example of a virtual system is Conway's Game of Life (Gardner, 1970) as well as Langton's Loops (Langton, 1984) and other cellular automata. In such work, a wide range of different systems with different rules and/or initial conditions have been shown to exhibit similar features of emergent and complex behaviours. This has caused such virtual investigations to attract a large amount of interest. The down-side of work on a virtual system is that results and conclusions from virtual environments are not directly applicable to reality. This is because there is no specific mapping between the real and virtual, similar to the mapping that is key to modelling approaches.

In addition to the lack of a specific mapping, there is a further problem with virtual environments. Many people adopt a philosophical position, either

explicitly or implicitly, that virtual environments are intrinsically incapable of supporting a life-like organism of any kind. The reasons for taking this position are varied, but have some common points. The major issues are that computers are too simple to support a virtual environment capable of containing living systems, or that such a virtual environment is too complicated and/or difficult for a computer to process, or for a person to program. The continuous nature of reality and the digital representation within a computer are also often given as reasons why such virtual environments are not capable of supporting life. There is little evidence to support either a pro- or anti-position; modern computing has been in existence for less than a century and shows no sign of slowing its exponential growth in capacity, whilst at the same time particle physics has demonstrated how discrete and unintuitive the universe is on the smallest scales. Although it is true that there has been no clear-cut demonstration of a virtual environment containing artificial living beings, that lack of evidence does not mean that such systems are impossible — particularly given the relative youth of computer technologies.

However, these drawbacks have not stopped studies of virtual environments feeding back into real-world systems. A common approach is to try to implement some of the principles gained by studying virtual environments in physical systems such as robots. These approaches encounter constraints enforced by the embodiment that were avoided / ignored in the virtual system. The continuous "fuzzy" nature of three-dimensional objects compared to discrete two-dimensional grids often used in virtual environments is an example of this. Resolving such constraints forces systems to be less abstract 'virtual environments' and more 'computational models of reality'. This changes them from Artificial Life into replicating known living systems and the problems outlined above for mimics and models apply.

Related to, but independent from, "Artificial Life" is xenobiology (McKay, 2004) — life on other planets. Although there has been no direct evidence for extra-terrestrial living organisms, significant effort is being expended searching for such. This is challenged by some of the same issues as "Artificial Life", especially how to recognise living organisms. However, this can be partly addressed by focusing on forms of xenobiology similar to that found on Earth — carbon-based nucleic- and amino-acids — at the risk of missing other, more different, life forms. Unlike "Artificial Life" however, xenobiology is typically pursued for its own goals and not for other purposes such as robotics, psychology, or

artificial intelligence. This means that although xenobiology is less widely pursued, those studying it often do so with aims and interests that are more homogeneous than the Artificial Life community. The higher costs and limited access to suitable equipment, such as space-borne telescopes, reinforce this uniformity.

Outside of Artificial Life, two other topics are of particular relevance to this work: "complexity theory", and "emergence".

The behaviour of some systems are random, or pseudo-random; such as radioactive decay or rolls of a dice. Other systems are deterministic; knowing the initial state and how it changes over time, future states can be predicted. Some deterministic systems are chaotic; predictions are only accurate for the near future. In chaotic systems errors in measurements of the initial state grow exponentially, rapidly degrading any potential predictive power. Contrasting with chaotic systems are ordered systems where errors do not grow exponentially and the state of the system can be accurately predicted over long time-scales. Some systems fall between ordered and chaotic, and are said to be at the edge-of-chaos. Complex systems are neither ordered or chaotic, but have similarities to both. Thus complex systems are often edge-of-chaos systems.

The most famous example of a complex system is the Lorenz attractor (Lorenz, 1963). This is a simple set of three differential equations that, when evaluated deterministically, results in a trajectory that converges to a two-lobed pattern. It can be easily predicted that at some future point the state of the system will be in one or other of those lobes, but predicting where on the lobes the system is — or even which lobe the system is on — rapidly becomes difficult. Similar sets of differential equations exist for ordered systems where the state converges on a point or cycle, and for chaotic systems where the state diverges towards infinity.

It is important to note that 'complex' and 'complicated' have distinct meanings. Complicated systems have many different components, and while a complex system may be complicated (e.g. stock market) it does not have to be. Conversely, a 'simple' system can have complex behaviour.

The other topic relevant to Artificial Life is 'emergence'. This has already been mentioned as a desirable characteristic for Artificial Chemistries but, like life, emergence is difficult to define. Emergence is usually thought of as a collection of entities that exhibit a combined behaviour or feature that is

not part of any subset of those composing units — effectively forming a new kind of entity in the process. The objects composing the emergent system are typically of only one (or a handful) of types that are present in large numbers. The classical example of emergence is a flock of birds; each individual bird moves by itself, but together they form a flock that changes shape and size as if it was a single entity. Many examples of emergence are biological in nature — ants, cities, flocks/shoals/herds, etc. — but there are some non-biological ones as well such as sand-dunes. As with complexity, it is unclear if emergence is a requirement for life, or a consequence of being alive.

## 2.3 Introduction to Artificial Chemistry

Artificial Chemistry originated as a sub-topic of Artificial Life. In particular, Artificial Chemistry was originally seeking to investigate the origin of life and the transition from an abiotic to biotic environment (abiogenesis). However, since then ideas from Artificial Chemistry have been used on other topics such as computational search (Yamamoto and Banzhaf, 2010).

Artificial Chemistry for Artificial Life tests the hypothesis that life is an emergent property of lower-level chemistry and physics rather than a distinct and separable phenomenon. Therefore, wherever a suitable chemistry exists, life may emerge. This includes chemistries that are not natural, e.g. virtual computer environments.

The most widely-used definition of Artificial chemistry was given in Dittrich et al. (2001):

> "an artificial chemistry can be defined by a triple ($S$,$R$,$A$), where:
>
> - $S$ is the set of all possible molecules,
> - $R$ is a set of [reactions] representing the interaction among the molecules, and
> - $A$ is an algorithm describing ... how the [reactions] are applied to the molecules"
>
> (Dittrich et al., 2001)[b]

[b] Formatted for clarity.

This definition includes several terms that need further explanation and description. As it is analogous to chemistry, Artificial Chemistry has adopted much terminology from chemistry. Unlike real chemistry however, Artificial Chemistries may violate the fundamental assumptions underlying those same terms.

### 2.3.1 Molecules

The "set of all possible molecules" in the quote above presents the first issue. It is useful to define the term 'molecular species' to describe a unique molecular structure and reserve 'molecule' for a particular instance of an 'molecular species'. This is analogous to the distinction between an individual organism and the species that organism is classified as. Using the term 'molecular species' the above quote is equivalent to "$S$ is all possible molecular species", which makes it clear that no two molecular species are the same (though there may be many equivalent molecules of a single molecular species) and that each molecular species may have a number of individual molecules at any particular point in space and/or time ranging from none to many.

In some Artificial Chemistries, "all possible molecular species" is simple to describe as the chemistry is defined with a finite — if large — number of molecular species. An example is binary strings of a particular length (Banzhaf, 1993a,b). Some chemistries have an unbounded number of molecular species, such as strings of any length (Hickinbotham et al., 2011) or tree structures (Fontana, 1991). For such structural representations of molecular species, the terms 'element' and 'atom' may be appropriate to use to described the components. As with the terms 'molecular species' and 'molecule', the 'element' is the the unique type of every 'atom'.

The term 'atom' applies to molecules which cannot be decomposed into multiple molecules and cannot be composed from multiple molecules. A particular element may exist in multiple states and be able to undergo transitions between those states, but atoms of one element cannot change to or from another element[c]. These are definitions based on function rather than *a priori* statements, and this reflects how atoms and elements were historically identified. These definitions may conflict with the usage of these terms in other

[c]Nuclear fusion and fission enable changes of element but are outside the scope of this thesis.

works.

To construct molecules from 'atoms', some form of 'bond' is needed. The exact details depend on the chemistry, but it is useful to consider molecules as a graph where the 'atoms' are nodes and 'bonds' are edges. There may be additional constraints for particular Artificial Chemistries — a string can be thought of as a directed acyclic graph where every node has a maximum in-degree of 1 and a maximum out-degree of 1. Given a set of 'elements' to use for 'atoms' and 'bonds' to connect them, "all possible molecular species" can be defined. Note that this may be an unbounded number of molecules and so cannot be explicitly enumerated.

### 2.3.2 Reactions

The second component of an Artificial Chemistry is the "set of reactions representing interaction among the molecules". Each reaction is composed of two collections of molecules, 'reactants' and 'products'. A reaction represents the molecules of the reactants interacting (colliding) and becoming transformed into the products. Typically reactants are pairs of molecules, though Artificial Chemistries do not have to follow this. Products may be a fixed number of molecules, e.g. 2 or 3, or the products may be varying numbers of molecules for different reactions. The same molecular species may be represented more than once within the products or the reactants. If the reactants are the same as the products in a reaction then that reaction is said to be 'elastic'.

Reactions may be specified in different terms to those specified here. For example, Banzhaf (1993a) described an Artificial Chemistry in terms of enzyme synthesizing proteins from a template as well as proteins decaying. Most Artificial Chemistry can be translated into the framework described above — and of those Artificial Chemistries that cannot be mapped to this framework, it is questionable if they are indeed an Artificial Chemistry (Sayama, 2009). For the example above (Banzhaf, 1993a) the reactions can be mapped to the framework described here by having either reactions of two reactants and three products or reactions of one reactant and no products.

Some Artificial Chemistries explicitly list all reactions in a reaction table. Typically these Artificial Chemistries are hand-crafted after much trial-and-error, or based on information about real chemical processes. In other chemistries, reactions are generated by applying an algorithm to a set of reactants

to generate a set of products e.g. Banzhaf (1993a). If the number of molecular species is unbounded, then the set of reactions may also be unbounded. As each reaction is a transformation from a set of reactants to a set of products, a collection of reactions can be thought of as a type of graph. For example, a graph can be created by considering each molecular species or reaction as a node, and edges are drawn to a reaction from each of its reactants and out from a reaction to each of its products. Such a bipartite graph is also found in Petri Nets (Peterson, 1981).

### 2.3.3 Mixing

The final component of an Artificial Chemistry is an "an algorithm describing ... how the [reactions] are applied to the molecules" — known as 'mixing'. Natural chemistry is thought to be continuous three-dimensional spatial structure with continuous time[d]. Most Artificial Chemistries use a simpler spatial structure, commonly a well-mixed container (Banzhaf, 1993a,b; Fontana, 1991) or a two-dimensional space that can be discrete (Hutton, 2004b) or continuous (Hutton, 2007). These forms of space can be combined with either discrete time (Banzhaf, 1993a,b; Fontana, 1991; Hutton, 2004a) or continuous time (Hutton, 2007).

The number of times a reaction occurs within a unit of time is called the 'reaction rate'. Drawing on real chemistry, the reaction rate in well-mixed Artificial Chemistries is usually proportional to the number of times the collection of reactants can be constructed from the molecules present at that time ('concentration' of the reactants). For some algorithms, a 'rate constant' can be associated with each reaction. The 'rate constant' describes how the 'reaction rate' for different reactions varies given the same concentrations. Such rate constants are usually less than 1.0 and represent conformational and/or energetic requirements.

## 2.4 Conclusion

This chapter has described Artificial Chemistry in the context of Artificial Life, as well as some of the motivations for research in this field. Furthermore, the

[d] Ignoring quantum effects

basic concepts in Artificial Chemistry have been described as well as related topics relevant to this work.

The next step is to determine what the state-of-the-art is in Artificial Chemistry with regards to Artificial Life, and how this could be improved upon. This can then be used as a starting point for the exploration of novel Artificial Chemistries and their potential for Artificial Life.

# Chapter 3

# Preliminary Investigation

So far Artificial Chemistries have been discussed in the context of Artificial Life and some of the essential terms have been described. In order to provide a solid context for further discussion, a re-implementation of an existing Artificial Chemistry is described in this chapter.

Artificial Chemistries are studied from multiple different points of view and therefore it is useful to provide a baseline for comparison. This chapter describes a previously published Artificial Chemistry that was reconstructed in order to investigate the features and issues around Artificial Chemistries in general. This is motivated by a lack of developmental context to published Artificial Chemistries — often only a definitive description is published without exploration of alternatives.

The Artificial Chemistry chosen to re-implement is described in Hutton (2004b, 2007) and was selected for several reasons. Firstly, this Artificial Chemistry was created from a biological Artificial Life perspective in contrast with other chemistry such Binary String Chemistry (Banzhaf, 1993a,b,c, 1994) or Fontana's Algorithmic Chemistry (Fontana, 1991) that have a more mathematical approach. Secondly, it claims to demonstrate a replicating 'cell' as well as 'enzymes' encoded on a 'genome', which is much closer to accepted biological examples of life than other Artificial Chemistries.

The first part of this chapter describes Hutton's Artificial Chemistry using the terminology already introduced. After that the problems that were encountered reconstructing Hutton's Artificial Chemistry are discussed, followed by drawing some conclusions about this Artificial Chemistry as well as issues with previously published Artificial Chemistries in general.

## 3.1 Description of Hutton's Artificial Chemistry

The molecules in Hutton's Artificial Chemistry are composed of atoms connected by bonds. Each atom is a representation consisting of a letter and a non-negative integer, corresponding to the element and state of the atom respectively. Atoms can be changed to other states, but can never change element. Bonds can exist between any pair of atoms and connected groups of atoms form molecules. There are no explicit *a priori* limits on the number of atoms, number of distinct elements, number of atom states, or the number of bonds an atom may be part of (including allowing multiple bonds between the same pair of atoms).

There is an explicit list of reactions that each describe reactants of either 2 or 3 atoms and products of 2 or 3 atoms. Both reactants and products consist of 2 or 3 atoms of specific elements and in specific states; a reaction can change the state of an atom but cannot change the element nor create or destroy atoms. This ensures that mass is conserved within the Artificial Chemistry. The reactants and products also specify if a bond exists between the atoms before and/or after the reaction. Most reactions can only occur in a single direction; however, for two cases the reaction is reversible (i.e. the products can act as reactants and *vise versa*).

Hutton's Artificial Chemistry 'mixing' component is a space that has two dimensions and is either a discrete grid (Hutton, 2004b) or a smooth continuous space (Hutton, 2007). The movement of atoms is different in the two kinds of space. In discrete space, atoms are updated in a random order and move randomly into unoccupied sites that have a common edge with the current site (Hutton, 2007); i.e. only one step vertically or horizontally and not diagonally — this is commonly known as the Moore neighbourhood. This is further restricted by the limitation that it must stay within an adjacent site (this may include diagonally) to any atoms it is bonded to and movement cannot cause any bonds to cross. In the continuous space version, the position of atoms is updated by applying velocity vector that is changed by various forces acting upon the atoms, including bonds acting as springs between atoms and a "volume exclusion force" (Hutton, 2004b) between atoms.

The description above is for the general system, but does not describe the specific details that were used. In particular, an explicit list of reactions and

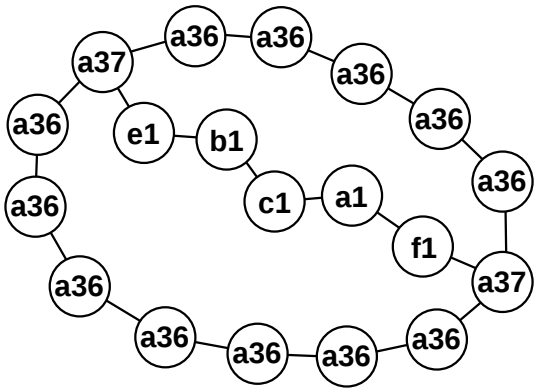



Figure 3.1.1: Initial cell configuration used by Hutton (2007) (reproduction of Figure 2).

an initial condition are required — see table 3.1.1 and figure 3.1.1 respectively for details. In the reaction table, several wild-card expressions are used (i.e. x & y); these could be removed by expanding all possible combinations they represent into single reactions. For example, reaction #6 is '$x5 + x0 \rightarrow x6$ — $x6$' but this represents the following 6 separate reactions:

| | | | |
|---|---|---|---|
| a 5 | a 0 | → | a 6—a 6 |
| b 5 | b 0 | → | b 6—b 6 |
| c 5 | c 0 | → | c 6—c 6 |
| d 5 | d 0 | → | d 6—d 6 |
| e 5 | e 0 | → | e 6—e 6 |
| f 5 | f 0 | → | f 6—f 6 |

In the same manner, each of the reversible reactions (#35 & #41) can be represented as two reactions with swapped reactants/products.

After initial work (Hutton, 2004b), Hutton extended the reaction table to enable enzymes and mutation. Enzymes are atoms with a state value above a specific threshold (38) which is interpreted as a reaction that can only occur when the enzyme is "present". Reactions #36-39 in table 3.1.1 control the production of enzymes in a manner inspired by transcription and translation. Reaction #40 in table 3.1.1 described how the state value of an enzyme encodes a reaction and, like other wild-card containing reactions, could be expanded into a large number of combinations. Mutation is defined by a reversible reaction (#41 in table 3.1.1) that has a low probability of occurring.

| # | Reactants | | Products |
|---|---|---|---|
| 1: | e 1—a37 | → | e 5—a10 |
| 2: | a10 e 6 | → | a37—e 3 |
| 3: | e 6—e 3 | → | e 2—e 3 |
| 4: | *x* 2—*y* 1 | → | *x* 7—*y* 4 |
| 5: | *x* 4 *y* 3 | → | *x* 5—*y* 7 |
| 6: | *x* 5 *x* 0 | → | *x* 6—*x* 6 |
| 7: | *x* 6 *y* 7 | → | *x* 3—*y* 4 |
| 8: | *x* 6—*y* 4 | → | *x* 1 *y* 2 |
| 9: | *x* 7—*y* 1 | → | *x* 2—*y* 2 |
| 10: | f 2—a37 | → | f 9—a11 |
| 11: | a11 f 3 | → | a11—f 9 |
| 12: | *x* 2—*y* 8 | → | *x* 9—*y* 1 |
| 13: | *x* 9—*y* 9 | → | *x* 8 *y* 8 |
| 14: | a11—a36 | → | a11—a12 |
| 15: | f 1 a12 | → | f13—a37 |
| 16: | *x*13—*y*01 | → | *x*14—*y*15 |
| 17: | a11 *x*15 | → | a11—*x*16 |
| 18: | *x* 4—*y*16 | → | *x*27—*y*16 |
| 19: | *x*27—a11 | → | *x*17 a11 |
| 20: | *x*17—*y*16 | → | *x*17—*y*13 |
| 21: | *x*13—e08 | → | *x*14—e15 |
| 22: | e13—a37 | → | e18—a19 |
| 23: | e13—a19 | → | e18—a20 |
| 24: | a20 a11 | → | a21—a22 |
| 25: | e18—a22 | → | e32 a23 |
| 26: | e18—a21 | → | e32 a24 |
| 27: | a24 a37 | → | a26—a27 |
| 28: | a27—a23 | → | a37 a28 |
| 29: | a26—a36 | → | a29—a30 |
| 30: | a29—a36 | → | a31—a30 |

| # | Reactants | | Products |
|---|---|---|---|
| 31: | a30 a28 | → | a25—a33 |
| 32: | a31—a25 | → | a32 a36 |
| 33: | a32—a30 | → | a34—a36 |
| 34: | a34—a33 | → | a37 a37 |
| 35: | a(i) — a(j)<br>\ /<br>a 0 | ↔ | a(i) a(j)<br>\ /<br>a36 |
| 36: | x32 z 0<br>\<br>y17 | → | x 1 z38<br>\ /<br>y35 |
| 37: | x35 — z(i)<br>\<br>y17 | → | x 1 z(j)<br>\ /<br>y35 |
| 38: | x(i) — f35<br>\<br>y 0 | → | x(i) f32<br>\<br>y(i) |
| 39: | f32—a37 | → | f 1—a37 |
| 40: | z(i) x(g)<br>b1/<br>y(h) | → | z(i) x(j)<br>b2/<br>y(k) |
| 41: | x 1 z 0<br>\<br>y 1 | ↔ | x 1 — z 1<br>/<br>y 1 |

Table 3.1.1: Reaction rules used by Hutton (2007). Letters a-f represent elements and the letters $x$, $y$ & $z$ are used as variables representing any element. Numbers represent the state of atoms and g-k are used as variables for any state. See text for details.

## 3.2 Implementation Challenges

In order to better understand Artificial Chemistries and the issues around them, replication of Hutton's work was attempted. There were a number of challenges associated with this in three categories — information missing from published material, fragility of Hutton's Artificial Chemistry to changes, and computational requirements to implement the chemistry.

As it has been developed over a number of years, the details of Hutton's Artificial Chemistry is split over multiple papers and have undergone various revisions which has resulted in inconsistent terminology. While (Hutton, 2007) contains sufficient detail to understand the reactions and other chemical properties, the information on the implementation of space used is less complete. For example, Hutton states the following to describe the continuous space physics:

> "Each atom has floating-point coordinates, and a system-wide *radius* value determines their size. Each atom has a *velocity* that in the absence of interactions will determine its next position. At each update step the velocities are first recomputed, by summing the forces that are acting on the atom, and then the positions of the atoms are updated. Forces come from overlap with other cells or the sides of the area, and from bond distances being greater than twice the radius. In each case the force is either proportional or inversely-proportional to the overstretch/overlap distance, with some user-determined factor $k$. While this form of motion force integration is unstable when large forces are involved (if two atoms get pushed together), simple measures such as limiting the velocity can prevent numerical problems. To permit small particles to diffuse through the membrane, we simply turn off the cell-overlap force for unbonded atoms. Atoms can react when they are overlapping, since this represents a collision."(Hutton, 2007)[a]

However, values for $k$, atom radius or maximum velocity are not mentioned in any published material and the source code is not publicly available. Therefore, when re-implementing Hutton's work these values had to be estimated by trial-and-error.

[a] In this quote, the term "cell-overlap force" is equivalent to the "volume exclusion force" mentioned previously.

These values then highlighted the second group of challenges encountered in this project — fragility of Hutton's Artificial Chemistry to changes. For example, one common deviation from the expected behaviour of the Artificial Chemistry was for cells to invert and eject their genome to the outside of the membrane. These genomes would still continue to interact, but this inversion event is not part of the behaviour Hutton reported and would be fatal to the cells if this was a biological system. Another common deviation observed was the entangling of the replicated genome, which resulted in an inability of the cell to proceed to division of the membrane. The underlying cause of both of these was an imbalance between the attraction of bonded atoms and the repulsion when overlapping. In addition to fragility in the physics, the reaction table is also sensitive to errors. If just one reaction is not correct, then the entire process does not proceed correctly.

The other group of challenges relate to the computational requirements of such a simulation. The re-implementation was based on the continuous space version, however, this has a large computational effort associated with it, for example, calculating the "volume exclusion force" is $O(n^2)$ and detecting triples of overlapping atoms that could potentially react is $O(n^3)$ where $n$ is the number of atoms. Although these are for naïve implementations, it illustrates that a large amount of computational effort is spent on things other than the reactions of the chemistry and this effort will increase as the size of the system is increased.

## 3.3 Conclusions

Hutton's Artificial Chemistry was, and still is, the only Artificial Chemistry that has demonstrated an emergent self-replicating organism without many special cases or other additional modifications. As discussed previously, all of the wild-card reactions used could be expanded into multiple non-wild-card reactions including those for enzymes, membranes, and mutations. Additionally, reactions #1-34 (those involving pairs of atoms) can be catalysed by enzymes which can be encoded on the genome of a cell.

However, despite its advantages and features, there are several issues with Hutton's Artificial Chemistry. Firstly, the system has large computational demands. This means that scaling up to run populations that could potentially interact with rich dynamics is not feasible. A single cell that encodes enzymes

for all required reactions would take days to complete a single division, for example. The major limiting factor appears to be the frequency with with suitable materials collide, especially for the three-way interactions such as catalysis. One aspect of the Artificial Chemistry that may restrict these interactions is that the genome partitions the cell into two regions (see figure 3.1.1). This could be solved by using a three-dimensional space which would allow for greater mixing of material within the cell, but would entail a higher computational cost per unit of time. This would be exaggerated by the larger spherical membrane which would be required for a three-dimensional cell.

Another issue with Hutton's Artificial Chemistry is that the system has difficulty interacting with novel materials, such as a previously unseen elements or states. This is because all reactions are listed explicitly, so if there is no reaction concerning the novel material (and if there was then it would not be novel) it does not react and therefore the presence of novel materials does not change the Artificial Chemistry. Furthermore, because enzyme activity is encoded in the state of the atom, enzymes cannot participate in reactions other than their enzymatic activity after they are created. This limits the behaviour that can evolve; for example, it is not possible to produce enzymes that are inactive when produced but become active after being exported from the cell. This fixed enzyme encoded also prevents them from interacting with novel materials.

An issue related to this is that Hutton's Artificial Chemistry does not have the ability to have varying environments, such as high or low temperature. Some of the parameters could have different values at different regions of space, but the high sensitivity of the system to particular parameter values may not allow this to be effective. This could limit the potential for evolution by enabling a single "perfect" cell to dominate at all viable parameter values.

There are further questions over the evolvability of the cells in this system. Enzymes are encoded onto the genome using a simple scheme, and mutation can only add or remove bases from the genome. This means that the fitness landscape may have significant obstacles to a smooth evolutionary path. Additionally, there is no scope for the 'evolution of evolvability' because the behaviour of enzymes themselves are 'hard-coded' into the Artificial Chemistry via reactions #36-39. As these reactions have three reactants (four with a catalyst), they cannot be encoded into the state of the enzyme in the same manner as other reactions.

#### 3.3.0.1 Revised aims

Due to the issues described above, it was decided that the re-implementation of Hutton's Artificial Chemistry should not be developed further. Instead, the desired ingredients and properties of an ideal Artificial Chemistry should be investigated and a novel approach determined from that basis.

In particular, there are two common issues with Hutton's Artificial Chemistry that should be addressed. Firstly, because Hutton's Artificial Chemistry is designed in a 'top-down' manner, it cannot incorporate or develop novelty and is highly tuned. It is assumed that this issue can be addressed by developing a system that emerges from underlying properties in a 'bottom-up' direction. However, designing emergent systems with specified properties is challenging. The second issue with Hutton's Artificial Chemistry is that due to the spatial representations there are high computational demands for aspects other than the reactions. Both of these issues are addressed in this thesis.

# Chapter 4

# Ingredients and properties of Artificial Chemistry

The previous chapter describes in detail an existing Artificial Chemistry and a re-implementation of that chemistry. This highlights problems with both that specific Artificial Chemistry and issues with Artificial Chemistries in general.

Along with the various Artificial Chemistries that have been proposed over the years, there has been some work suggesting properties that are desirable and should be considered when designing an Artificial Chemistry (Suzuki et al., 2003; Hickinbotham et al., 2010) as well as work attempting to analyse and/or classify Artificial Chemistries based upon properties that emerge from them (Dittrich and di Fenizio, 2007).

This chapter defines and summarises these emergent properties. It also suggests which ingredients may lead to which properties. This is then used to inform the design of an Artificial Chemistry for Artificial Life.

## 4.1 Ingredients

As Dittrich et al. (2001) describes, there are three core components of an Artificial Chemistry — 'molecules', 'reactions', and 'mixing'. However, for each of these components many different ingredients have been suggested and/or used previously. In many cases the selection of one core feature is independent of the other core features; e.g. most of the different ingredients for 'mixing' are independent of most of the ingredients for 'molecules'. This results in a large degree of interchangeability between different ingredients of Artificial

Chemistries — e.g. the 'mixing' component of an Artificial Chemistry can often be replaced with the 'mixing' component of another Artificial Chemistry resulting in the creation of a third Artificial Chemistry that is distinct from the previous two.

To assist in the design of Artificial Chemistries, some ingredients are described here — organized by the three core components of Dittrich et al. (2001): 'molecules', 'reactions', and 'mixing'. It does not attempt to enumerate all possible ingredients, nor list all ingredients that have been used previously; rather this section seeks to describe some of the choices to be considered and what some of their consequences might be.

### 4.1.1 Molecules

Molecules are the most fundamental part of an Artificial Chemistry. This is reflected in the multitude of different representations. In general, most of the molecular representations used in Artificial Chemistries can be placed into one of two types — symbolic or structured — though the original authors may not have described them in that terminology.

#### 4.1.1.1 Symbolic

In a symbolic representation of molecules, each molecular species is assigned a unique symbol (such as a word or number) and there is no internal structure or additional information contained within that symbol. This type of representation is typically used in Artificial Chemistries that model a well-described and closed real-world system, such as the photochemistry of planetary atmospheres (Yung and DeMore, 1998).

While such systems are easier to understand than other Artificial Chemistries, a symbolic molecular representation has a number of limitations. These limitation are due to the fact that molecules contain no information, and therefore it is not possible to use implicit reactions (see §4.1.2.2) which limits the possibilities of open-ended novelty and emergent behaviours. Whilst this may not be a problem for some applications of Artificial Chemistry, in the context of Artificial Chemistries for Artificial Life this is a critical flaw.

#### 4.1.1.2 Structured

This is the most common representation of molecules found in Artificial Chemistries. Each molecular species consists of one or more atoms arranged into a structure such as a string, matrix, tree or graph. The structures used can either be the same for all molecules — e.g. a string of fixed length (Banzhaf, 1993a) — or can be different for different molecules (potentially within certain bounds) such as strings with variable lengths (Banzhaf, 1994).

In most structured molecular representations used in Artificial Chemistries there are multiple types of atoms that can exist; each type is known as an element. Typically, the elements of an Artificial Chemistry are symbols of a finite set; e.g. $\{0, 1\}$ of a binary string (Banzhaf, 1993a) or the 'variable' and 'algorithmic' operators from AlChemy (Fontana, 1991). Hypothetically, an Artificial Chemistry could be constructed from a countably infinite set of elements such as the natural numbers ($\mathbb{N}$) or prime numbers ($\mathbb{P}$). Indeed, even an uncountable infinite set of elements such as the real numbers ($\mathbb{R}$) could be used in an Artificial Chemistry. A large number of possible elements may be a desirable feature in an Artificial Chemistry, as simple combinatorics increases the chance that a small subset of elements contains the desired properties.

In addition to atoms of elements, structured molecular representations contain bonds. These are associations between groups of atoms, typically pairs, within a molecular species. Bonds are used to delimit which atoms form a molecule — those atoms that form a connected component.

Structured molecular representations may have alternative representations that are more easily recorded; e.g. the parentheses characters in AlChemy (Fontana, 1991) are not considered to be symbolic atoms used in a string structure, but rather the variables and operators are symbolic atoms in a tree structure. A structured molecular representation can be expressed as a symbolic molecular representation, but not necessarily *visa versa*.

Compared to symbolic molecular representations, structured molecular representations allow for a richer set of molecules, but this is both an advantage and a disadvantage. On the one hand, a structured molecular representation allows for implicit reactions between molecules as well as the possibility of functional groups. But this richness can lead to a combinatoric explosion in the number of possible molecular species, which can subsequently cause computational problems for investigation and analysis. In addition, a structured molec-

ular representation is harder to visualise and discuss unless symbolic names are applied to the molecular species — analogous to real chemistry where a molecular species is called "caffeine" rather than the specific, but lengthy, "1,3,7-trimethyl-1$H$-purine-2,6(3$H$,7$H$)-dione3,7-dihydro-1,3,7-trimethyl-1H-purine-2,6-dione"[a].

#### 4.1.1.3 Sub-symbolic

In the process of creating this categorization of Artificial Chemistry ingredients, it became apparent that a significant gap exists in previous work on this topic. In natural chemistry, all of the properties of a molecule are emergent from the sub-atomic components; nuclei of protons and neutrons with precisely shaped cloud of electrons surrounding them. Indeed, these sub-atomic components are themselves an emergent consequence of still smaller components - quarks, bosons etc.

An molecular representation in Artificial Chemistry that captures this emergence from lower-level structures has been named "sub-symbolic". This draws upon analogy from artificial intelligence fields in the invention of neural networks (Smolensky, 1988). Some previous Artificial Chemistries have used symbolic molecules that are interpreted by the reaction algorithm in order to determine the products. However, these have never used an explicit separation of atomic and sub-atomic structures and therefore these would not be classed as sub-symbolic molecular representations.

Compared to the symbolic and structured molecular representation described previously, in a sub-symbolic molecular representation the 'symbols' are units with internal structure. This allows the molecular properties and implicit reactions to be emergent properties of the underlying units.

A sub-symbolic representation is analogous to the sub-atomic structures underlying biochemistry. However, unlike quantum-mechanical molecular dynamic simulations, in an Artificial Chemistry the sub-symbolic representation can be selected to exhibit desirable properties yet be computationally tractable.

As with a structured molecular representation, the additional richness of a sub-symbolic representation is a double-edged sword. The main advantage is that it enables molecules and functional groups to be more than the sum of their parts without much additional specification of the chemistry, but this potential

[a]International Union of Pure and Applied Chemistry (IUPAC) systematic name

emergence comes with a computational cost to calculate and depends upon using both a sub-symbolic representation and an implicit reaction algorithm that can take advantage of it.

#### 4.1.1.4 Intra-molecular space

When using a structured or sub-symbolic molecular representation, the relative positions of the atomic units within the molecular species is implied. This space can be of a number of different types, which have different properties and implications for Artificial Chemistries that use them.

The simplest type of intra-molecular space is an aspatial structure where all atoms are zero distance apart, such as sets or multi-sets. In such a representation, there can be no 'bonds' in the usual sense as every atom in a molecule is equally bonded to every other atom. This implies that each molecule is the sum of its atoms without any difference between alternative arrangements of those atoms (isomers).

Intra-molecular spaces can also be 1-, 2-, or 3- dimensions [b] of either continuous or discrete space. For example, a string based molecular representation can be though of as a 1-dimensional discrete intra-molecular space.

Another possibility for the structure of the intra-molecular space is a graph; each atom is a node in the graph and each bond is an edge. Here the distance between atoms is the graph distance between them. Whilst this may appear similar to conventional chemical structures, a strictly graph-based intra-molecular space has significant topological differences. For example, if three atoms are in a linear molecule with two bonds, the distance between the atoms at the ends is always the sum of the length of the bonds in a graph-based intra-molecular space. In a discrete or continuous space, the distance depends on the angle between the bonds.

The stick-and-ball model of conventional Chemistry implies a graph embedded in a continuous 3-dimensional euclidean space. This means that angles and relative position of molecules is important, and that atoms that are far apart when the molecule is viewed as a graph are not necessarily far apart in the intra-molecular space; this is a key feature of protein folding and other secondary- and tertiary- structures. However, such types of space can be computationally expensive to work with due to this phenomenon.

[b] Or more than 3-dimensions, but these are challenging to visualize and conceptualize, and therefore not considered here.

The list presented here does not claim to be an exhaustive list of every intra-molecular spatial representation possible. However, the ones outlined in this thesis have been previously used in Artificial Chemistries.

#### 4.1.1.5 Molecular states

In conventional Chemistry, atoms or molecules may have an associated state or charge (also known as 'oxidation state'), e.g. $Na^{+}$, $Cl^{2-}$, $OH^{-}$. The state of a molecule can have important consequences for the reactions it undergoes, for example a molecule may be non-reactive in one state and highly reactive in another.

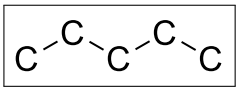
In this work, 'molecular states' are defined as two or more molecular species that have the same atomic composition in the same structure but undergo different reactions. In cases where not all reactions are known, 'molecular states' may instead refer to the same structure but with identifiable differences. It is important to note that this definition of 'molecular states' can only be used in Artificial Chemistries which have a structured molecular representation. Further, whilst the possibility of different 'molecular states' is a feature of the molecular representation used, it is possible for there to be no reactions in the Artificial Chemistry that lead to different 'molecular states'.

#### 4.1.1.6 (Stereo-)isomers

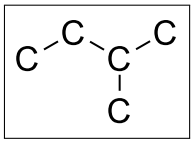
In conventional chemistry, it is possible to have multiple molecular species that have the same atomic components but arranged in a different structure. For example, in natural chemistry the molecular species pentante, 2-methylbutane and 2,2-dimethylpropane are all isomers of each other (see figure 4.1.1)

In this work, a molecular species is an 'isomer' if at least one other molecular species exists that has the same elemental composition but a different structure. An isomer may also be a 'stereo-isomer' if the other molecular species has the same elemental composition and the same bonds but in a different arrangement, such as a mirror image. Note that for some forms of intra-molecular space (such as a graph), there may only be one possible arrangement and therefore stereo-isomers are impossible.

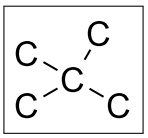
Artificial Chemistries that exhibit these properties are required to have certain characteristics. For example, isomers and stereo-isomers can only exist in an Artificial Chemistry where molecules are structure from multiple atoms.




(a) Pentane




(b) 2-methylbutane




(c) 2,2-dimethylpropane

Figure 4.1.1: Three example isomers from organic chemistry; pentante, 2-methylbutane, and 2,2-dimethlypropane. All contain 5 carbon atoms and 4 carbon-carbon bonds but there are 3 different arrangements.

### 4.1.2 Reactions

Although molecules provide the structures of an Artificial Chemistry, without reactions an Artificial Chemistry is not interesting. Reactions can be defined in two main ways — explicitly when the Artificial Chemistry is designed or implicitly by the application of a rule to a particular collection of reactants.

#### 4.1.2.1 Explicit

Explicitly defined reactions in an Artificial Chemistry are those where all possible reactions are specified and defined in advance, usually in a reaction table, e.g. Hutton's Artificial Chemistry (Hutton, 2004b, 2007) discussed in chapter 3. Explicitly defined reactions are used in Artificial Chemistries aiming to model a real-world systems where there is prior knowledge based on evidence for some or all of the reactions and this should be reflected in the Artificial Chemistry. In other cases where explicit reactions are used it is typically because a high degree of control is desired and this is reflected by stating all reactions in advance.

Such work often follows an iterative approach; create one set of reactions, see if they do what is expected, tweak the reactions and retry — repeat until the Artificial Chemistry does what is intended. In some cases, the reaction rate constants or reactions themselves could be changed by an external search algorithm in order to best match the target.

The explicit listing of reactions in a reaction table has a number of disadvantages; for interesting Artificial Chemistries, the number of reactions to specify can be large; emergent behaviours can only exist in the dynamics, not molecular structures or reactions; a reaction table can define reactions that will never occur; etc. . However, explicit reactions do have a number of advantages; mostly that they are easy to understand (particularly when compared to some

| Reactants | | | Products |
|---|---|---|---|
| A | + | B → | AB |
| AB | + | A → | ABA |
| ABA | + | B → | ABAB |
| ABAB | + | A → | ABABA |
| ABABA | + | B → | ABABAB |

Table 4.1.1: Example of an explicit reaction table using string molecules of letters A and B up to length 6.

| Reactants | | | Products |
|---|---|---|---|
| A | + | B → | AB |
| *AB | + | A → | *ABA |
| *BA | + | B → | *BAB |

Table 4.1.2: Example of an implicit reaction table using string molecules of letters A and B, up to length 6. The * character represents between zero and three characters.

implicit reaction schemes), and computationally straight-forward to work with as any reaction is a table lookup.

#### 4.1.2.2 Implicit

Implicit reaction schemes are more complicated than explicitly listing all reaction and tend to be of two types; wild-card or algorithmic.

Specifying an implicit Artificial Chemistry by the use of wild-cards is the easiest approach. Similar to an explicit chemistry, an implicit wild-card reaction table lists all reactions as reactants and products. The difference from an explicit reaction scheme is that the reactants and products may include abstract representations of other parts of the molecules. For example, an explicit chemistry shown in the reaction table 4.1.1 can be represented implicitly using wild-cards as shown in table 4.1.2. Although this example can be interpreted as describing reactions not present in the explicit reaction table (e.g. CBA + B → CBAB), if only A and B molecules are present initially the explicit and implicit chemistries are equivalent. More complicated wild-card schemes are possible, for example using regular expressions, and these may also be used to define implicit chemistries.

As well as specifying an implicit Artificial Chemistry through wild-cards,

an implicit Artificial Chemistry can be defined by an algorithm or function that is applied to a collection of reactants in order to determine the products. An example of this is binary string chemistry (Banzhaf, 1993a, 1994, 1993c) where the products are determined by folding the reactants into matrices and applying a function to create a matrix as the product.

Although the simplest implicit Artificial Chemistry functions map each collection of reactants to one collection of products, it is possible to specify divergent reactions (where multiple alternative products exist for one collection of reactants). Such divergence is often the result of a stochastic aspect to the algorithm.

In some cases, the algorithm of an implicit Artificial Chemistry can be interpreted as occurring in a spatial structure, which may be different both from the spatial structure of the individual molecules and from the spatial structure between molecules. An example of such a 'reaction space' is in StringMol (Hickinbotham et al., 2011) where each molecule is a one-dimensional string and the 'mixing space' in a well-mixed aspatial container, yet the 'reaction space' can be thought of as a graph with the various pointers forming edges between the strings. This approach has additional complexity to understand and visualize, yet combines the computational tractability of a well-mixed aspatial container with the richness of more intricate molecular interactions.

Algorithmic approaches require only a single algorithm that can be applied to all possible collections of reactants, rather than many separate reactions. However, algorithmic approaches can exhibit large responses to minor changes of the algorithm — this may be due to amplifying feedback through the population of molecules.

Despite their fragility, implicit algorithmic Artificial Chemistries can determine products for a wide range of reactants, potentially including reactants that were not part of the original specification. This is one mechanism by which novelty and emergence could occur; properties that are thought to be required for Artificial Life.

#### 4.1.2.3 Reaction rate constants

In addition to reactants and products, a reaction has a rate constant associated with it. In chemistries where this is not specified, it can be assumed to be the same for every reaction and therefore of no consequence. The rate constant of a

reaction is a real number value which represents deviation between the reaction rate and the rate at which the reactants come into close proximity (collide). This deviation could depend on orientation or energetics of the reaction, or due to attractive forces between reactants such as opposite ionic charges.

As with the reactants and products of a reaction, the reaction rate constant can be specified explicitly for each reaction in advance, or implicitly generated for a specific reaction when required.

The actual reaction rate depends on how the reaction rate constants are interpreted by the mixing algorithm of the Artificial Chemistry, which can include the influence of environmental factors such as temperature or pressure. The reaction rate is important in several features including catalysis, equilibria, and competing reactions.

### 4.1.3 Mixing

The mixing component of an Artificial Chemistry is an algorithm which describes the order of and intervals between reactions, starting from an initial collection of molecules; this can be termed a "mixing space". There are a range of options used in different Artificial Chemistries, and some features can be recombined in different ways. The mixing component of an Artificial Chemistry is generally independent from the reaction and molecular representation components, except in cases where the molecular and/or reaction space is also the space used for mixing such as in Hutton (2007). There are several broad classes of commonly used mixing components, depending on the dimensionality and granularity of the space.

#### 4.1.3.1 Well-mixed / aspatial

The simplest mixing space assumes that either space does not exist or that it is so well-mixed that everything can react with everything else at any time. This can be combined with either a discrete or continuous time. A typical algorithm that uses discrete time is as follows (Banzhaf, 1993a, 1994, 1993c):

- Given a collection of molecules ...
- ... pick and remove a number of reactants (typically 2) at random with uniform probability.

- ... determine the products of those reactants. If they do not react, the products are the reactants.
- ... add the products back into the collection of molecules.
- ... repeat.

A continuous time aspatial algorithm was described for the simulation of natural chemistry in Gillespie (1977). However, this approach works well for Artificial Chemistries too. If the Artificial Chemistry uses explicit reactions where all the reactions are known in advance, then it can be implemented as described in Gillespie (1977) or an improved version that saves computational time at the expense of computational memory (Gibson and Bruck, 2000). For an implicit Artificial Chemistry, the reactions are not known in advance and therefore the same procedure cannot be used. However, the principle of picking an interval and a reaction can still be used.

The main advantage of aspatial mixing spaces in Artificial Chemistries is their computational speed. For some Artificial Chemistries with more complicated spatial structures, a large amount of effort is spent tracking the locations of each molecule rather than determining products of reactions. However, an aspatial mixing algorithm has the disadvantage that membranes and cell cannot exist as there is no space to be partitioned. This means that whilst Artificial Chemistries using only aspatial mixing algorithms may be useful for pre-biotic environments, it is unlikely that anything resembling known biology could emerge. It may be possible, however, to create a hybrid of several 'mixing' algorithms — see §4.1.3.5

#### 4.1.3.2 Two-dimensional space

An often used form of mixing space in Artificial Chemistries is a two-dimensional discrete grid of sites where each site is occupied by zero or one molecule. Each discrete time step, each molecule (in a random order) tries to move into one of the 4 adjacent sites (usually wrapping around both directions at the edges of the grid). If the target site is occupied, a reaction is attempted instead of the movement. The movement of a molecule may be further constrained by additional links to nearby molecules, analogous to the non-covalent bonding that holds a membrane together (Hutton, 2004a).

Another mixing space is to use a two-dimensional continuous space. Each molecule is treated as a point, and reactions occur whenever two molecules are sufficiently close. In addition, each molecule has a velocity which is used to update the molecules position over time (which is usually a small discrete time-step to approximate continuous time). In contrast to the discrete grid of sites, this continuous space has a much higher computational cost, and inaccuracies in the floating-point representation of position which can lead to numerical issues.

Two-dimensional Artificial Chemistries have some issues, whether discrete or continuous. These generally stem from the topology of the space and shapes within it, for example in Hutton's Artificial Chemistry (Hutton, 2007) the linear genome acts as a barrier separating the cell into two halves that cannot freely mix.

A one-dimensional mixing algorithm for an Artificial Chemistry is possible. However, it is unclear what the benefits of such would be compared to the more intuitive two-dimensional.

#### 4.1.3.3 Three-dimensional space

Extending two-dimensional mixing spaces to three or more dimensions is possible. However, visualization and interpretation of the Artificial Chemistry becomes difficult with four or more dimensions. In practice, even three dimensional mixing spaces are rarely used in Artificial Chemistries.

Compared to two-dimensional mixing spaces, three-dimensional mixing spaces are more computationally challenging. Consider a three-dimensional discrete chemistry for example; in order to achieve the same mean distance between molecules, the ratio of unoccupied to occupied sites needs to be much higher.

#### 4.1.3.4 Sequential or simultaneous reactions

In discrete time mixing spaces in Artificial Chemistries, multiple reactions can occur in the same time-step. This can be addressed in several ways; firstly, one reaction can be completely resolved before any others are considered (sequentially), which can lead to the products reacting again in the same time-step. An alternative is for the products of a reaction to appear in the following time-step so that all the reactions in the same time-step happen simultaneously.

This means that each molecule has to exist for at least one time-step between reactions.

It is not clear which of these approaches to mixing algorithms in Artificial Chemistries is better, or indeed what the consequences of these differences are. However, this difference should be investigation in future work on Artificial Chemistries.

#### 4.1.3.5 Multiscale representations

It is possible for Artificial Chemistries to combine several mixing space representations at different scales in order to construct a multi-scale spatial representation. For example, a "global space" could consist of a number of discrete sites on a grid, but within each site the mixing space is treated as aspatial. Special reactions could represent molecules moving to nearby sites in the 'global space".

Such multicale representations tend to be used with simple Artificial Chemistries due to the computational requirements involved (Jeschke et al., 2008) and are primarily aimed at simulating known biochemical systems in which space is a feature of interest. However, with the increasing parallelism of modern computational hardware, multiscale mixing space representations may be one approach that can take advantage of such technology.

## 4.2 Design properties

Building an Artificial Chemistry requires a number of design decisions. These decisions will entail some trade-offs between different properties of the resulting Artificial Chemistry, such as increased spatial detail at the expense of computational tractability. However, the consequences of these trade-offs on the behaviour of the Artificial Chemistry are not fully known in advance. In order to guide the design process, some properties have been suggested for Artificial Life and should be considered in the construction of Artificial Chemistries.

One set of desirable properties for an Artificial Chemistry was developed by Suzuki et al. (2003) based on personal experience in developing Artificial Chemistries such as Network Artificial Chemistry (Suzuki, 2004), and by analogy with real-world biological systems. Recently, some additional properties were proposed to extend those of Suzuki *et al.* (Hickinbotham et al., 2010).

### 4.2.1 Suzuki et al. properties

Suzuki et al. described 10 properties which they considered necessary (but not sufficient) for Artificial Life. These properties can be separated into three groups (Hickinbotham et al., 2010); #1-4 relate to molecules, #5-8 and #10 relate to membranes, and #9 relates to mutation.

Some of the terminology used is not directly compatible with other terms used in this work. For example, Suzuki et al. use the term "symbol" to refer to molecules, atoms, and functional groups as a collective, without specifying if they are using a symbolic, structured, or sub-symbolic representation in the Artificial Chemistry. Where possible, the usage of the terms below has been adjusted to match the rest of this thesis.

#### 4.2.1.1 Conservation of mass

> "The [molecules] or [atoms] be conserved (or quasi-conserved) in each elementary reaction, at least with the aid of a higher-level manager." (Suzuki et al., 2003)

The property of conservation of mass is seen by Suzuki et al. as key for open-ended evolution. If it is not present, then it is easy to imagine systems which generate their own food supply — and could then continue to grow unbounded. Conservation of mass prevents this unbounded growth, and also provides competition for resources as the available raw materials are consumed. This may lead to various evolutionary dynamics, such as predation or the consumption of waste. The concept of a "higher-level manager" may or may not be useful, particularly as Artificial Chemistry for Artificial Life is thought to depend upon bottom-up emergence rather than top-down control.

Related to conservation of mass, but not directly addressed in Suzuki et al. (2003), is the idea of Conservation of Energy. Indeed, much of the biochemistry of living cells is restricted by available energy rather than mass as almost all atoms in a cell are either carbon, hydrogen, oxygen, or nitrogen. This issue will be addressed at a future point (see §4.2.3.2).

#### 4.2.1.2 Unbounded molecular size / structure

> "An unlimited amount of information be coded in [an atom] or a sequence of [atoms]." (Suzuki et al., 2003)

As with property #1, an unbounded molecular size and/or structure is regarded as useful for open-ended evolution by allowing the size of a genome (or similar) to grow over time. There are many ways to encode an unlimited amount of information, and not all of them may be useful. For example, one could encode an unlimited amount of information by using an unlimited number of elements — but this may not be useful without structuring the elements. However, on balance, the ability for an Artificial Chemistry to encode molecules of any size is a desirable one.

When combined with conservation of mass, this property implies a trade-off. If molecules are larger than they have to be, then they are using mass which could be used elsewhere, e.g. for replication. However, if all mass is strictly required, then there is less room for neutral mutation that could lead to evolutionary favourable adaptations in the future.

#### 4.2.1.3 Catalysis

> "Particular [molecules] that specify and activate reactions be present." (Suzuki et al., 2003)

Whilst catalysis is undoubtedly critical to an Artificial Chemistry for Artificial Life, Suzuki et al.'s definition of it may not be the most suitable. Catalaytic reactions are not intrinsically different from any other reactions, they are merely the subset where a product is also a reactant — either directly or indirectly through a series of reactions. Further, catalysis is not an all-or-nothing process in real-world chemistry; it is due to relative rates and depends upon a variety of factors.

Due to these reasons, it is suggested that catalysis should not be thought of as a design property to be built into an Artificial Chemistry. Rather, it is an emergent consequence of one or more reactions. As catalysis is critical to Artificial Chemistry for Artificial Life it is discussed in more detail with other emergent properties in §4.3.1.

#### 4.2.1.4 Embedded gene expression

> "The translation relation from genotypes to phenotypes be specified as a phenotypic function." (Suzuki et al., 2003)

At the core of this property is the idea of the evolution of evolvability. By placing the mechanisms of evolution within the evolving system, selection for

those organisms that can evolve to fitter forms faster will have an advantage — as long as the environment is sufficiently dynamic. This is a well-known phenomena, and applies not only to genotype-phenotype but also to mutation, cell division, gene regulation and other processes.

When applying this idea to Artificial Chemistries, it is not clear in an emergent system what is a genotype and what is a phenotype. For example, consider an RNA world; it is straightforward to consider the sequence of each RNA molecule as the genotype, although it is questionable if each molecule should be considered as a separate individual or if they should be grouped together somehow. The definition of phenotype in an RNA world is less clear-cut; is it the shape of the folded molecule, the reactions it participates in, or something else? As a consequence of this inability to cleary define both phenotype and geneotype in arbitary Artificial Chemistries, the usefulness of this property as described in Suzuki et al. (2003) is limited.

#### 4.2.1.5 Cells

> "The [mixing] space be able to be partitioned by semi-permeable membranes, creating cellular compartments in the space." (Suzuki et al., 2003)

It is well established that, in the absence of spatial partitioning, well-mixed populations of simple evolving self-replicators become overrun with parasites — resulting in mutual extinction. If a simple means of temporary spatial partitioning exists, such as spontaneously formed membrane vesicles, then group-level selection begins to apply; those vesicles that contain too many parasites replicate less or die out entirely.

In addition, local spatial structure that restricts movement causes molecules to repeatedly interact with neighbouring molecules, which may have been the products of previous reactions. This means that any emerging life-like organisation has to cope with the consequences of its behaviours, both good and bad. Through mechanisms such as this, the impact of evolutionary fitness is focused back onto the individuals and the overall rate of evolution within the population is enhanced.

Although Suzuki et al. seem to know of these features, they have focused on semi-permeable membranes as the mechanism for achieving this spatial structure. This is the compartmentalization mechanisms used between, as

well as within, all living cells that we know of. It is easy, however, to imagine other features that could achieve the same effects; e.g. tidal pools, freeze-thawing of ice, micro-pores in rocks, and other surface/solvent borders. These other mechanisms may be as important, if not more so, than semi-permeable membranes — at least for the origin of life. Therefore, while the partitioning of space into compartments[c] can be regarded as essential, Suzuki et al.'s assertion that these should be semi-permeable membranes is more contentious.

#### 4.2.1.6 Dynamic cellular contents

> "The number of [molecules] in a cell can be freely changed by [molecule] transportation" (Suzuki et al., 2003)

It is claimed that this property allows for evolution of higher functions because cells with higher functions have more molecules in them. Whilst this may be true, it is trivial to imagine a cell that continually expands to contain more molecules or molecular species without developing any higher functions.

In light of this, it may be better to describe this property as the following instead:

> "The number of molecules in a cell can be freely changed by molecule transportation and the rate of such change is influenced by local conditions."

This not only reflects that compartments should be of varying sizes, but also that organisms should both have influence on and be influenced by the size of such compartment. However, it would be interesting future work to determine the extent to which this applies and how it influences the emergence and evolution within such systems.

#### 4.2.1.7 Cell movement

> "Cellular compartments mingle with each other by some random process." (Suzuki et al., 2003)

This property is described as applying not only to mingling of cells, but also to mingling of sub-compartments within a cell.

[c] mixing between compartments is slower than within a compartment

It is obvious that cellular compartments are used by many biological organisms for a range of purposes within a cell; e.g. nutrient uptake, waste storage, assembly-line metabolism, toxin transport, etc. However, it is not clear if this is a requirement for life or an adaptation of early living systems. In known biological cellular processes, the movement of sub-cellular compartments is rarely "random"; it is controlled by proteins and enzymes.

When this property is applied to mingling between cells, it is again unclear if it forms a requirement or a common consequence. The example Suzuki et al. use is that of Tierra (Ray, 1992); within that system, a parent reproduces into a random adjacent unoccupied site. This is claimed to lead to the emergence of a parasitic relationship between creatures. However, a wide range of mathematical biology studies have shown that the consequences of neighbourhood spatial structure is more complicated than this, with different degrees of relatedness, cooperation and competition between organisms all having an influence on population dynamics under different spatial structures.

Further, many organisms are able to influence their movement — either in a highly controlled manner such as walking or flight, or in the stay-or-go behaviour seen in simpler organisms such as bacteria. This indicates that the mingling between cells is not just a random process, but one organisms can influence through a range of mechanisms.

Because of these reasons, this particular property may or may not be a requirement for Artificial Life. Although the principle of movement both within and between cells is seen as desirable, it is questionable what the consequences of these properties are and how the emergence of Artificial Life is influenced by it. Therefore, this would be another interesting area for future work.

#### 4.2.1.8 Intra- and inter-cellular signaling

> "In-cell or between-cell signals be transmitted in the manner of [molecule] transportation." (Suzuki et al., 2003)

This property can be seen as a consequence of dynamic cellular contents (§4.2.1.6) and conservation of mass (§4.2.1.1). As compartments gain and lose molecules, conservation of mass ensures that these molecules are present to be transported out of one compartment and into another compartment, potentially via the wider environment. If these molecules trigger a change in

the receiving compartment, then the molecule can be interpreted as having been a 'signal'.

However, it should be noted that in biology there is no *a priori* separation of signaling molecules from non-signaling molecules. It is the consequences of their interactions that determines this, though certain types of molecules may be more suited depending on the details of the Artificial Chemistry.

#### 4.2.1.9 Spontaneous mutation

> "There be a possibility of [molecules / atoms / functional groups] being changed or rearranged by some random process." (Suzuki et al., 2003)

It is obvious that without a source of novelty, evolution cannot occur. In modern cells, novelty usually means changes in the DNA sequence of the organism and is typically caused by rare events such as damage or copying / repair errors. However, although these events are highly stochastic in nature, they are not entirely random. Biological organisms have some influence over DNA repair and replication mechanisms, and hence their mutation rate. This can be seen, for example, in bacterial colonies which enter a state of 'hyper-mutation' when faced with some forms of environmental danger. Further, it is a widely accepted hypothesis that the need to generate novelty upon which evolution can act was a contributing factor in the evolution of diploidy, polyploidy and sex. Evidence such as this has lead to the concept of the 'evolution of evolvability', where an evolutionary system can be self-tuning. Although Suzuki et al. acknowledge this with respect to the genotype-phenotype mapping, it is surprising that they do not apply this concept to reproduction, and hence evolution, directly.

The description of this property in Suzuki et al. (2003) also refers to the exposure of organisms to locally novel molecules having been transported between cells. Such external sources of material can be important in the adaptation and evolution of living systems; however, unless such molecules are of a sufficiently stable supply, any evolution triggered by them is not sustainable. Without repeated exposure, such interactions are rare random events and cannot be adapted to. If such sources are stable, at least in the medium-term, then they can perturb the evolutionary landscape and potentially trigger novel circumstances if the supply of molecules disappears in the future. However,

such evolutionary dynamics are not a consequence of the supply *per se*, but rather a property of changes in the environment which could also include space, temperature, pressure, etc. Therefore, these are considered in more detail in §4.2.2.3.

#### 4.2.1.10 Catalysis of position

> "[Molecules] be selectively transferred to specific target positions by particular activator [molecules] (strongly selective), or at least selectively transferred by [molecular] interaction rules (weakly selective)."(Suzuki et al., 2003)

The final property Suzuki et al. discuss relates to relative movement of molecules as a consequence of their interaction. A typical example of this is transport across a membrane — over and above the passive diffusion of smaller molecules. This can be thought of as catalysis of location change, rather than changes in structure. Like structural catalysis, it may be more beneficial for this property to be an emergent consequence rather than an property specified *a priori* . This would allow for adaptation through evolution.

Like several of the properties proposed in Suzuki et al. (2003), it is not clear if catalysis of position is a strict requirement for the emergence of Artificial Life in an Artificial Chemistry. Pores and pumps across membranes are found in all living cells, from bacteria through to humans, but they may be so widespread due to an early adaptation rather than being a prerequisite. Determining if catalysis of position is a requirement or not is not an easy task; the strongest evidence would be two almost equivalent systems — one with catalysis of position, and the other without — where emergence of life-like properties occurs only in the system that includes catalysis of position. Another possibility is that life can emerge without catalysis of position, but is then limited in how far or how fast it can evolve. However, investigating this would be future work.

### 4.2.2 Hickinbotham at al. properties

In addition to the properties described by Suzuki et al., 8 properties were recently suggested by Hickinbotham, Faulconbridge, and Nellis (2010) based on the authors' development of three Artificial Chemistries. In contrast to the properties described by Suzuki et al., the properties described by Hickinbotham

et al. are less specifications and more akin to guidelines or potential topics of investigation.

#### 4.2.2.1 Novelty and innovation

This property is similar to the "unbounded molecular size / structure " property of Suzuki et al. (see §4.2.1.2). However, rather than simply being the number of possible molecular species the concept of "novelty and innovation" applies to Artificial Chemistries where different molecular species have different functional characteristics. New functional characteristics may appear over time in such an Artificial Chemistry as previously unseen molecular species are created and/or introduced from outside the system. These novel functions may provide an evolutionary advantage, and thus drive open-ended development of the system.

However, it is unclear how such a property could be detected and measured in an Artificial Chemistry. One possibility is to identify the appearance of previously unseen molecules over time, and then determine if the appearance of novel molecules is correlated with a change in the population dynamics. However, such work would be computationally challenging and practically difficult.

#### 4.2.2.2 Range of scales

Living systems exhibit recuring dynamical motifs; the ecosystem dynamics between gut bacteria is similar to the dynamics between the organisms they live within. This self-similar structure requires that a sufficiently large range of scales is spanned to contain these self-similarities. These scales can be spatial, temporal, or environmental parameters such as heat, light, or pH. For example, stationary organisms that extract nutrition from the local environment is an archetype demonstrated by trees, coral reefs, and deep-sea vent tubeworms. Over evolutionary time, a species may move over several of these scales; the ancestors of dolphins once walked on land, and before that were water-based.

Whilst this may not be a strict requirement for the emergence of Artificial Life, a range of scales could be an important factor in separating Artificial Chemistries that can support open-ended evolution from those that reach a steady state.

#### 4.2.2.3 Dynamic environment

Changes in the environment have been critical to evolution of life on Earth — for example mass-extinctions or more localised changes such as day/night cycles and temperature gradients. These environmental differences exist in both spatial and temporal scales, and exhibit varying degrees of consistency and stability. Evolving systems have an incredible adaptability for exploiting novel conditions, and these conditions can lead to the development of novel biological features that can re-invade the original ecological niches they evolved away from.The evolution of whales is an example; originally evolving from fish that moved onto land, they subsequently moved back into water over evolutionary time. Such dynamic environments are also more resistant to extinction by a common parasite or other factor — it is likely that something, somewhere will survive almost anything that the environment can produce. Further, a dynamic environment over time selects for the evolution of evolvability, and such self-tuning systems are better able to adapt to other challenges.

Without dynamic environments, any system that looks at the emergence of Artificial Life has to ensure that the environment is suitable. In a dynamic environment, a life-like system only has to emerge once within the range of the environment and will adapt as it spreads to other regions.

#### 4.2.2.4 Redundancy and degeneracy

The properties of redundancy and degeneracy may appear to run counter to the properties of novelty and innovation. Redundancy means that there are two or more molecules that are completely interchangeable in their reactions, and degeneracy applies when two or more molecules are interchangeable in some of their reactions. In a system with redundancy and degeneracy, there must be more molecular species to enable the same degree of novelty and innovation as a system without redundancy and degeneracy. However, redundant/degenerate molecular species may be the products of different reactions and this can allow for the generation of 'hidden' variation within an evolutionary system. This variation can then be selected from when evolutionary conditions change. For example if temperatures rise and cause formerly degenerate molecules to undergo different reactions, or if a precursor of a molecule becomes rarer.

#### 4.2.2.5 Emergent complex properties

Artificial Chemistry for Artificial Life is fundamentally a search for emergent properties, and the emergence of life in particular. Additionally, living systems are typically complex systems where a small change in a component may have large repercussions throughout the system, or may have no measurable effect at all. In light of this, it seems obvious that an Artificial Chemistry for Artificial Life should have the property of emergence of complexity.

However, systems with emergence are difficult to approach analytically because a traditional reductionist approach of analysing individual types of components fails to capture the emergent properties resulting from the interactions. Rather than developing tools and techniques to deal with this emergence, it is easier to use a system without emergent properties — but this undermines the purpose of studying Artificial Chemistries for Artificial Life. The same argument applies to complex properties as well, and systems with emergent complex properties are particularly challenging for traditional reductionist approaches.

#### 4.2.2.6 Unified molecular representation

The idea of a unified molecular representation is that there is no *a priori* separation between role and molecule. This means that the same rules and constraints apply equally to membranes, enzymes, genomic material, and metabolic molecules. The function of a particular molecule within the organism should be an emergent property, rather than a top-down separation.

It is claimed in Hickinbotham et al. (2010) that a unified molecular representation enables more self-optimization potential and further open-ended evolution. However, whilst this seems to be an intuitively correct statement, there is little evidence either for or against.

#### 4.2.2.7 Stochasticity

Another property suggested by Hickinbotham et al. is stochasticity. In particular, they contrast stochastic Artificial Chemistries with deterministic ones and assert that stochastic Artificial Chemistries have more evolutionary potential due to more novelty generation. This may be true, but this novelty does not have to be generated by stochasticity. Divergent reactions can generate many different molecular species from a small collection of initial reactants,

and although selecting between possible products is usually done at random for simplicity, it does not have to be; for example, divergent reactions could alternate between outcomes rather than choosing at random.

The other benefit of stochasticity suggested by Hickinbotham et al. is a smoothing of evolutionary changes by enabling sampling of the reaction space. For example, from multiple stochastic interactions, rare combinations of events can happen occasionally. This provides an intermittent individual selection pressure for those rare events, which are inevitable over longer time-scales. This selection of responses to rare events may allow a population to gradually overcome an evolutionary challenge, rather than being faced with an all-or-nothing situation. The reverse is also possible; a population could persist despite adverse environmental conditions for a limited time by being 'lucky' rather than successful.

#### 4.2.2.8 Emergent mutation rates

The final property proposed by Hickinbotham et al. is 'emergent mutation rates'. Like the 'embedded gene expression' property of Suzuki et al. (see §4.2.1.4), 'emergent mutation rates' enables an organism to evolve evolvability and adapt its own evolutionary process to best fit its environment.

In particular, Hickinbotham et al. suggest that some mutation should be due to error-on-copy rather than an absolute external mutation rate. This would enable a system to manipulate its own mutation rates in more intricate and subtle ways — for example, by encouraging some kinds of genome base substitution rather than others. Error-on-copy is one mechanism by which mutation rates can be influenced by emergent properties.

#### 4.2.2.9 Conclusions

Overall, the properties proposed by Hickinbotham et al. as desirable for an Artificial Chemistry for Artificial Life are interesting. However, there is little concrete evidence either for or against many of them and therefore they have to be taken with a pinch of salt until such time as evidence is available. The process of performing such studies could shed valuable light, not only on the properties suggested by Hickinbotham et al. but on general properties of Artificial Chemistries.

### 4.2.3 Other design properties

In addition to the properties discussed above, there are more properties that should be considered when designing an Artificial Chemistry.

#### 4.2.3.1 Functional groups

In real chemistry, sub-structures of some molecules have consistent emergent properties common to that sub-structure independent of the rest of the molecule. For example, a hydroxyl group, an ester group, a benzyl group, etc. In biological systems, functional groups provide a general framework for the manipulation and/or construction of larger structures. Proteins are examples of this on multiple scales; at the smallest scale the peptide bonds joining amino acids together allows breakage/formation without dependence on the sequence. At larger scales, secondary structures form such as alpha helices and beta sheets. As with peptide bonds, these are not unique to specific molecules but rather are the result of local sub-structure. Finally, the catalytic activity of many enzymes is due to the side-chains of a few specific amino acids, and the rest of the protein sequence can be changed dramatically as long as those key amino acids remain constant.

For Artificial Chemistries, the ability to describe functional groups enables similar reactions to happen to a range of molecules with common local sub-structures, which could be important for emergent evolution and other life-like characteristics. However, the potential of functional groups has not been investigated to date nor has any comparison of similar Artificial Chemistries with and without functional groups been described. Again, this would be an interesting topic for future work.

#### 4.2.3.2 Conservation of energy

In real chemistry, both energy and mass are conserved. This is demonstrated by exothermic or endothermic reactions — as the reaction proceeds, the products have more/less thermal energy than the reactants which results in heating/cooling respectively. In turn, temperature has an effect on the rate that reactions occur at, both by changing the number of molecular interactions and by changing the average free energy of those interactions. This effect on reaction rate is not uniform across different reactions and can therefore change the composition of the molecules.

On a smaller scale, such as within the complicated biomolecular machinery of a cell, conservation of energy may have resulted in the evolution of currency metabolites such as Adenosine Tri-Phosphate (ATP). These act as a common chemical resource that can be produced from a wide range of food sources (e.g. sugars, starch, protein) and is consumed in a wide range of cellular processes (e.g. DNA replication, membrane transport, etc. )

Applied to Artificial Chemistries however, the effects of conservation of energy are not known. Many Artificial Chemistries do not have conservation of mass, and while it may be possible to have conservation of energy without conservation of mass it would be very unintuitive. Of those Artificial Chemistries that do have conservation of mass, none of them have addressed conservation of energy directly. Therefore, this would be an interesting topic for investigation in future work.

## 4.3 Emergent properties

Using the concept of emergence discussed previously (§2.2), a number of interesting properties of Artificial Chemistries can be classified as emergent properties. These are not properties of any individual Artificial Chemistry, nor of any individual molecular species. Rather, emergent properties are properties of groups of molecular species within an Artificial Chemistry.

An approach towards identifying Artificial Chemistries based on some of their emergent properties is Chemical Organization Theory (Dittrich and di Fenizio, 2007). In Chemical Organization Theory, sets of molecules are classified based on properties such as closure or self-maintenance. There are a number of other emergent properties described, such as autocatalytic sets (Kauffman, 1969), hypercycles (Eigen and Schuster, 1977, 1978a,b), reflexive auto-catalytic food-generated (RAF) sets (Hordijk and Steel, 2004) and autopoesis (Maturana and Varela, 1980). Some of these were developed in the context of a historical hypothesis of pre-biotic Earth, but are also applicable to Artificial Chemistries for Artificial Life.

In addition to these specific emergent properties, a range of other terms are used when describing properties of Artificial Chemistries with often imprecise and/or varying definitions. The confusion is partly due to: terms being 'borrowed' from other disciplines such as chemistry or biology; different interpretations of terms by researchers from different backgrounds; by application

of the same term to Artificial Chemistries with radically different foundations from chemistry. Such terms are re-defined in the context of this thesis so that they can be applied appropriately to Artificial Chemistries.

### 4.3.1 Catalysis

Catalysts are a category of molecules that increase the overall reaction rate of one or more reactions without being consumed; the catalyst catalyses the reaction(s). In biology, proteins that are catalysts are known as enzymes. Artificial Chemistries may or may not have different rate constants for different reactions. A reaction that occors either with or without the catalyst is known as 'rate catalysis'[d]. Some reactions only happen in the presence of a catalyst, in which case the rate constant can be though of as being zero without the catalyst; such cases are termed 'absolute catalysis'.

#### 4.3.1.1 Direct and indirect catalysis

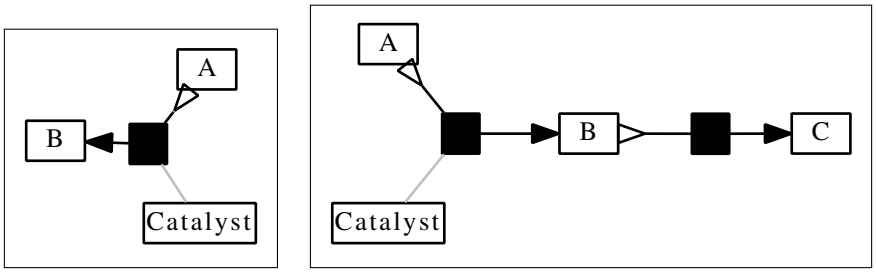
If a catalyst increases the rate of production of one or more molecular species, then that catalyst is said to catalyse the production of those molecular species. Usually, this is 'direct catalysis' because the molecules being produced are in the same reaction as the catalyst. Alternatively, this can be 'indirect catalysis' if the catalyst increases the production of a precursor molecular species which is a limiting factor, or if the molecular species produced by the catalysed reaction molecule undergoes spontaneous reactions (e.g. decay) resulting in the production of the product molecular species.

#### 4.3.1.2 Multi-stage catalysis

Catalysis can come from a sequence of multiple reactions, and in such cases the catalyst can be consumed by some of the reactions as long as they are subsequently produced, thus not being consumed by the sequence overall. These typically form a circular shape when drawn as a reaction graph, but should not be referred to as a 'catalytic cycle' as this term is used by Eigen and Schuster (1977) for a different phenomenon.[e] Instead, these reactions are termed 'multi-stage catalysis'. Figure 4.3.2c shows an example of a sequence of

[d] if the reaction is faster with the catalyst

[e] This is the convention used in this work; other works may vary.




(a) Direct catalysis of a single reaction (Catalyst + A → Catalyst + B)

(b) Indirect catalysis. An additional reaction (B → C) means that the catalyst indirectly catalyses the production of C

Figure 4.3.1: Examples of various forms of catalysis. Reaction graphs where nodes represent either a molecular species or a reaction. Open arrows away from a node represent reactants, and closed arrows towards a node indicate it is a product. Grey lines indicate a catalyst. Rate constants are indicated by a number; where not indicated it is assumed to be 1.0.

three reactions that collectively have the same overall outcome as the simpler case shown in figure 4.3.1a.

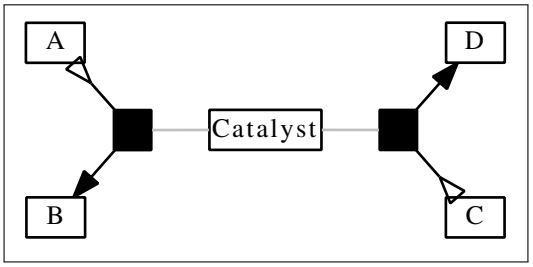
#### 4.3.1.3 Multiple catalysis

A catalyst that only catalyses a single reaction can be termed a 'mono-catalyst'. By extension, a 'bi-catalyst' catalyses two reactions (e.g. 4.3.2a), a 'tri-catalyst' three reactions, and so on; a 'poly-catalyst' describes a catalyst that increases the rate constant of two or more reactions.

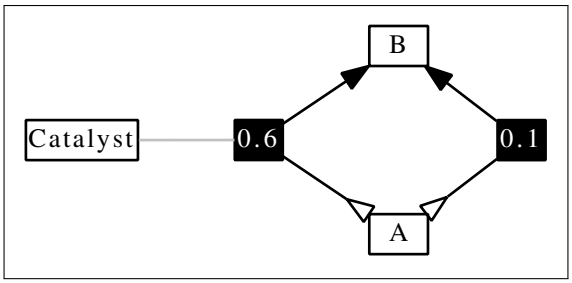
A single molecular species can have multiple types of catalysis. For example, a molecular species could directly catalyse one reaction and form a multi-stage catalytic reaction with a different reactant — this would also mean that the catalyst was a bi-catalyst.

### 4.3.2 Competing reactions

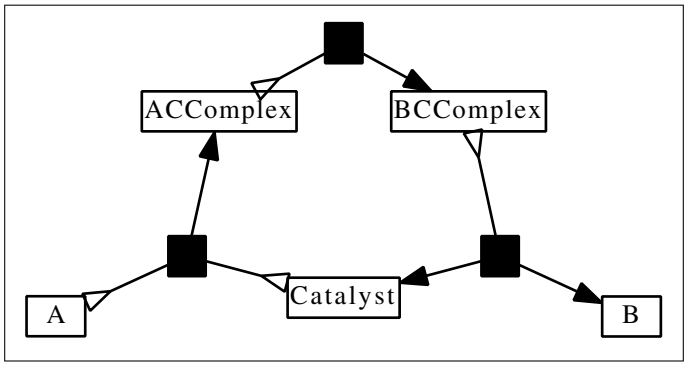
Most molecules in most Artificial Chemistries can participate in several different reactions, depending on what the other reactants are. In such cases, reactions that share one or more reactants can be said to be competing reactions for that reactant. How many reactions compete for a reactant is an emergent property of the chemistry as a whole, rather than a property of any one reaction or molecular species. The distribution of the number




(a) Bi-Catalyst catalyses two separate reactions (Catalyst + A → Catalyst + B; Catalyst + C → Catalyst + D)




(b) Different rates of catalysis. The rate constant of the reaction with the catalyst is greater than the rate constant of the reaction without the catalyst.




(c) Three-stage catalysis. The catalyst and the other initial reactant (A) undergo two intermediate steps before the catalyst is released with the other product (B).

Figure 4.3.2: Additional examples of various forms of catalysis. Representation as in figure 4.3.1.

and size of sets of competing reactions may have some relationship to other emergent properties seen in an Artificial Chemistry, however, this has not been investigated to date.

### 4.3.3 Divergent reactions

Typically in Artificial Chemistries, any collection of reactants can undergo at most one reaction. However, it is also possible to have an Artificial Chemistry in which a single collection of reactants undergo one of several reactions that each produce different collections of products. Most commonly which of the reactions occors is determined randomly, or may be linked to other details of the chemistry such as relative orientations.

For example, in Banzhaf (1993a) reactions are the application of one molecule to another, and any molecule can be both operator and subject (e.g. A + B and B + A). This leads to divergent reactions if these result in different products (e.g. A + B $\rightarrow$ A + B + C and B + A $\rightarrow$ B + A + D).

The presence of divergent reactions in an Artificial Chemistry can lead to an increase in diversity of molecular species, which could enable emergence and selection processes to occur.

### 4.3.4 Reversible reactions

Some reactions can be said to be reversible. These are reactions for which there a second reaction exists where the collection of reactants of the first reaction is the collection of products of the second reaction and *vise versa* (see figure 4.3.3). In such cases, an equilibrium point will exist between them where the rate of the first is equal to the rate of the second — resulting in no overall change in the amounts of the molecules involved. Such an equilibrium may or may not be reached in any particular instance of an Artificial Chemistry, depending upon factors such as the rates of other reactions. Pairs of reversible reactions can be linked together through shared molecular species to form chains or other structures, increasing the number of molecular species that are stabilized.

Reversible reactions are an interesting emergent property as they can provide a degree of stability to the molecules present in the chemistry at any point in time. This is because the rates converge on the equilibrium point in the absence of other factors, given that the rates of the reactions are dependant

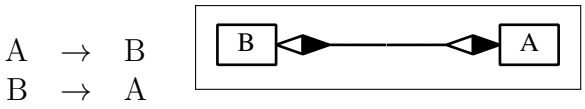



Figure 4.3.3: Example reversible reaction. Nodes represent molecular species and edges represent reactions. The half-filled diamond indicates that there are two reactions over the same edge with opposite reactants/products.

on the concentration of the reactants. Such stability can lead to persistence over time of some types of molecules, which could be a requirement for life-like properties.

### 4.3.5 Rearrangement reactions

Reactions where one or more products are isomers (see §4.1.1.6) of reactants are said to be "rearrangement reactions". For example, ABC → BAC.

The consequences of rearrangement reactions in an Artificial Chemistry are not clear; in real chemistry, molecular species which undergo rearrangement reactions often undergo convergent and/or reversible reactions too. In addition, many intra-molecular reactions — such as protein folding — can be classified as rearrangement reactions according to this definition.

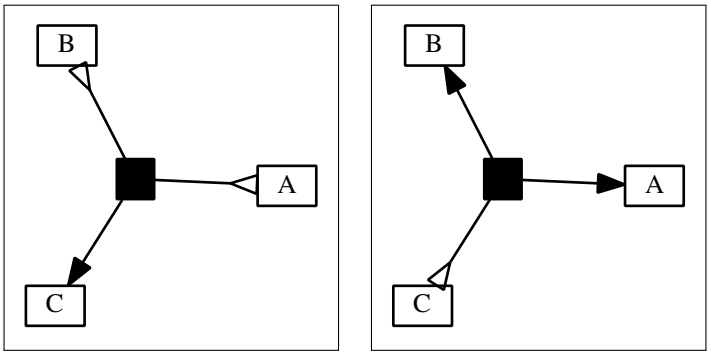
### 4.3.6 Synthesis and decomposition

In synthesis reactions multiple reactant molecules combine to form fewer product molecules, typically just one product molecule. Synthesis reactions are also known as a 'combination reaction'. Most commonly there are two reactants and a single product, but the term could apply to any reaction where there are fewer products than reactants. See figure 4.3.4a for an example.

Decomposition is the opposite of synthesis - a reaction where there are more product molecules than reactant molecules. This may be a spontaneous process such as decay, or may require the participation of other molecules. Decomposition may also be called 'degradation', especially when catalysed. An example of decomposition is shown in figure 4.3.4b.

### 4.3.7 (Hetero-)polymers

Molecular structures that are of interest in the context of Artificial Chemistry for Artificial Life are 'polymers' and in particular 'heteropolymers'. A polymer




(a) Synthesis (b) Decomposition

Figure 4.3.4: Examples of synthesis and decomposition reactions.

is a molecule composed of repeated subunits (monomers), usually formed into a linear structure of unspecified length. Many plastics are polymers, as are many structural biomolecules such as cellulose and actin. Note that according to the definition of molecular species given previously, many different molecular species can be the same polymer depending on exactly how many monomers are involved. A simple example of part of a heteropolymer is shown in figure 4.3.5a.

Polymers that are composed of different monomers are known as 'heteropolymers'[f]. Heteropolymers in biology are composed of many subunits of a few different types.

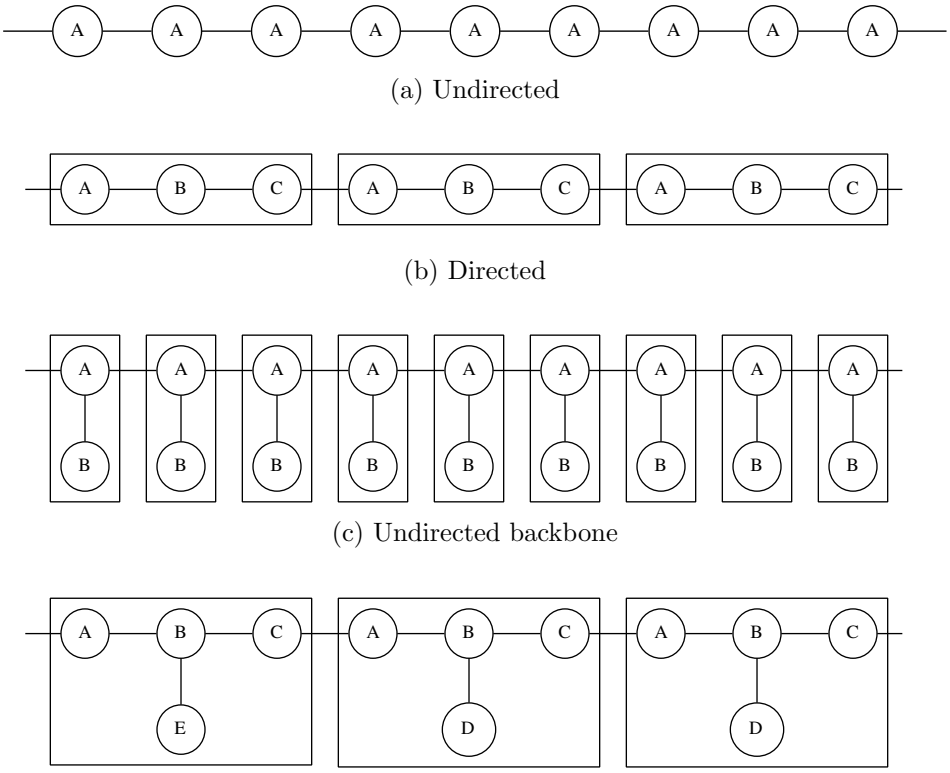
#### 4.3.7.1 Undirected and directed

The subunits of a polymer may be symmetrical or asymetrical with respect to the direction of the polymer, and this results in an 'undirected' or 'directed' polymer respectively. The directedness of a polymer can be detected by local interactions and does not require global information about the entire molecule. This feature is critical in biological systems to enable molecules to move along the polymer. The example shown in figure 4.3.5a is undirected but figure 4.3.5b shows an example directed polymer.

#### 4.3.7.2 Backbone

A common structure of heteropolymers is to have a shared 'backbone' component with different side-chains attached to each monomer (e.g. figure 4.3.5c).

[f]Polymers composed of one type of subunit could be called 'homopolymers'.




(a) Undirected

(b) Directed

(c) Undirected backbone

(d) Directed backbone hetero-polymer

Figure 4.3.5: Several examples of polymers. Circles indicate atoms within the molecule. Boxes around atoms indicates the repeating multi-atom groups (monomers) within the overall molecule. Note that these represent only part of the molecular structure, as indicated by the bonds extending sideways.

Molecules of this type include DNA, RNA and protein — all of which are key building blocks of life as we know it and all of which are directed heteropolymers with a backbone structure. Whilst such molecular structures may not be a strict requirement for Artificial Life, it is difficult to envisage how Artificial Life would function without them.

#### 4.3.7.3 Role of heteropolymers

Directed heteropolymers are thought to be important because information can be encoded in the order of the subunits. This information can be 'read' by moving along the molecule in a consistent direction. This movement is simpler if there is a shared backbone component to each monomer. Discovering the importance of a sequence of subunits is the hallmark of modern biology.

In many Artificial Chemistries, all molecules are heteropolymers modelled

as strings (Kauffman and Farmer, 1986; Banzhaf, 1993a; Hickinbotham et al., 2011). However, this approach removes much of the subtlety of biological heteropolymers with a shared backbone, such as the double-helix of DNA or the many folded structures of proteins. Further, representing all molecules are strings reduces the possibilities for non-heteropolymer molecules — such as food, toxins and waste.

Although heteropolymers could be built into an Artificial Chemistry directly, if they were an emergent property of the Artificial Chemistry then it may have greater potential for the emergence of Artificial Life. Detecting the existence of heteropolymers may be problematic, and could be investigated in future work.

### 4.3.8 Autocatalytic sets

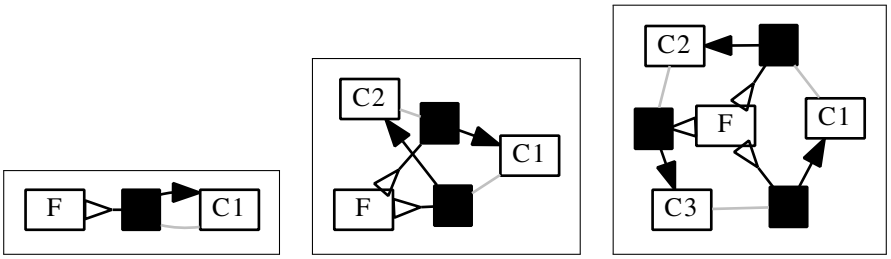
Autocatalytic sets have been a concept in biology for a considerable time, however in the context of Artificial Chemistry for Artificial Life they were best described by Kauffman and Farmer (1986).

An autocatalytic set is defined as a group of catalysts where the formation of each member of the group is catalysed by at least one member of the group. The simplest autocatalytic set is a single molecular species that catalyses its own formation ($A \rightarrow A + A$). In the case of an autocatalytic set of size one, that molecular species is referred to as an autocatalyst.

Autocatalytic sets can overlap, either partially or completely. This, theoretically, provides a point where natural selection could apply. Autocatalytic sets would compete for a limited, but periodically replaced, supply of 'food'. If all molecules are removed equally by an mechanism, for example outflow, then eventually autocatalytic sets that are more efficient at their formation would more rapidly increase the number of their molecules.

#### 4.3.8.1 RAF sets

In early work on autcatalytic sets (Kauffman and Farmer, 1986), it was suggested that any sufficiently large system with a sufficient level of catalysis will be highly likely to contain autocatalytic sets. However, subsequent work (Hordijk and Steel, 2004) has suggested that this is due to increasing probability of catalysis with the size of the system. Further, Hordijk and Steel (2004) expanded the concept of autocatalyic sets to RAF sets - Reflexively




(a) Size 1 (b) Size 2 (c) Size 3

Figure 4.3.6: Example autocatalytic set of various sizes (see §4.3.8). F refers to an abstract 'food' molecular species. C$i$ are the catalysts that are members of the autocatalytic set.

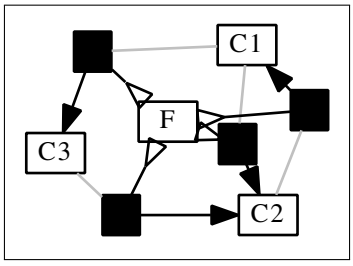



Figure 4.3.7: Example of overlapping autocatalytic sets (see §4.3.8). The molecular species C1 and C2 form an autocatalytic set, as do C1, C2 and C3. This can be written as {{C1, C2}, C3}

Autocatalytic sets generated from a Food source[g]. The concept of RAF sets focuses on the dependence of an autocatalytic set on a sufficient supply of food molecules in order to continue growing.

#### 4.3.8.2 Irreducible RAF sets

The concept of an 'irreducible RAF set' is introduced by Hordijk and Steel (2004). An irreducible RAF set is a RAF set where the removal of any molecular species results in a non-RAF set. The idea of an irreducible RAF set is used to discriminate against RAF sets that are composed of two or more separate irreducible RAF sets, and the union is therefore only a RAF set because all of its components are.

[g]In previous work (Steel, 2000) these were known as CRA sets

#### 4.3.8.3 Food sets

Intuitively, it can be seen that a RAF set that requires a large number of different molecular species as food is less life-like than a RAF set that requires fewer different molecular species, or is stable over multiple sets of food molecular species. This could lead to a comparison of RAF sets based upon the number and range of food molecular species it can be supported by, as well as the size of a RAF set and its replication rate. Unfortunately, this opportunity is missed and instead Hordijk and Steel (2004) assume that an *a priori* specifies set of molecular species is the food supply. Whilst this may be useful for investigating properties of known RAF set generating chemistries, it limits the application of RAF sets to Artificial Chemistries in general.

### 4.3.9 Catalytic cycles and hypercycles

A desirable property of Artificial Chemistries is that of the hypercycle (Eigen and Schuster, 1977, 1978a,b). This is a scheme in which self-replicating entities combine to produce a 'higher' level of organization. However, the original works are obfuscated by a lack of succinct and precise definitions, as well as loose analogues to more complicated modern real-world chemical reaction networks both biological and non-biological.

#### 4.3.9.1 Catalytic cycles

The first term Eigen and Schuster use is 'catalytic cycle' (see figure 4.3.8); this appears to be a subset of autocatalytic sets. Eigen and Schuster add a further limitation on an autocatalytic sets; that the members of the autocatalytic set must form a directed ring where each molecular species catalysts the formation of the next molecular species along the ring. This topological requirement is not used elsewhere in their work, and Artificial Chemistries without poly-catalysts (i.e. where catalysts can only catalyse a single reaction) will always meet this requirement. Therefore, 'catalytic cycle' can be discarded in favour of 'autocatalytic set'.

#### 4.3.9.2 Catalytic hypercycles

The second term Eigen and Schuster use is 'catalytic hypercycle' (often shortened to 'hypercycle') which they describe as:

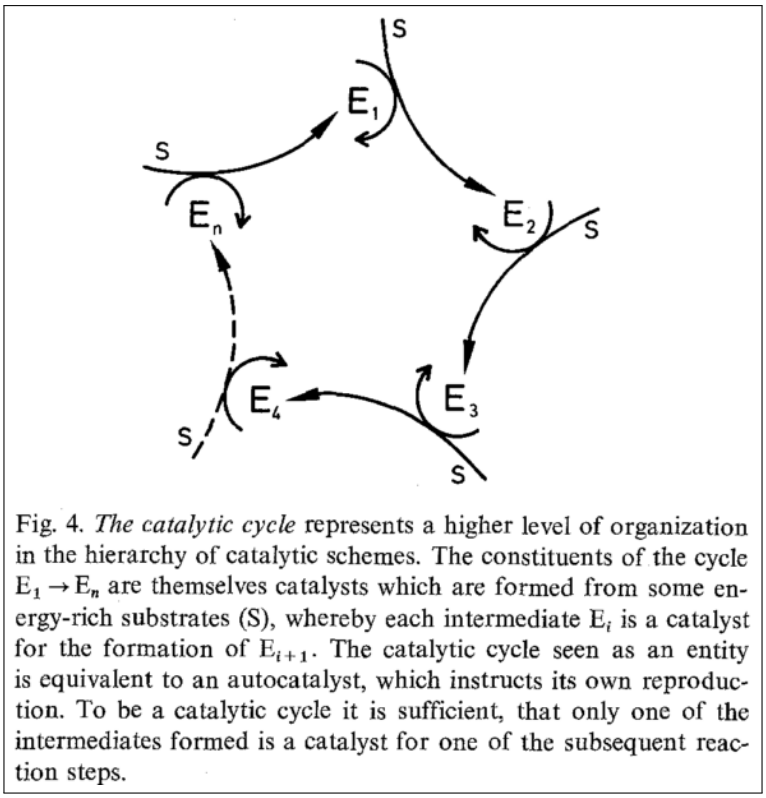



Fig. 4. *The catalytic cycle* represents a higher level of organization in the hierarchy of catalytic schemes. The constituents of the cycle $E_1 \rightarrow E_n$ are themselves catalysts which are formed from some energy-rich substrates (S), whereby each intermediate $E_i$ is a catalyst for the formation of $E_{i+1}$. The catalytic cycle seen as an entity is equivalent to an autocatalyst, which instructs its own reproduction. To be a catalytic cycle it is sufficient, that only one of the intermediates formed is a catalyst for one of the subsequent reaction steps.

Figure 4.3.8: Reproduced from Eigen and Schuster (1977)

> "A catalytic hypercycle is a system which connects autocatalytic or self-replicative units through a cyclic linkage."(Eigen and Schuster, 1977)

Although this concept of 'hypercycle' is widely regarded, this quote is the most definitive description that Eigen and Schuster give and yet is open to multiple interpretations. "Autocataltic or self-replicative units" can be interpreted as autocatalytic sets which have been discussed previously, or as other hypercycles. The description of a "linkage" is less clear; it is usually interpreted as molecules shared between the self-replicative units — e.g. a by-product of one autocatalytic set that is a substrate of another autocatalytic set. The term "linkage" can also be interpreted as catalysts that are in multiple autocatalytic sets.

Note that a hypercycle itself is a self-replicative unit, and therefore one hypercycle can act as a unit within a larger hypercycle. This is seen as an important feature of hypercycles, and has been suggested as explaining some of the multi-layer organization seen in biology — cells forming tissues, forming organs, forming individuals, forming populations, etc.

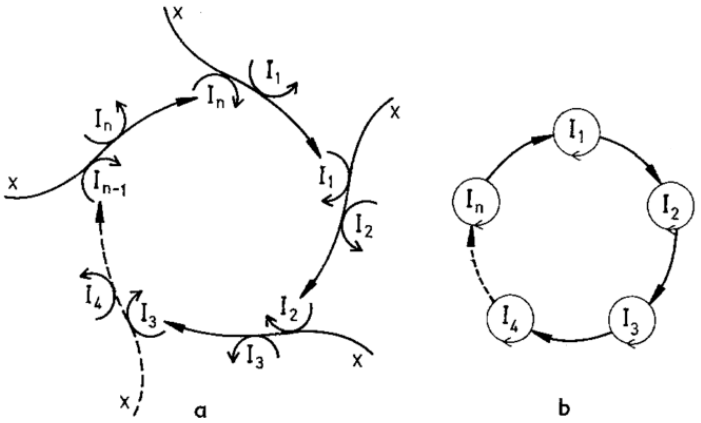



Fig. 7. *A catalytic hypercycle* consists of self-instructive units $I_i$ with two-fold catalytic functions. As autocatalysts or—more generally—as catalytic cycles the intermediates $I_i$ are able to instruct their own reproduction and, in addition, provide catalytic support for the reproduction of the subsequent intermediate (using the energy-rich building material X). The simplified graph (b) indicates the cyclic hierarchy

Figure 4.3.9: Reproduced from Eigen and Schuster (1977)

#### 4.3.9.3 Linkage in hypercycles

There are a number of subtle details of hypercycles that are often misinterpreted. Although the quote above specifically describes "autocatalytic or self-replicative units", in the caption to figure 9 (reproduced in figure 4.3.10) Eigen and Schuster state "it is sufficient if one of the intermediates posses autocatalytic or self-instructive function[h] presuming the other partners feed back upon it via a closed cyclic link". The first quote implies that every component in the hypercycle must be autocatalytic in the absence of any of the other components. However, the caption of figure 9 (reproduced in figure 4.3.10) contradicts this by stating that only one component must be autocatalytic. For the purposes of this document, the caption is treated as correct and reconciled with the quote through the term "linkage". As "linkage" is also ill-defined, in this work a linkage between two molecules is defined as a sequence of reactions that each share one or more products with the subsequent reaction's reactants, such that the first molecule in the linkage is a reactant of the first reaction and the last molecule in the linkage is a product of the last reaction. The linkage can consist of zero reactions, in which case it must be

[h]This is assumed to be equivalent to self-replication.

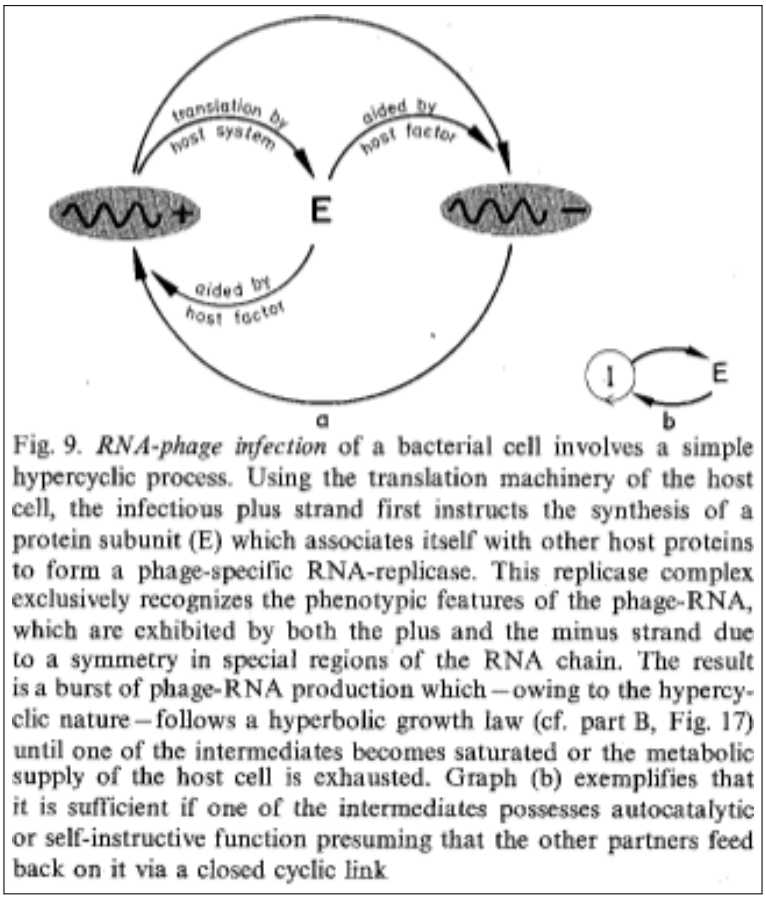



Fig. 9. *RNA-phage infection* of a bacterial cell involves a simple hypercyclic process. Using the translation machinery of the host cell, the infectious plus strand first instructs the synthesis of a protein subunit (E) which associates itself with other host proteins to form a phage-specific RNA-replicase. This replicase complex exclusively recognizes the phenotypic features of the phage-RNA, which are exhibited by both the plus and the minus strand due to a symmetry in special regions of the RNA chain. The result is a burst of phage-RNA production which – owing to the hypercyclic nature – follows a hyperbolic growth law (cf. part B, Fig. 17) until one of the intermediates becomes saturated or the metabolic supply of the host cell is exhausted. Graph (b) exemplifies that it is sufficient if one of the intermediates possesses autocatalytic or self-instructive function presuming that the other partners feed back on it via a closed cyclic link

Figure 4.3.10: Reproduced from Eigen and Schuster (1977)

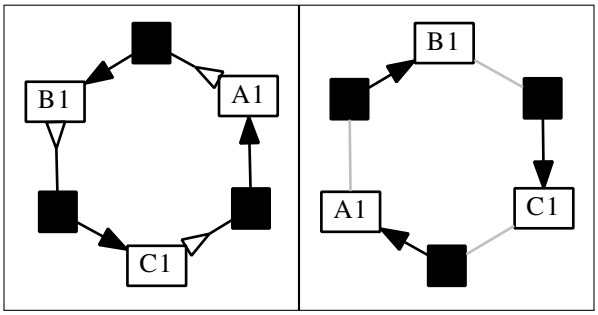
the same molecular species that is in both linked units.

The use of the qualifier "cyclic" in "cyclic linkage" describes the overall topology that the linked units are connected into, rather than a property of any individual linkage.

#### 4.3.9.4 Cycles and loops

A source of confusion about hypercycles is the use of the terms 'cycle' and 'catalytic'. A collection of molecules that form a loop through shared products/reactants of reactions (i.e. each molecule has a linkage to itself through the other molecules) does not count as a 'catalytic cycle' as defined by Eigen and Schuster. This is the case even though each repetition of the loop does not consume the molecules that that compose the loop overall, and if the loop increases the rate of the one or more reactions overall.

A consequence of this error is that what would be hypercycles are linked




(a) Loop (b) Cycle

Figure 4.3.11: Examples of loops and cycles on the left and right retrospectively. Note that for each molecular species in the loop the previous molecular species is consumed as a reactant and that in the cycle it is not consumed and acts as a catalyst instead.

loops. Such structures may be interesting in their own right, but they are not hypercycles by the intended definition. Instead, the term "hyperloop" should be applied.

An example of misinterpreting the term 'hypercycle' is in Dittrich and di Fenizio (2007). In this case, none of the components of the so-called 'hypercycle' are capable of self-replication without other components of the hypercycle. This means that Dittrich and di Fenizio (2007) describes a catalytic cycle (autocatalytic set) rather than a hypercycle. Although for the purposes of their work the distinction was not critical, it serves to demonstrate how such terms can be misinterpreted without clear definitions.

The presence of one or more hypercycles within an Artificial Chemistry would be a very interesting property. However, due to there unclear definition and the potentially computationally intensive task of searching for them within a given Artificial Chemistry of significant size and richness, there has been little success in discovering hypercycles within published Artificial Chemistries.

### 4.3.10 Chemical Organization Theory

Chemical Organization Theory (Dittrich and di Fenizio, 2007) was developed to categorise sets of molecular species in some types of Artificial Chemistry. Chemical Organization Theory is designed for "algebraic chemistries" which is the subset of Artificial Chemistries that are explicit §4.1.2.1, symbolic §4.1.1.1, aspatial §4.1.3.1 and without reaction rate constants §4.1.2.3. In addition,

Chemical Organization Theory simplifies individual molecules to the presence/absence of molecular species. Using these presence / absence sets as building blocks, several properties are developed for classifying them, and these properties form a lattice hierarchy due to dependencies. The properties of molecular species presence / absence sets specified in Chemical Organization Theory are described below and their usefulness in the context of Artificial Chemistries for Artificial life is discussed.

#### 4.3.10.1 Closed

The simplest property Dittrich and di Fenizio apply to sets of presence / absence of molecular species is the property of 'closed'. Every set with this property contains all the molecular species that can be produced by any reaction amongst the molecular species of this set; i.e. there is no novelty. Given a set that does not have this property, a closed set containing the given set can be found by repeatedly adding all products of reactions that can occur between all combinations of molecular species in the set until a closed set can be found.

Whilst the requirement of this property can be understood as an analytical tool, in the context of Artificial Chemistry for Artificial Life, novelty and emergence are seen as important properties (see §4.2.2.1). In particular, implicit unbounded Artificial Chemistries may be practically impossible to close, as each additional molecular species adds more possible reactions that could generate additional molecular species. For example, consider a growing polymer; each reaction that grows the polymer by a monomer subunit enables another reaction of the polymer with another monomer subunit. This suggests that the application of Chemical Organization Theory to Artificial Chemistries exhibiting life-like properties may be problematic. Dittrich and di Fenizio do acknowledge this and suggest that the application should be limited to a particular subset of an implicit Artificial Chemistry — such as the reactions seen during a particular simulation, or those molecular species below a certain size.

#### 4.3.10.2 Semi-self-maintaining

The property Dittrich and di Fenizio term 'semi-self-maintaining' is applied to a set of molecular species within an Artificial Chemistry in which each

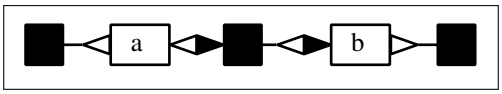



Figure 4.3.12: Example of a semi-self-maintaining but not self-maintaining algebraic chemistry. Based on §2.3.1 from Dittrich and di Fenizio (2007).

molecular species that is a reactant of a reaction more than a product of that reaction (consumed) is also in the products more than the reactants of another reaction (produced). In other words, for each molecular species that is consumed, it must also be produced. Note that this allows for direct catalysts to exist without being produced, as long as they are not consumed.

#### 4.3.10.3 Semi-organisation

A 'semi-organization' is a set of presence / absence of molecular species within an Artificial Chemistry that is both closed (§4.3.10.1) and semi-self-maintaining (§4.3.10.2).

#### 4.3.10.4 Self-maintaining

A semi-self-maintaining set may also have the property of being 'self-maintaining'. In a semi-self-maintaining set, every molecule which is consumed by a reaction is also produced. However, a semi-self-maintaining does not ensure that the set is stable. The example given is the reaction table below and in figure 4.3.12:

$$\begin{aligned} a &\rightarrow b \\ b &\rightarrow a \\ a &\rightarrow \\ b &\rightarrow \end{aligned}$$

Although all molecular species (a and b) are both consumed and produced (making the set of a and b semi-self-maintaining), both molecular species can spontaneously decay. This means that eventually both a and b will disappear from the system. Without the decay reactions (a $\rightarrow$, b $\rightarrow$) a and b can be interconverted without loss *ad infinitum*.

To address this, the term 'self-maintaining' applies to semi-self-maintaining sets of molecular species where a 'flux vector' exists that does not result in the overall decay of any molecular species. A 'flux vector' is the rate[i] of each

[i]Note that this is not a rate constant but a net rate

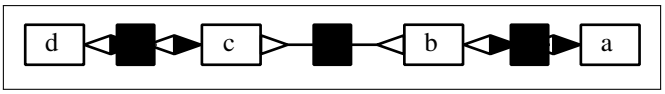



Figure 4.3.13: Example reaction network of a two self-maintaining sets ( {a,b}, {c,d} ) whose union is not self-maintaining.

reaction. For reactions whose reactants are not present in the set of molecular species, the rate of that reaction must be zero. For reactions whose reactants are present, the rate must be greater than zero.

#### 4.3.10.5 Organization

The term 'organization' applies to sets of presence / absence of molecular species that are both closed (§4.3.10.1) and self-maintaining (§4.3.10.4). As a self-maintaining set is also a semi-self-maintaining set, an organization is also a semi-organization.

#### 4.3.10.6 Semi-consistent

As well as properties of sets of presence / absence of molecular species, several terms that apply to the algebraic chemistry as a whole are discussed. The first of these is that a 'semi-consistent' algebraic chemistry. This applies to algebraic chemistries where the union of any two semi-self-maintaining sets of presence / absence of molecular species is also semi-self-maintaining, and the the union of any two self-maintaining sets is also self-maintaining.

Although this property appears ubiquitous, examples can be constructed that do not posses this property. For example, consider the following algorithmic chemistry reaction table (shown graphically in figure 4.3.13).

$$\begin{aligned} a &\rightarrow b \\ b &\rightarrow a \\ c &\rightarrow d \\ d &\rightarrow c \\ b + c &\rightarrow \end{aligned}$$

There are two self-maintaining sets, {a,b} and {c,d}. The union of these ({a,b,c,d}) is not self-maintaining as b and c react to decay both reactants, which will causes molecules to disappear. This situation is analogous to the example for self-maintaining set given above (figure 4.3.12).

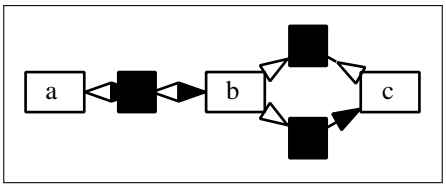



Figure 4.3.14: Example reaction network of a self-maintaining set ( {a,b}) whose closure ({a,b,c}) is not self-maintaining.

#### 4.3.10.7 Consistent

Building on semi-consistent, another possible property of algebraic networks is 'consistent'.

> "A semi-consistent algebraic chemistry is called consistent if the closure of a semi-self-maintaining set is semi-self-maintaining and the closure of a self-maintaining set is self-maintaining.(Dittrich and di Fenizio, 2007)"

The processes of 'closure' is the generation of the smallest closed set containing the set undergoing closure, e.g. by repeatedly expanding the set with all products that come from all possible reactions between members of the set.

As with the property 'semi-consistent', this description appears to be straightforward at first glance, but contains hidden traps. In particular, consider the example reaction table given below and in figure 4.3.14:

$$\begin{aligned} a &\rightarrow b \\ b &\rightarrow a \\ b &\rightarrow c \\ b + c &\rightarrow \end{aligned}$$

In this example, there is a self-maintaining set {a,b}. However, this set is not closed due to the reaction b $\rightarrow$ c. When c is added to the set {a,b}, a closed set {a,b,c} is generated. However, {a,b,c} is not self-maintaining due to the mutual decay of b and c (b + c $\rightarrow$ ).

Because of these potential issues with semi-consistent and consistent algebraic chemistries, it is difficult to determine if a particular algebraic chemistry has these properties. For larger systems, examining each possible set of molecular species is a computationally prohibitive task due to the $O(2^n)$ complexity where n in the number of molecular species. To address this,

Dittrich and di Fenizio classify algebraic chemistries based on the types of reactions they contain, and then assert that all algebraic chemistries in some classes are consistent. These classes are described below.

#### 4.3.10.8 Catalytic flow system

A 'catalytic flow system' is an algebraic chemistry with two additional limitations; every molecule spontaneously decays to nothing; no molecule is otherwise consumed by a reaction – molecules may only be catalysts or produced by reactions. A consequence of these features is that every semi-self-maintaining set is self maintaining, and therefore every semi-organization is an organization.

However, the limitations that a catalytic flow system has would appear to limit emergent life-like properties. For example, without decomposition the only way for a system to remove undesirables, such as waste products or parasites, is to wait for them to decay spontaneously. Further, conservation of mass is only possible if nothing is consumed and therefore nothing is produced; this prohibits interesting chemistries.

#### 4.3.10.9 Reactive flow system

A 'reactive flow system' is a generalization of a catalytic flow system. Every molecule must undergo spontaneous decay to nothing, but any other reaction is also possible. Unlike catalytic flow systems however, not every semi-self-maintaining set is also self-maintaining (and not every semi-organization is an organization).

#### 4.3.10.10 Reactive flow system with persistent molecules

Combining aspects of catalytic flow systems and reactive flow systems, a 'reactive flow system with persistent molecules' has two classes of molecules. Non-persistent molecules all undergo spontaneous decay to nothing as in a reactive flow system. Persistent molecules are never consumed by any reaction. Any catalytic flow system or reactive flow system is also a reactive flow system with persistent molecules, so Dittrich and di Fenizio prove that those types of systems are consistent by first proving that a reactive flow system with persistent molecules is consistent. Consistency is desirable because it means there must be a single largest organization that can be identified in a more computationally efficient manner.

Whilst this classification of algebraic chemistries is useful to Chemical Organization Theory, it is questionable whether reactive flow systems with persistent molecules is capable of containing life-like systems. The requirement that non-persistent molecules spontaneous decay to nothing is at odds with conservation of mass as well as the nutrient recycling thought to promote open-ended evolution within ecosystems though dynamic niches. In biological systems all molecules are consumed sooner or later, which suggests that there are no 'persistent molecules'.

#### 4.3.10.11 General reaction system

Algebraic chemistries that are not a reactive flow system with persistent molecules are termed 'general reaction systems'. These may or may not be (semi-)consistent, and therefore to find which sets of presence / absence of molecular species have which properties requires testing possible sets. As the number of sets is $O(2^n)$ with respect to the number of molecular species, exhaustive testing is not practical for any algebraic chemistry of a reasonable size.

#### 4.3.10.12 Conclusions

Whilst Chemical Organization Theory has a number of useful concepts, applying it in its entirety to every Artificial Chemistry is not feasible. Not only is the restriction to algorithmic chemistries of explicit symbolic representations limiting, but also much of the dynamics of an Artificial Chemistry (e.g. space, stochasticity, reaction rates) are removed as well.

However, unlike some other described properties of Artificial Chemistries, Chemical Organization Theory is an accurate and well-defined collection of properties and features. Therefore, while identifying interesting subsets (e.g. autocatalytic sets, hypercycles, etc. ) of an Artificial Chemistry using Chemical Organization Theory is impractical, classifying and comparing interesting subsets once they have been found by other methods may yield useful insights.

## 4.4 Requirements

The ingredients described in §4.1 are possible ways that an Artificial Chemistry can be constructed, and the properties described in §4.2 are desired characteristics of an Artificial Chemistry for Artificial Life. Using these some suggestions

| Property | Requirement |
|---|---|
| Autocatalysis | Catalysis |
| Autocatalytic Set | Catalysis |
| Catalysis | Direct Catalysis *or*, |
| | Indirect Catalysis *or*, |
| | Rate Catalysis |
| Rate Catalysis | Rate Constants |
| Rearrangement Reactions | Stereo-Isomers |
| Stero-Isomers | Isomers |
| Isomers | Intra-molecular space |
| Intra-molecular space | Structured molecular representation *or*, |
| | Sub-symbolic molecular representation |
| Unbounded Molecular Size / Structure | Implicit reactions |
| Cell Movement | Cells |
| Dynamic Cell Contents | Cells |
| Dynamic Cell Contents | Partially Permeable Membranes *and/or*, |
| | Catalysis of Position |
| Catalysis of Position | Catalysis |
| Catalysis of Position | *not* Aspatial Inter-Molecular Space |
| Cells | *not* Aspatial Inter-Molecular Space |

Table 4.4.1: Table of some requirement relationships between properties and/or ingredients of an Artificial Chemistry.

can be made about which properties require which ingredients. Relationships like these have been briefly discussed in the context of Chemical Organization Theory, but here such relationships are discussed in wider and more abstract context. Some suggested requirement relationships are represented in table 4.4.1

At the highest level, hypercycles require that multiple autocatalytic sets must be present. In order to contain autocatalytic sets an Artificial Chemistry must exhibit catalysis. Catalysis in an Artificial Chemistry may be in several different forms; direct or indirect, and absolute or by rate constant.

The properties proposed by Suzuki et al. have some links between them. Two of the properties ('dynamic cell contents' and 'cell movement') require that cells must exist, and a third ('partially permeable membranes') informs how such cells should be formed. In addition, the property of 'unbounded molecular size / structure' cannot be achieved with explicit reactions or with symbolic molecular representation; these would require an unbounded list of reactions and molecular symbolic respectively.

An Artificial Chemistry that uses implicit reactions allows for a number of other properties. Not only does it allow for 'unbounded molecular size / structure' by continuously enabling larger molecules to interact, but it also potentially enables 'novelty and innovation'. In an explicit reaction scheme, there can be no novelty because all reactions must be stated *a priori*. In order for an Artificial Chemistry to use implicit reaction the molecules must be structured or sub-symbolic — otherwise there is no molecular information for the implicit reactions to operate on. Compared to a structured molecular representation, a sub-symbolic molecular representation may be more able to provide 'emergent complex properties' as a consequence of the lower-level representation.

The mixing of molecules within an Artificial Chemistry can also impose significant limitations on its properties. In an Artificial Chemistry without inter-molecular space there can be no cells because the 'inside' and 'outside' cannot be separated. Further, in order for a chemistry to exhibit 'rate constants', 'stochasticity' or 'divergent reactions' the mixing algorithm used must be compatible.

Bringing these relations between properties and/or ingredients together, it can be suggested that a sub-symbolic molecular representation with an implicit reaction scheme would allow for a large number of properties. These include catalysis, autocatalytic sets, and hypercycles as well as novelty and innovation through unbounded molecular size / structure.

## 4.5 Conclusions

In this chapter, a wide range of concepts and terminology have been discussed. In particular, possible ingredients of Artificial Chemistries were separated into three groups — molecular representation, reactions, and mixing algorithms — and alternatives within these groups were described and compared.

As well as ingredients, potential properties of Artificial Chemistries were examined and described from two main groups: design properties and emergent properties. Design properties are properties that are built into an artificial Chemistry as part of its specification. Two main groups of design properties were previously described in Suzuki et al. (2003) and Hickinbotham et al. (2010). In addition to design properties, emergent properties were also examined in this chapter. These range from simple properties such as

reversible reactions and direct catalysis to more complicated properties such as autocatalytic cycles and RAF sets, as well as hypercycles and the properties described in Dittrich and di Fenizio (2007).

Building on these ingredients and properties, some guidelines for an Artificial Chemistry for Artificial life were outlined. The next chapter describes an Artificial Chemistry that satisfies the guidelines.

# Chapter 5

# RBN-World: Sub-Symbolic Representation

In this work I have discussed how Artificial Chemistries can be used as an approach for Artificial Life, and outlined some of the previous work in this area as well as a preliminary investigation that reimplemented one of them. Building on this, I described relevant properties, and ingredients that may provide them.

In this chapter I describe a new Artificial Chemistry named RBN-World. This was designed drawing upon the desirable properties and ingredients described in chapter 4 in order to construct a suitable environment for life-like systems to emerge spontaneously.

One of the key ingredients identified was the use of a sub-symbolic structured molecular representation (§4.1.1.3 & §4.1.1.2). However, in order to use this ingredient a sub-symbolic representation must be selected. To assist in that process a number of properties can be hypothesised to be desirable.

It should be noted that, as previously discussed (§4.1.1.3) a sub-symbolic representation may be expressed as a symbolic shorthand. The example used was "caffeine" rather than "1,3,7-trimethyl-1$H$-purine-2,6(3$H$,7$H$)-dione3,7-dihydro-1,3,7-trimethyl-1H-purine-2,6-dione". Although sub-symbolic representations will be used where possible, symbolic shorthand may be unavoidable in places.

## 5.1 Sub-symbolic Representation Properties

There are a large number of possibilities that could be used as a sub-symbolic representation within an Artificial Chemistry. However, this range can be reduced by the identification of a number of desirable properties, and then restricting the possible sub-symbolic representations to those Artificial Chemistries exhibiting the desired properties. As there is no previous work on sub-symbolic representations for Artificial Chemistries, the properties described here are an initial assessment and will need to be revised in the light of future work.

### 5.1.1 Computationally Tractable

One of the most important characteristics a sub-symbolic representation should have for use in an Artificial Chemistry is that it should be computationally tractable to calculate the properties of the representations. Deterministic sub-symbolic representations are also preferable as they can potentially be cached for future reuse.

### 5.1.2 Composability and Decomposability

Key properties of a sub-symbolic representation is that it should be both composable and decomposable. This means that not only can multiple representations be combined into one or more larger representations but that those larger representations should be able to break apart at some future point into smaller representations, not necessarily the same as those that the larger representation was formed from. In the context of an Artificial Chemistry this is in order to construct function groups and molecules, as well as allowing them to break apart - potentially in a different way to how they were formed.

### 5.1.3 Unlimited Size

It is preferable that there is not an *a priori* limit on the maximum size of a sub-symbolic representation. This is desirable so that it is possible for different emergent behaviours to appear at different size scales; a key property of biological molecules such as proteins.

### 5.1.4 Complexity and Emergence

The sub-symbolic representation used in an Artificial Chemistry should have emergent properties with complex behaviour. This means that the relationship between the structure of the sub-symbolic representation and one or more of its overall properties should be that similar structures generally have similar properties but in some cases small changes in structure have a large affect on the properties.

### 5.1.5 Alternative Representations

As this is a novel approach to Artificial Chemistries, the selection of a sub-symbolic representation requires a degree of exploration of the possibilities. This process can be made easier by selecting a sub-symbolic representation that has a number of similar alternatives because if the original selection does not work then the alternatives can be substituted into the Artificial Chemistry with relatively little effort, compared to changing to a very different sub-symbolic representation.

### 5.1.6 Graph Representations

Sub-symbolic representations that are graph-based, or can be formulated as a graph such as a character string, are promising for use in an Artificial Chemistry. Graphs can be composable and decomposable by adding and removed edges to form fully connected components. Further, there is not an *a priori* limit on the number of nodes and edges in a graph, which allows for a potentially unlimited size.

## 5.2 Random Boolean Networks

Based on the above description of desirable properties of sub-symbolic representations for use in an Artificial Chemistry, Random Boolean Networks (RBNs) were selected. RBNs (Kauffman, 1969) were originally designed to describe gene expression networks to demonstrate the emergence of multiple cell types, and subsequent work has continued in that vein. However, RBNs can also be seen as a collection of computationally efficient and deterministic

systems, with complex and emergent behaviours that can be summarised in a range of different properties.

### 5.2.1 Construction of a RBN

To describe RBNs, it is easiest to described how a RBN is constructed. A RBN consists of $n$ nodes, each of which has $k$ inputs from nodes (including itself and duplicate inputs), a Boolean state and a function mapping the Boolean state of the nodes $k$ inputs into the nodes updated Boolean state. RBNs can be created by the following procedure:

1. Create $n$ nodes. Fixed values of $n$ are used here, but alternatively $n$ could be drawn from a probability distribution for each RBN.

2. Assign an initial Boolean state to each node. Here all nodes had probabilities of 0.5 for an initial 'true' state and 0.5 for an initial 'false' state, but other alternatives are possible such as all nodes starting in the 'false' state.

3. Assign a $k$-input Boolean function to each node selected at random from all possible functions with equal probability. All $k = 2$ Boolean functions are shown in table 5.2.1.

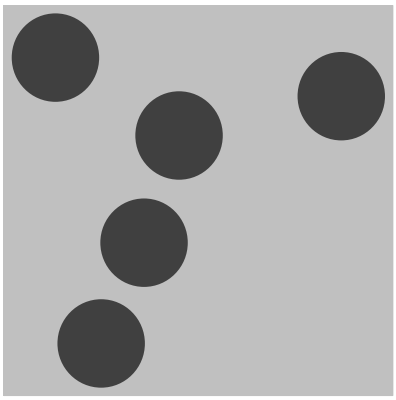
4. Assign $k$ inputs to each node. These are independently selected at random with equal probability from all $n$ nodes[a].

This is a completed RBN, similar to those described by Kauffman. There are a few minor differences due to interpretation of some subtle details, in particular if multiple parallel inputs from the same node are permitted or if inputs from self are permitted. This procedure is summarised in figure 5.2.1.

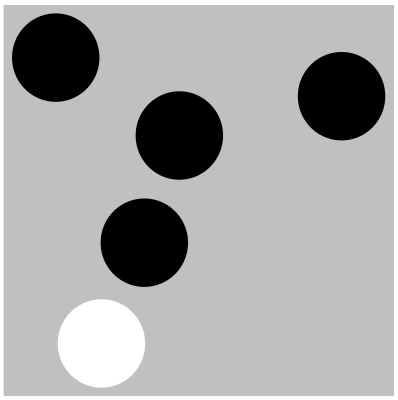
[a]In Kauffmans original work(Kauffman, 1969), inputs to each node could not be from themselves and were not independent.

| Input node states | | | | |
|---|---|---|---|---|
| (False, False) | (False, True) | (True, False) | (True, True) | |
| False | False | False | False |  |
| False | False | False | True |  |
| False | False | True | False |  |
| False | True | False | False |  |
| True | False | False | False |  |
| False | False | True | True |  |
| False | True | False | True |  |
| True | False | False | True |  |
| False | True | True | False |  |
| True | False | True | False |  |
| True | True | False | False |  |
| False | True | True | True |  |
| True | False | True | True |  |
| True | True | False | True |  |
| True | True | True | False |  |
| True | True | True | True |  |

Table 5.2.1: Symbols used to represent all two-input Boolean functions. The four left-hand columns indicate the output state for each of the four possible input states. The symbols can be divided into four quadrants, horizontally by the state of the first input and vertically by the state of the second input, that represent outputs of True/False if they are convex or concave respectively.

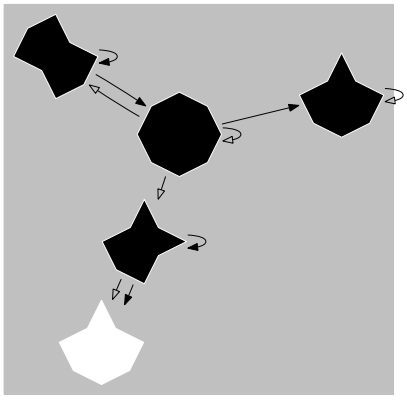


(a) Create $n$ nodes ($n = 5$)



(b) Assign initial Boolean state randomly. Black/white represent true/false repsectively.

(c) Assign function randomly. See table 5.2.1.



(d) Assign $k$ inputs per node ($k = 2$). Filled/empty represent first/second inputs respectively.

Figure 5.2.1: Steps for construction of an example random Boolean network.

### 5.2.2 RBN dynamics

Due to the updating process of RBNs, their state can change. The state of the RBN is a vector of the state of all its nodes, and the RBN's state is updated by synchronously applying the function of each node to the previous state of the inputs to that node. For example, consider figure 5.2.2.

Although the next state that a RBN will move to from its current state is deterministic, the same state can be reached from multiple prior states. This means that a RBN can reach the same future state from multiple different initial states. As a consequence of this, cycles where one or more states are repeatedly visited (see figure 5.2.2) will always be reached by repeated updates. The cycle of a RBN is an attractor and all the states that lead to a particular cycle are the basin of attraction of that cycle. A specific RBN can have multiple basins of attraction, and these can not only be of varying sizes but also with varying cycle lengths. The state space of a RBN can be exhaustively mapped to reveal all possible attractors, and the trajectory from a particular initial state identified within this state space. This is shown in figure 5.2.3 where all possible states of the RBN from figure 5.2.1 are the nodes and the single deterministic next state is indicated by an edge. The light grey edges describe how the state changes from the initial state (outlined) until it has completed the cycle.

From the state space shown in figure 5.2.3, there are several disconnected components, each of which forms its own separate basin of attraction. Without external influence, a RBN cannot move between basins of attraction. Each basin of attraction has a cycle at its centre (though that may be a cycle of one state) and different basins of attraction within the same RBN will have different cycles which may have different properties such as length.

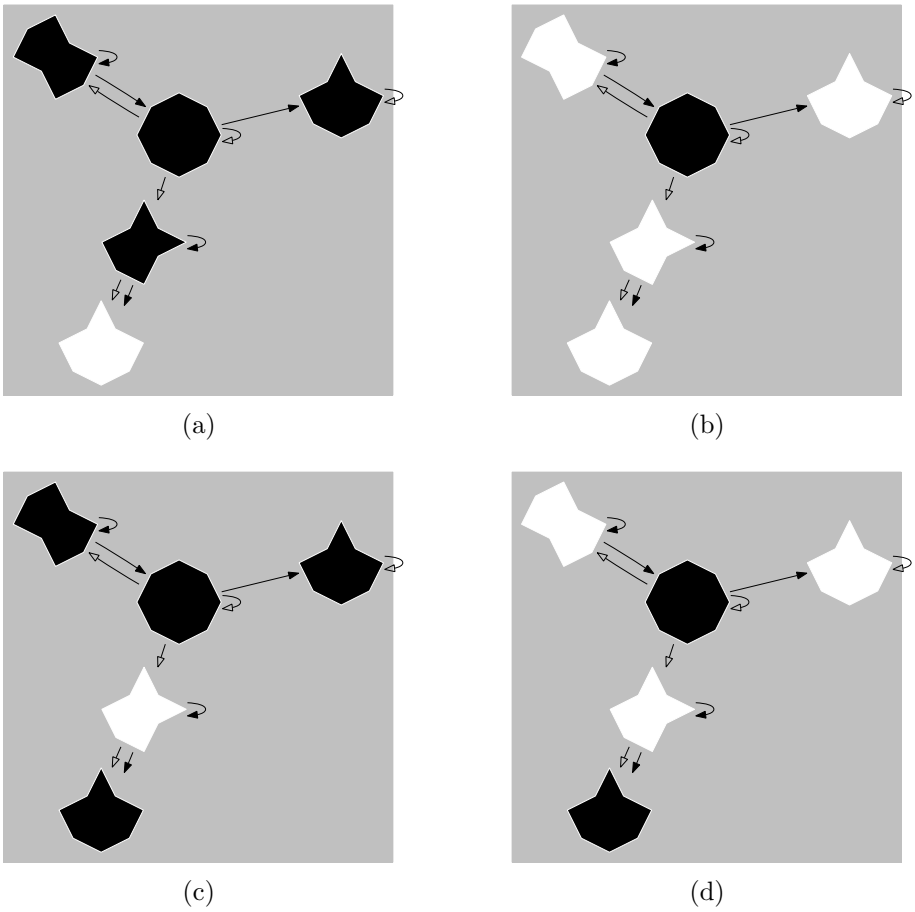



Figure 5.2.2: Change over time of an example random Boolean network. From step d the network repeats a cycle from step c.

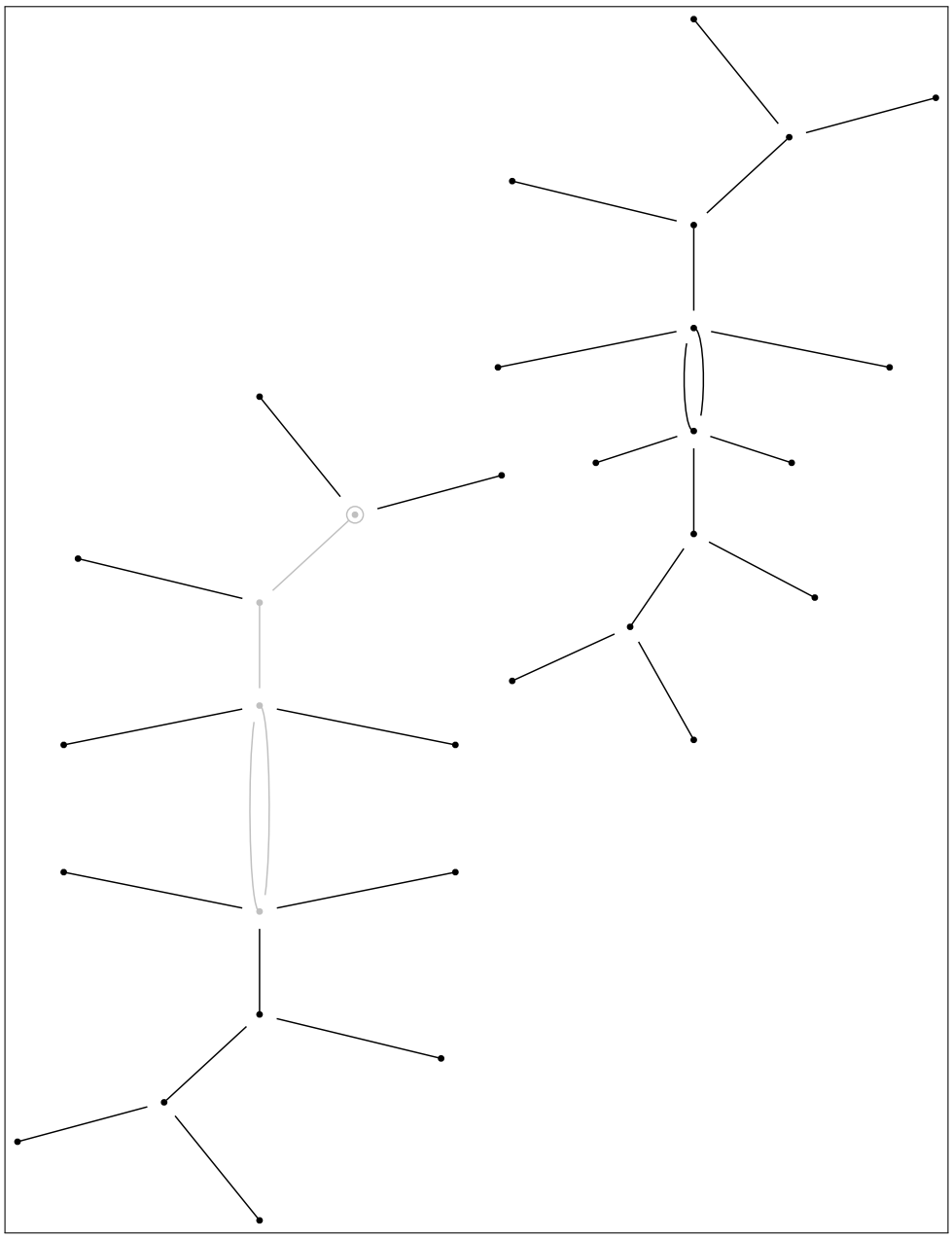


Figure 5.2.3: State space of an example random Boolean network. Points represent distinct states of the network and grey indicates the trajectory shown in figure 5.2.2 from the circled initial state.

### 5.2.3 RBN properties

In order for RBNs to be used in a sub-symbolic Artificial Chemistry, it is desirable that there should be several emergent properties that bond formation and breakage can be related to. In the original description of RBNs by Kauffman, the number of states on the cycle of the attractor (cycle length) was the emergent property measured. It is an emergent property because it is not easily predictable from any single or subset of nodes. The distribution of cycle lengths is known to be highly skewed with a median of approximately $0.3log(n)$ and a maximum of $2^n$ (Kauffman, 1969), but with a rugged distribution on fine scales. This can be seen in figure 5.2.4 (reproduced from Kauffman (1969)). This also demonstrates a many-to-one mapping between RBNs and cycle length; the same cycle length exists in RBNs with different structures and organizations. Further, the composition of the node states during the cycle is highly variable, some are very dynamic with most nodes changing in most steps and some are almost static with only a few nodes that change each step and/or few nodes changing at all.

In addition to the richness of the emergent properties, RBNs exhibit both sensitivity and resilience to changes. This can be seen by changing the state of one node from false to true and *visa versa* and comparing the cycle lengths before and after (figures 5.2.6 and 5.2.7). Although the vast majority of RBNs investigated did not change to a different basin of attraction ($710,685$ out of $1,000,000$) and those that did often had the same cycle length($225,164$ out of $289,315$). In some cases changing the state of one node resulted in a different value of the emergent property and these did not change in a simple and easily predictable manner.

These behaviours indicate that RBNs have rich and complex emergent properties, which was an important desired characteristic when selecting a sub-symbolic representation for use in an Artificial Chemistry.

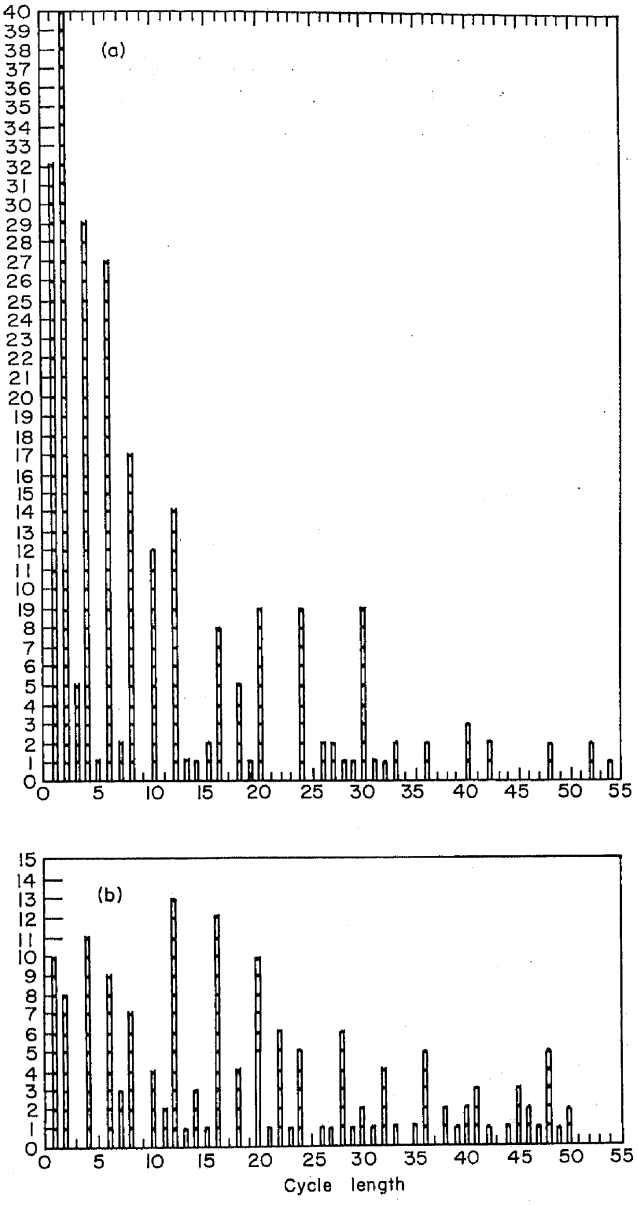



FIG. 3. (a) A histogram of the lengths of state cycles in nets of 400 binary elements which used all 16 Boolean functions of two variables equiprobably. The distribution is skewed toward short cycles. (b) A histogram of the lengths of state cycles in nets of 400 binary elements which used neither tautology nor contradiction, but used the remaining 14 Boolean functions of 2 variables equiprobably. The distribution is skewed toward short cycles.

Figure 5.2.4: Reproduced from Kauffman (1969)

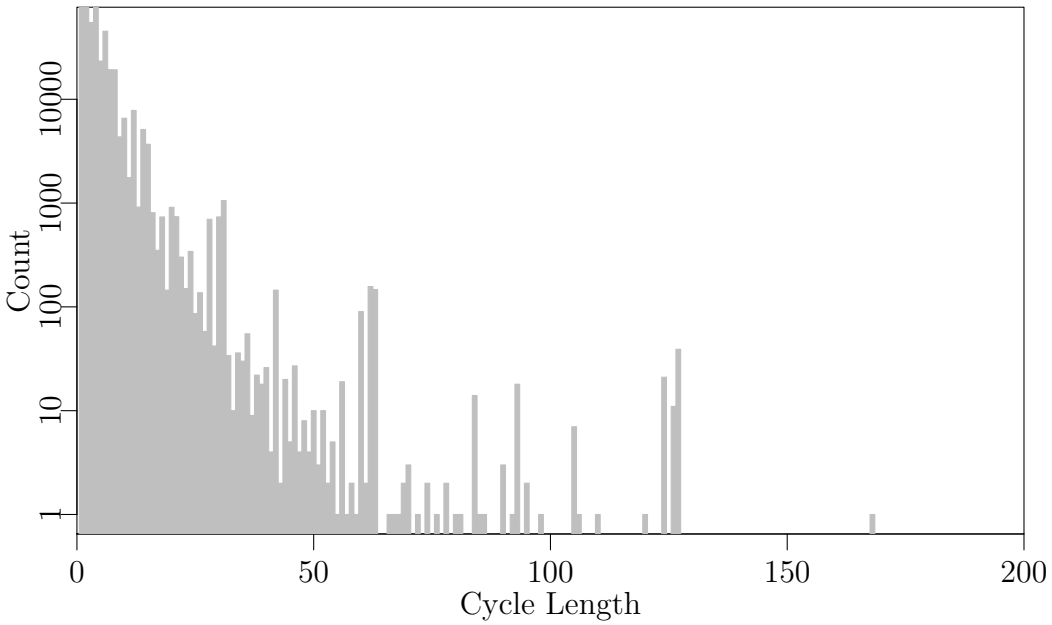



Figure 5.2.5: Graph of counts of cycle lengths seen from $1,000,000$ RBNs ($n = 10$).

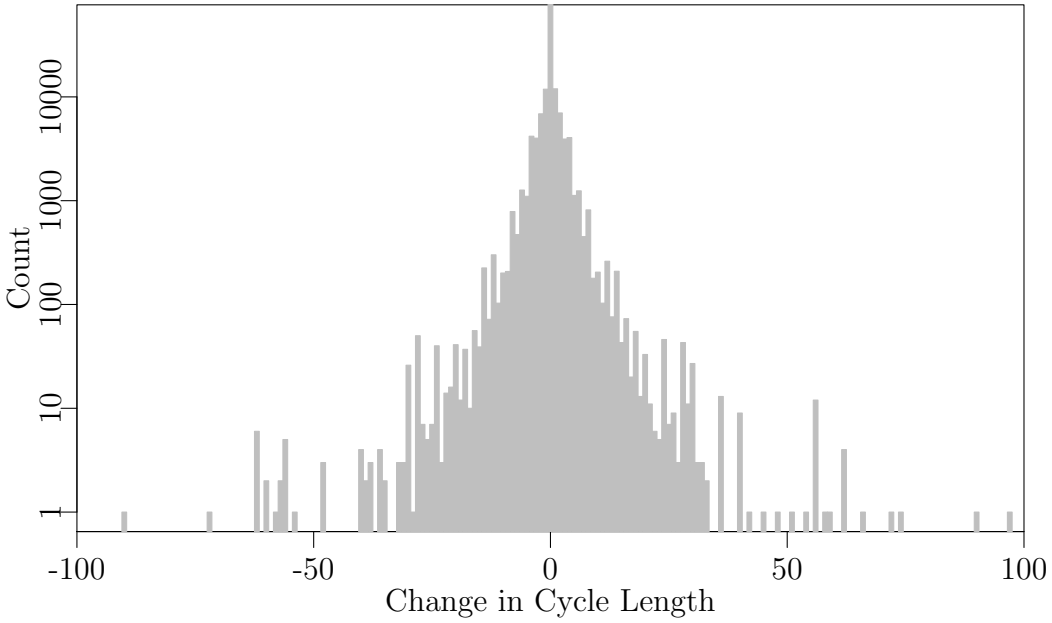



Figure 5.2.6: Plot of changes of cycle length in 1,000,000 RBNs when flipping the state of one node ($n = 10$).

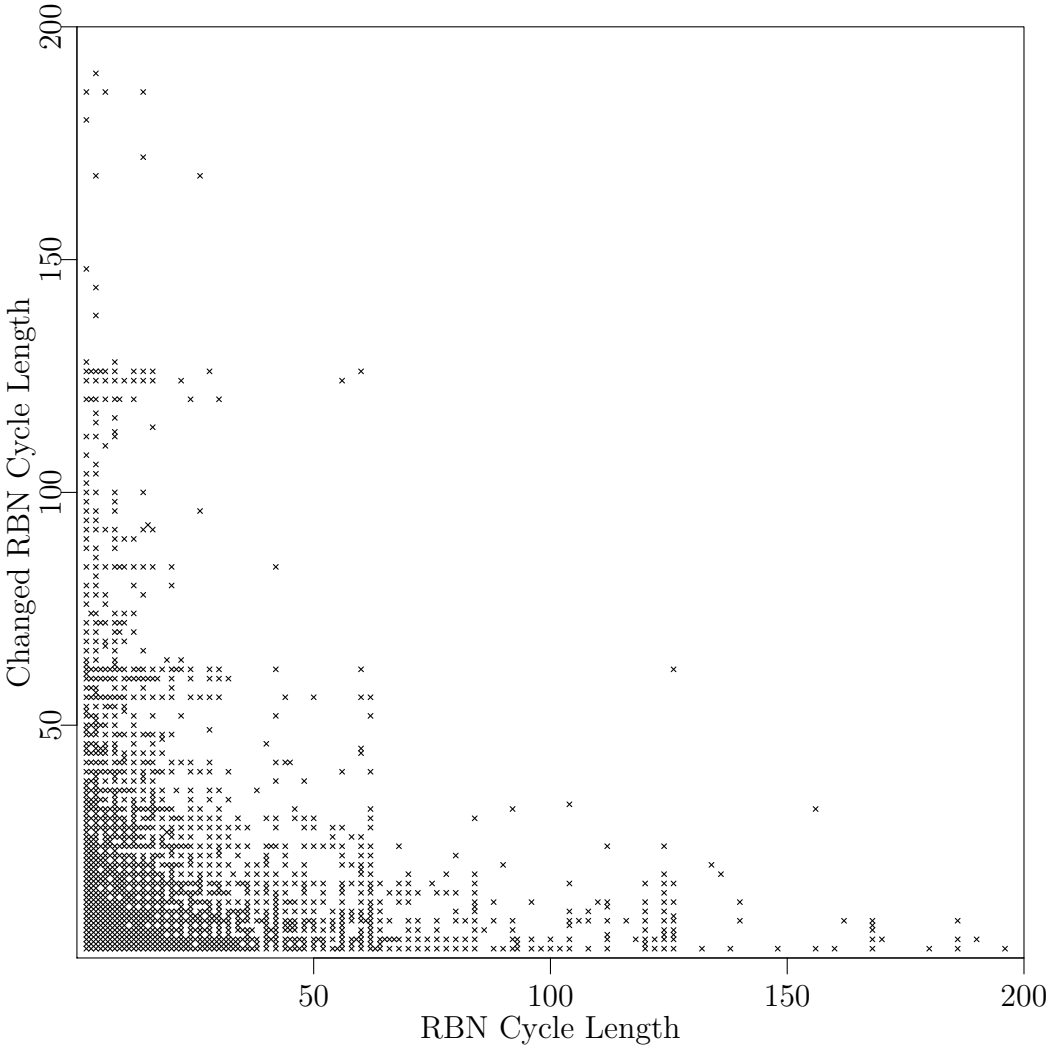



Figure 5.2.7: Scatter plot of cycle length before and after the state of one nodes is flipped in 1,000,000 RBNs ($n = 10$).

## 5.3 Bonding Random Boolean Networks

Although RBNs have the properties thought to be desirable for use as a sub-symbolic representation in an artificial chemistry, it is not obvious exactly how they should be composed and decomposed. To address this, a variant of RBNs was developed called Bonding RBNs (bRBNs) which is described below.

The bRBNs described here are not the only way that RBNs could have been adapted to be used in an Artificial Chemistry. However, as there have been no previous sub-symbolic artificial chemistries to perform these initial investigations a number of choices ad to be made based on limited information and investigation of alternatives — it is better to do something that might be wrong than to spend too much time investigating specific details initially.

### 5.3.1 Construction of a bRBN

Creating a bRBN involves taking a normal RBN and applying two additional steps (see figure 5.3.1):

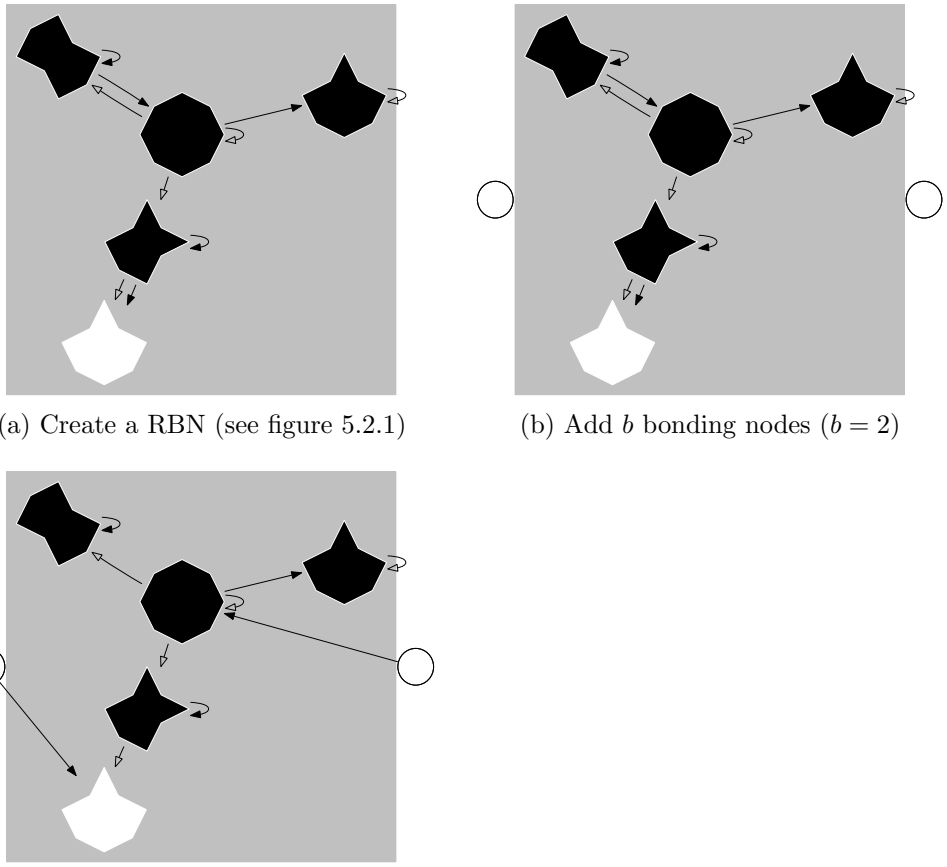
1. Add $b$ bonding nodes ($b = 2$). These are nodes that do not have a Boolean function or any inputs and have a fixed Boolean state (initially false).

2. Redirect one input to each node. For each bonding node, one of the inputs to a non-bonding node (selected at random) is set to be from that bonding node.

This process produces bRBNs that are similar to RBNs whilst also having additional features useful for composition and decomposition in a sub-symbolic Artificial Chemistry.

### 5.3.2 bRBN dynamics

As with RBNs, a bRBN can be updated by simultaneously applying the Boolean function of each node to the state of that nodes inputs. This is demonstrated in figures 5.3.2 and 5.3.3, which are the bRBN equivalents of figures 5.2.2 and 5.2.3. The computational tractability and deterministic dynamics that were desirable features of RBNs for use in a sub-symbolic structures molecular representation are still present in the bRBNs. However, one of the



(a) Create a RBN (see figure 5.2.1)

(b) Add $b$ bonding nodes ($b = 2$)

(c) Redirect one random input to each bonding node

Figure 5.3.1: Steps for construction of an example bonding random Boolean network.

reasons why RBNs are selected was their rich complex emergent properties; these may have been disrupted by the addition of bonding nodes to make bRBNs.

In order to confirm that the changes require to make bRBNs have not disrupted the underlying rich complex emergent properties, an equivalent experiment to Kauffman's (figure 5.2.4) was conducted. 1,000,000 RBNs and bRBNs ($n = 10$) were created at random and for each of them the length of the cycle was measured. This distribution can be seen in figure 5.2.5 and 5.3.4; the extreme tails are not plotted for clarity. As in Kauffman's work, this produces a highly skewed distribution with a long tail and a rugged distribution on fine scales in RBNs. When bRBNs are investigated, the distribution is different ($p \ll 0.05$) but bRBNs have a similar overall distribution and, importantly, a

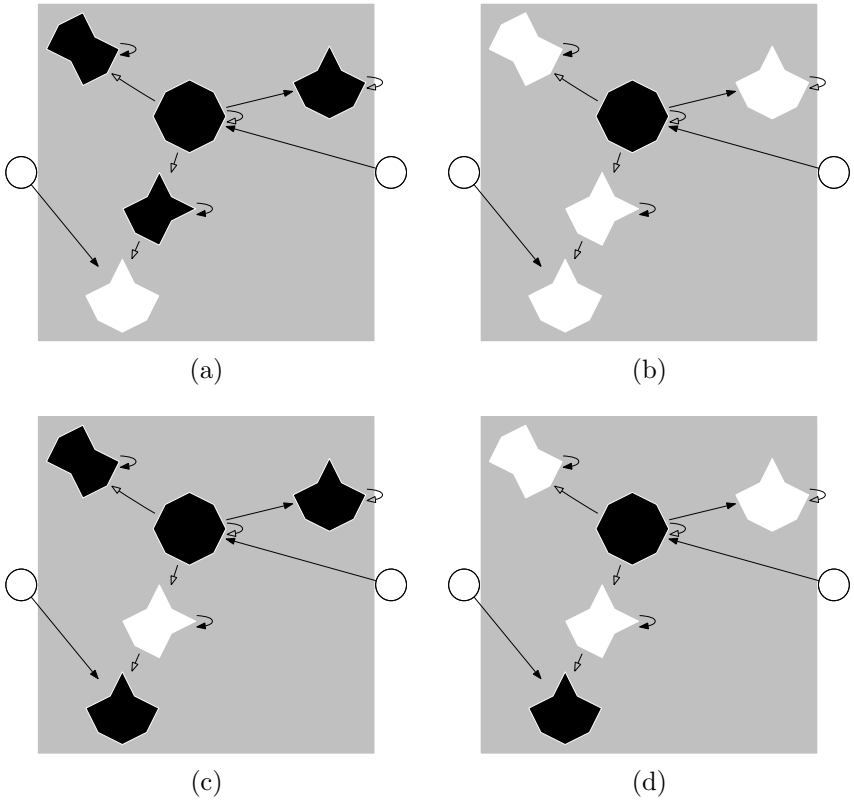



Figure 5.3.2: Change over time of an example bonding random Boolean network.

similar rugged distribution on fine scales.

A further feature of RBNs that was also desired for bRBNs was their sensitivity and resilience to changes. This was investigated by changing one of the bonding sites on the bRBN from an 'empty' state where its value is always false to a 'filled' state where its value is always true; cycle length before and after this change was compared and can be seen in figures 5.3.5 and 5.3.6. As with the equivalent experiment with RBNs shown in figures 5.2.6 and 5.2.7, bRBNs exhibited complex responses to perturbation.

These results demonstrate that bRBNs have the rich and complex emergent properties desired for a sub-symbolic representation in an Artificial Chemistry.

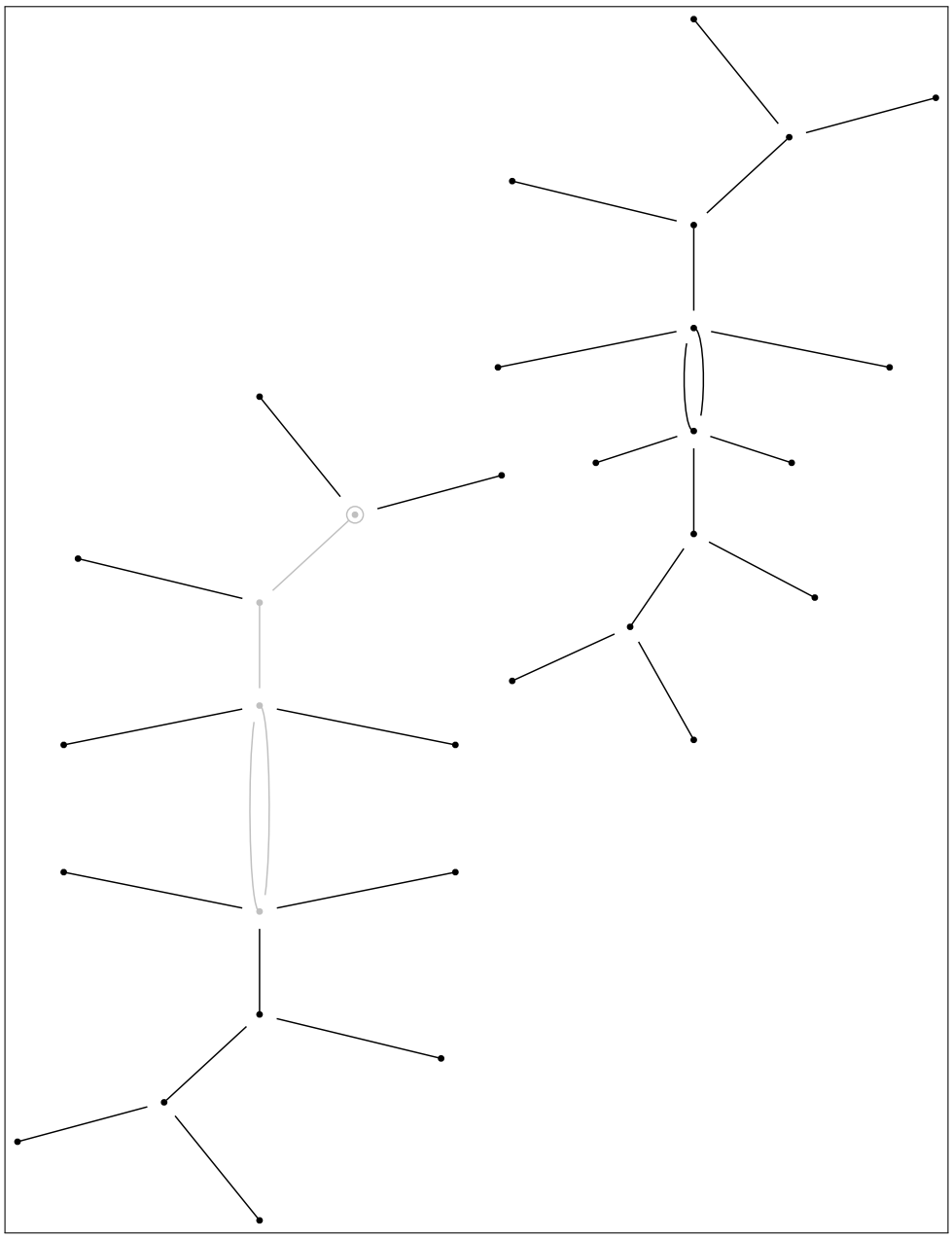


Figure 5.3.3: State space of an example bonding random Boolean network. Points represent distinct states of the network and grey indicates the trajectory shown in figure 5.3.2 from the circled initial state.

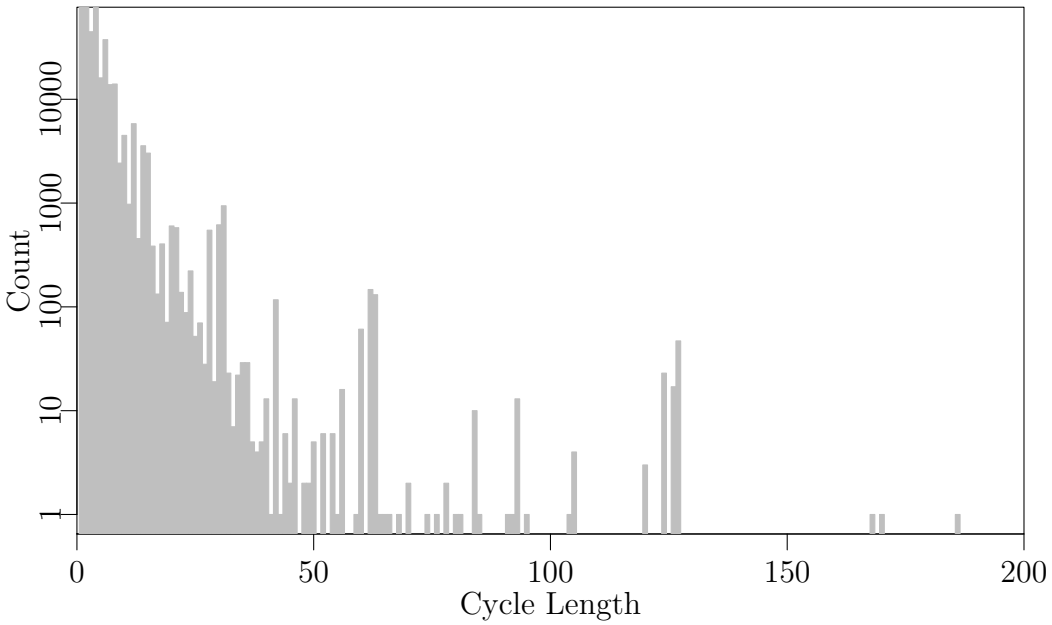



Figure 5.3.4: Graph of counts of cycle lengths seen from 1,000,000 bRBNs ($n = 10$, $b = 2$).

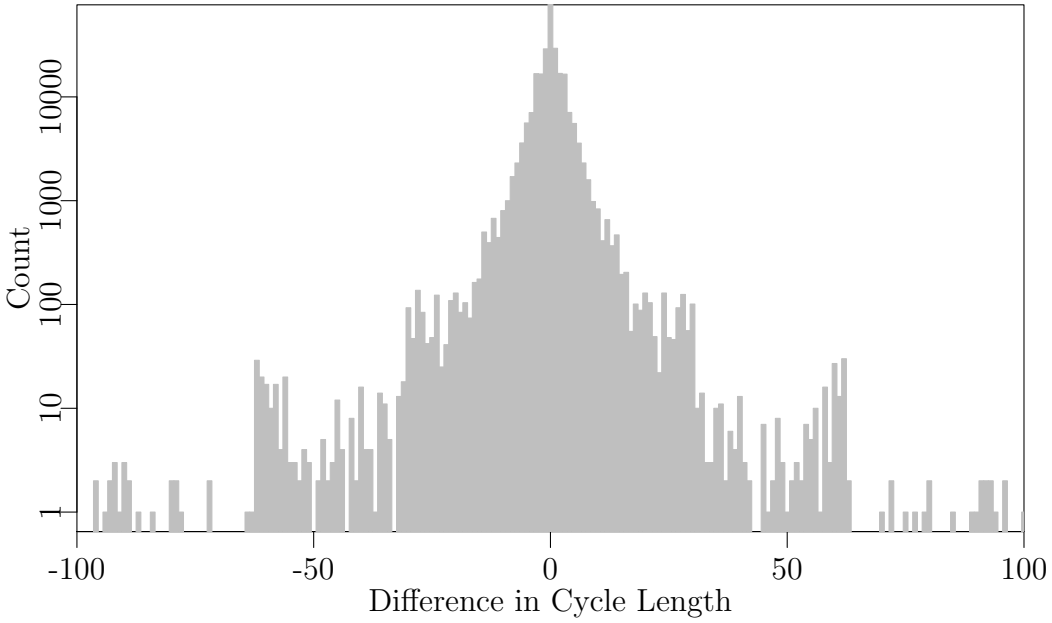



Figure 5.3.5: Plot of changes of cycle length when one bonding node filled from 1,000,000 bRBNs ($n = 10$).

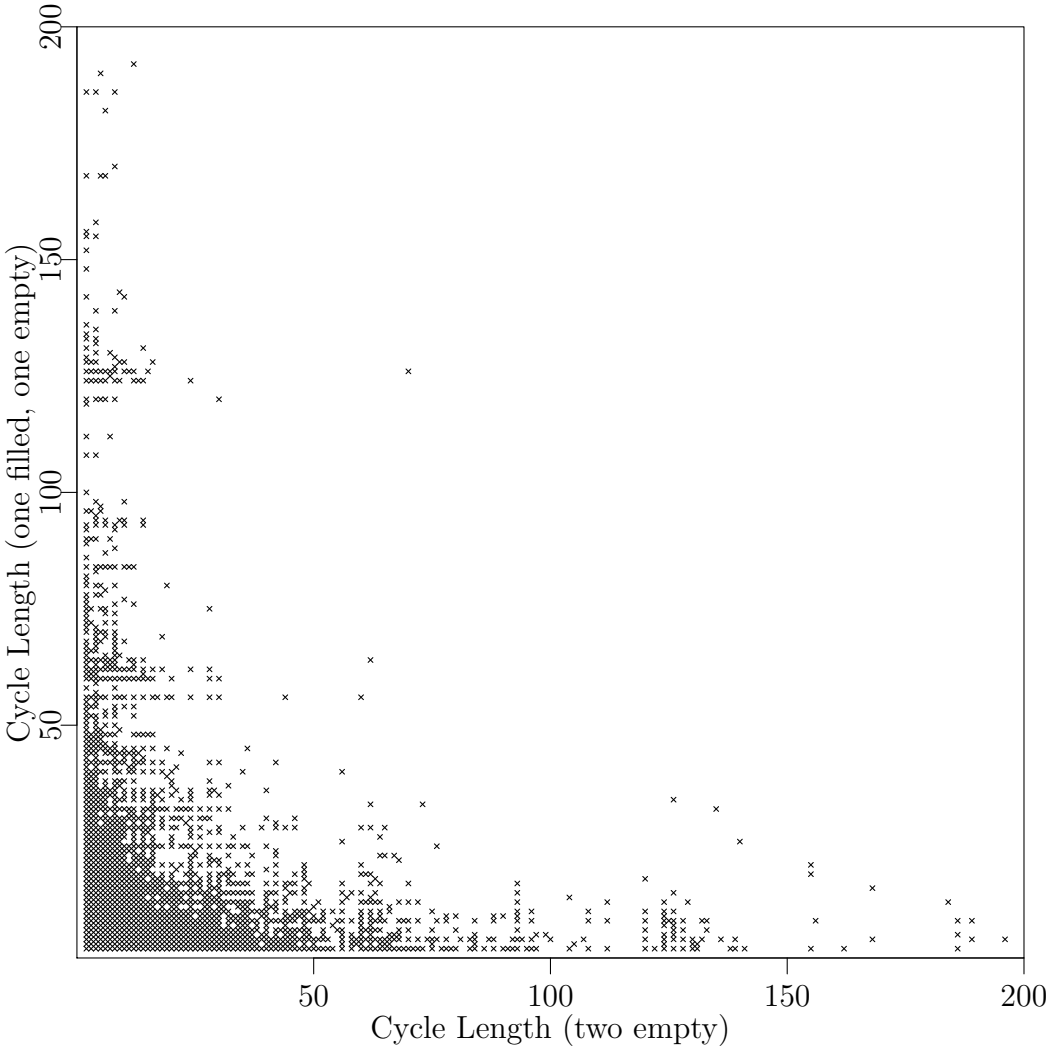



Figure 5.3.6: Scatter plot of cycle length before and after when one bonding node is filled from 1,000,000 bRBNs ($n = 10$).

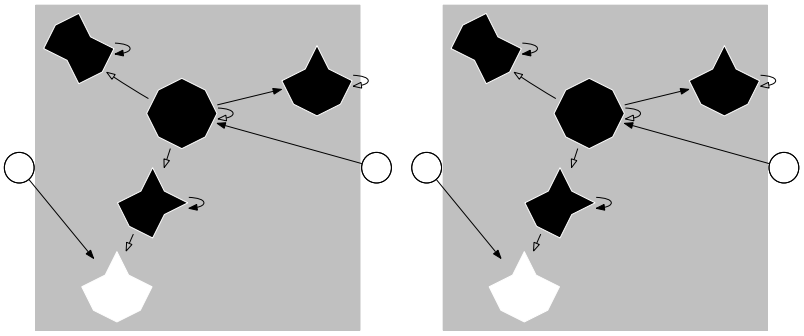
### 5.3.3 Composing bRBNs

A key requirement for a sub-symbolic molecular structures in Artificial Chemistry is a way to compose the sub-symbolic components into larger structures. In the bRBN representation used in RBN-World, the process is described below and shown in figure 5.3.7.

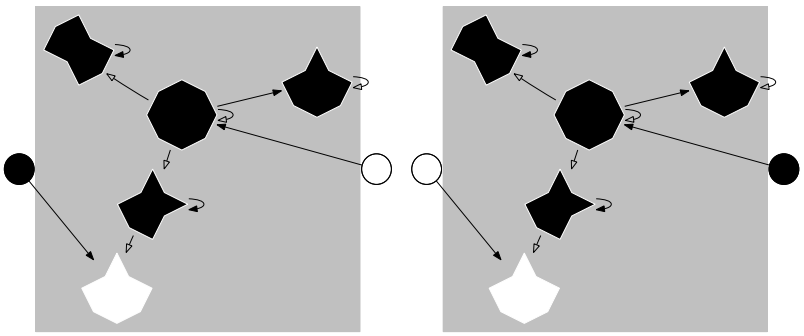
1. There must be two bRBNs to be composed, and each bRBN must have at least one bonding node.
2. Select one bonding node on each bRBN
3. For input to a node from a bonding node, replace it with an input from the other node that takes an input from a bonding node. Remove the now unused bonding nodes.

The outcome of this process is itself equivalent to a bRBN and therefore the composition process can continue to act upon it. Examples of this are shown in figure 5.3.8. As composite bRBN can be composed further to make larger structures there is no upper limit on the size of the bRBNs and no intrinsic restrictions on which bRBNs can be composed. This is desirable for use in an Artificial Chemistry as these attributes can be emergent properties of the Artificial Chemistry rather than *a priori* limits from the sub-symbolic representation.

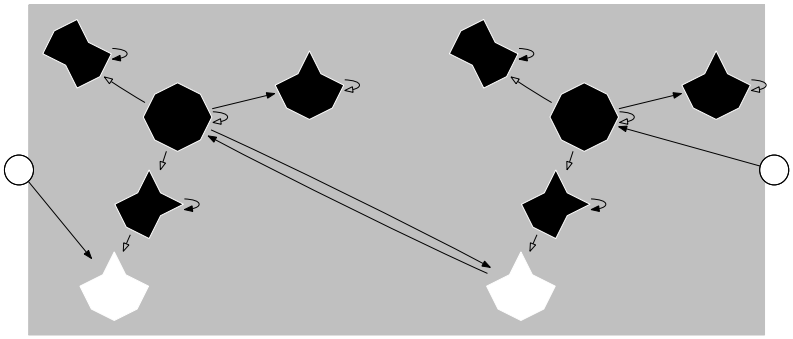
As with the individual bRBNs, composite bRBNs have dynamics from simultaneously updating each node, in the same manner as the individual bRBNs. This includes basins of attraction and cycles, as well as related emergent properties. This is tested by generating 1,000,000 two-atom bRBNs ($n = 10$) and measuring their cycle length, as in figures 5.2.4, 5.2.5 and 5.3.4. The results of the composed bRBNs are shown in figure 5.3.9; as with figure 5.3.4, the distribution is not the same as in figure 5.2.5 but it maintains the desirable features of being a rich complex emergent property. Indeed, it is important for an Artificial Chemistry that larger molecules do not have the same properties as smaller ones as this allows for emergent behaviours to appear at different structural scales due to the differences in size.



(a) Initially there must be two or more bRBNs to form composite

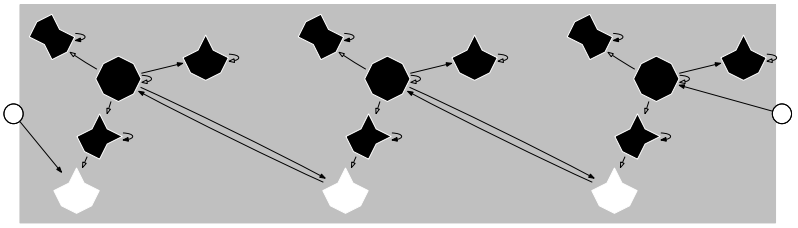


(b) Select a bonding node in each composing bRBN

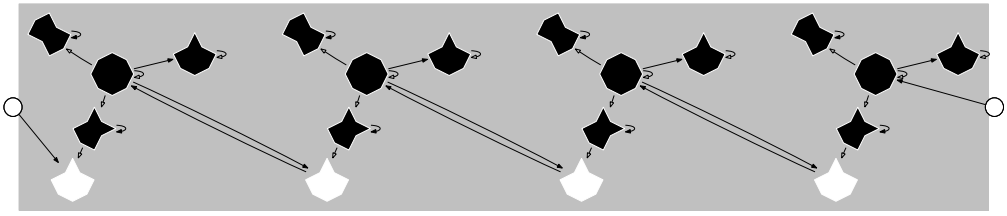


(c) Redirect one input to each bonding node to be from the other node that takes input from the selected bonding node

Figure 5.3.7: Example of the steps of composing two bRBNs into a single larger bRBN.



(a) A bRBN composed of 3 bRBNs



(b) A bRBN composed of 4 bRBNs

Figure 5.3.8: Example of repeated construction of larger bRBNs.

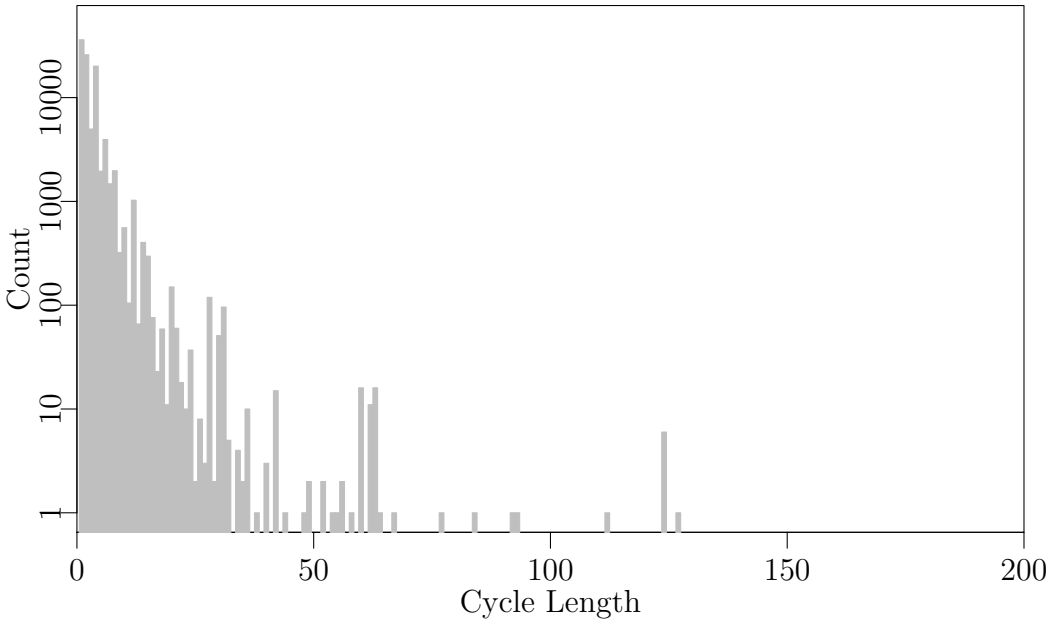



Figure 5.3.9: Graph of counts of cycle lengths seen from 1,000,000 bRBNs composed from two smaller bRBNs ($n = 10$, $b = 2$).

### 5.3.4 Decomposing bRBNs

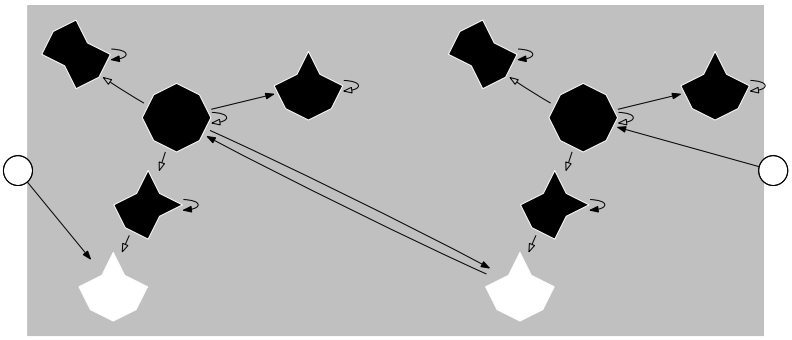
In order to balance composition, decomposition should be possible in a sub-symbolic representation suitable for an Artificial Chemistry.

However, in the composition scheme described above, it is not possible to determine with certainty the structure of the original bRBNs before they were composed and therefore they cannot decompose with those units guaranteed to be intact and therefore atomic. The replacement of the bonding node with reciprocal inputs (shown in step 5.3.7c) creates a subgraph structure that is rare but not impossible to find elsewhere. If this was identified as a place where they could be decomposed, then the atomic units of the original bRBNs could be recreateable. Additionally, bRBNs can have disconnected components which would make it difficult to define when a molecule had broken into smaller fragments. In such an Artificial Chemistry, rather than being a sub-symbolic structured representation, each of the individual bRBN nodes would have to be an 'atom' and therefore the Artificial Chemistry would be a simpler symbolic structured representation.

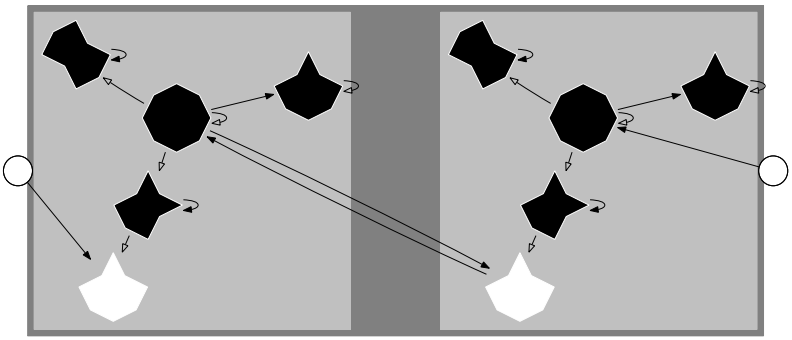
If the composition of the bRBNs is recorded in addition to their structure, as shown in figure 5.3.10, then they can be easily decomposed into their atomic components when required.

#### 5.3.4.1 Functional Groups

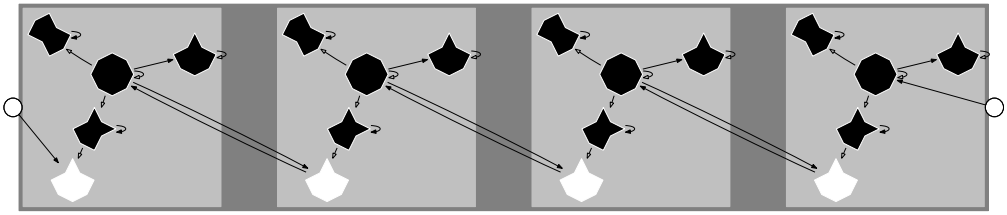
An additional benefit of decomposing bRBNs is that structures within a composed bRBN can be formed. For example, one of the simplest strucutres is a composed bRBN that is itself part of a composed bRBN, shown in figure 5.3.11. These intermediate composed bRBNs have properties that are the result of their smaller-scale composing bRBNs, but are not necessarily the same as the properties of those composing bRBNs. This means that an Artificial Chemistry built using composed bRBNs with recording of structure has the potential to allow the emergence of functional groups (§4.2.3.1) as molecular properties, where those functional groups are represented as intermediate composed bRBNs.



(a) Compsed bRBN without components recorded



(b) Compsed bRBN with components recorded



(c) Larger composed bRBN with components recorded

Figure 5.3.10: Examples of composition of a bRBN both without (a) and with (b and c) recording components.

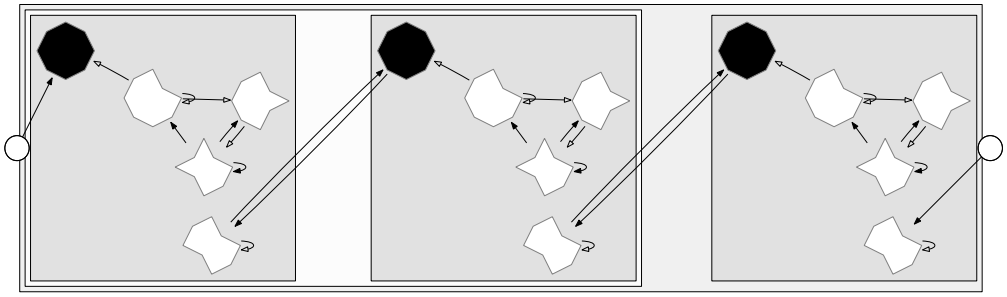


Figure 5.3.11: Example of composition of multiple composed bRBNs.

## 5.4 Conclusion

In this chapter RBNs were described, and then extended to bRBNs for use in Artificial Chemistries. Not only do bRBNs have rich complex internal dynamics, but those dynamics can be changed by composition processes that can form other, larger bRBNs.

Using bRBNs as a molecular representation ingredient of an Artificial Chemistry, the next step is to describe the reaction ingredient. This will be done in the following chapter.

# Chapter 6

# RBN-World: A Sub-Symbolic Artificial Chemistry

Now that bRBNs have been described as a sub-symbolic representation, an Artificial Chemistry using bRBNs can be described - RBN-World. This is not one single Artificial Chemistry, but rather a related and overlapping collection of Artificial Chemistries that have a range of different details within their construction, and consequently may have a range of emergent life-like properties.

The aim of these Artificial Chemistries is to produce a rich and dynamic collection of molecules with emergent properties that may lead to the emergence of life-like properties. After comparing the different ingredients of Artificial Chemistries and the properties they provide (as discussed in chapter 4), a sub-symbolic structured molecular representation using bRBNs as atomic units appears to be promising.

This chapter describes the molecules and reactions of RBN-World. Chapter 7 describes some of the alternative Artificial Chemistries within RBN-World and identifies interesting ones. Chapter 8 uses one of those interesting Artificial Chemistries to discover rich emergent behaviours from a small subset of elements.

All molecular species in RBN-World are composed of bRBN atoms. These atoms can be joined by bonds to form molecules, which have a sub-symbolic representation as composite bRBNs. Reactions are pairwise interactions between two molecules (which may be of the same or different molecular species)

and involve the formation and/or breakage of bonds between atoms. Bonds only exist between bRBNs whose emergent bonding property meets the bonding criterion at all times; if the bonding criterion between two bonded bRBNs is not valid at any point then that bond is broken. Bonds can be formed between single atom bRBNs, or composite bRBNs forming function groups or molecules, or combinations of these.

To demonstrate RBN-World, several example reactions are shown of increasing complexity. Although there are many possible bonding properties and many bonding criteria that could be used, the initial examples in this chapter use the cycle length as the bonding property and equality as the bonding criterion. This means that bonds only exist between bRBNs that have the same cycle length. Additionally, for these initial examples there is only one element which is a bRBN with $n = 5$ and $b = 2$. These were the arbitrary choices used for early development; other similar Artificial Chemistries can exist with the RBN-World framework and are investigated in chapter 7.

## 6.1 Reaction algorithm

RBN-World is the first Artificial Chemistry to use a sub-symbolic molecular representation (§4.1.1.3) and uses an implicit reaction algorithm (§4.1.2.2) to determine the products for collections of potential reactants. All reactions are between two reactants; it is assumed that more complicated reactions can be expressed as a series of two-reactant reactions with intermediate structures.

The pseudocode for the main reaction algorithm is described in figure 6.1.1. There are three major functions used in this algorithm that are each described in subsequent figures; TestForInteraction in figure 6.1.2, MakeStableBond in figure 6.1.3, and Decomposition in figure 6.1.4. To summarize, first it is determined if a temporary interaction between the bRBNs occurs, if so then it either collapses immediately or is transformed into a more stable bond. At the end of the reaction algorithm, decomposition is determined in order to yield the final product(s). An in-depth explanation of the reaction process is given in the subsequent examples.

```
Require: a and b are the reactant molecules.
 1: abonding := pick an empty bonding site in a
 2: bbonding := pick an empty bonding site in b
 3: result := TestForInteraction(a, b, abonding, bbonding)
 4: if result != NULL then
 5:    abrbn, bbrbn := result
 6:    fill bonding site abonding in abrbn
 7:    fill bonding site bbonding in bbrbn
 8:    abrbnprop := BondingProperty(abrbn)
 9:    bbrbnprop := BondingProperty(bbrbn)
10:    if not BondingCriterion(abrbnprop, bbrbnprop) then
11:       empty bonding site abonding in abrbn
12:       empty bonding site bbonding in bbrbn
13:    else
14:       c := MakeStableBond(abrbn, bbrbn)
15:    end if
16: end if
17: if a permanent bond was formed then
18:    return Decomposition(c)
19: else
20:    return Decomposition(a), Decomposition(b)
21: end if
```

Figure 6.1.1: Pseudocode of implicit reaction algorithm.

## 6.2 Example reactions

As described previously, RBN-World is not a single Artificial Chemistry, but rather a collection of similar Artificial Chemistries with different properties. There is also a collection of reaction methods as there are several alternatives to different aspects. For example, the testing of bonds for decomposition could also be done after the bonding sites have been filled but before the bond forms a stable composite. The example reactions described below follow the reaction algorithm given above, but it is important to consider what could be done differently in an alternative Artificial Chemistry.

```
Require: abonding is a bonding site in molecule a
Require: bbonding is a bonding site in molecule b
Require: a != b
 1: abrbn := smallest bRBN containing abonding
 2: bbrbn := smallest bRBN containing bbonding
 3: abrbnprop := BondingProperty(abrbn)
 4: bbrbnprop := BondingProperty(bbrbn)
 5: while not BondingCriterion(abrbnprop, bbrbnprop) do
 6:   if not HasLargerbRBN(abrbn) and not HasLargerbRBN(bbrbn) then
 7:     return NULL
 8:   else if HasLargerbRBN(abrbn) and HasLargerbRBN(bbrbn) then
 9:     asize := GetSize(GetLargerbRBN(abrbn) )
10:     bsize := GetSize(GetLargerbRBN(bbrbn) )
11:     if asize == bsize then
12:       abrbn := GetLargerbRBN(abrbn)
13:       bbrbn := GetLargerbRBN(bbrbn)
14:     else if asize > bsize then
15:       abrbn := GetLargerbRBN(abrbn)
16:     else if asize < bsize then
17:       bbrbn := GetLargerbRBN(bbrbn)
18:     end if
19:   else if HasLargerbRBN(abrbn) and not HasLargerbRBN(bbrbn) then
20:     abrbn := GetLargerbRBN(abrbn)
21:   else if not HasLargerbRBN(abrbn) and HasLargerbRBN(bbrbn) then
22:     bbrbn := GetLargerbRBN( bbrbn )
23:   end if
24:   abrbnprop := BondingProperty(abrbn)
25:   bbrbnprop := BondingProperty(bbrbn)
26: end while
27: return abrbn, bbrbn
```

Figure 6.1.2: Pseudocode of TestForInteraction function.

**Require:** *abrbn* and *bbrbn* to have a temporary bond
1: **for all** smaller bRBN composing *abrbn* or *bbrbn* **do**
2: fill bondsites site in the smaller bRBN equivalent to the bonding sites used in the temporary bond.
3: **end for**
4: **while** HasLargerbRBN(*abrbn*) or HasLargerbRBN(*bbrbn*) **do**
5: **if** HasLargerbRBN(*abrbn*) and HasLargerbRBN(*bbrbn*) **then**
6: subtitute GetLargerbRBN(*abrbn*) and GetLargerbRBN( *bbrbn* ) for a bRBN composed of all components of GetLargerbRBN(*abrbn* ) and all components of GetLargerbRBN( *bbrbn* )
7: *abrbn* := previously larger bRBN of *abrbn*
8: *bbrbn* := previously larger bRBN of *bbrbn*
9: **else if** HasLargerbRBN(*abrbn*) and not HasLargerbRBN(*bbrbn*) **then**
10: subtitute GetLargerbRBN(*abrbn*) for a bRBN composed of all components of GetLargerbRBN(*abrbn*) and *bbrbn*
11: *abrbn* := previously larger bRBN of *abrbn*
12: **else if** not HasLargerbRBN(*abrbn*) and HasLargerbRBN(*bbrbn*) **then**
13: subtitute GetLargerbRBN(*bbrbn*) for a bRBN composed of *abrbn* and all components of GetLargerbRBN(*bbrbn*)
14: *bbrbn* := previously larger bRBN of *bbrbn*
15: **end if**
16: **end while**
17: *c* := create a new bRBN composed of from *abrbn* and *bbrbn*
18: **return** *c*

Figure 6.1.3: Pseudocode of MakeStableBond function.

```
Require: one molecule a
repeat
  for all brbn in a do
    if HasSmaller(brbn) then
      for all bond in GetBondsIn(brbn) do
        thisbrbn, thatbrbn := bond
        thisbrbnprop := BondingProperty(thisbrbn)
        thatbrbnprop := BondingProperty(thatbrbn)
        if not BondingCriterion(thisbrbnprop, thatbrbnprop) then
          add bond to tobreak set
        end if
      end for
    end if
  end for
  for all bond in tobreak do
    empty bonding sites involved in bond
    split all composing bRBNs that were composed of both bRBNs in
    bond
  end for
until no more bonds are broken
for all brbn do
  if HasLargerbRBN(brbn) then
    size := GetSize(brbn)
    largebrbn := GetLargerbRBN(brbn)
    largersize := GetSize(largebrbn)
    if size == largersize then
      if HasLargerbRBN(largebrbn) then
        set the larger bRBN of brbn to be GetLargerbRBN(largebrbn)
      else
        delete GetLargerbRBN(brbn)
      end if
    end if
  end if
end for
return one or more molecules
```

Figure 6.1.4: Pseudocode of Decomposition function.

### 6.2.1 Example Reaction: A + B → (A–B)

When an Artificial Chemistry in RBN-World is created, there are only a small subset of possible elements present and the only molecular species that exist are composed of single atoms. In order to make the examples easier to understand, the bRBN atoms are represented symbolically in the form A, B, etc. . For the purposes of these examples, these symbolic representations are not consistent between examples, i.e. A in the first example is not necessarily the same as A in any of the other examples. By default, Artificial Chemistries within RBN-World require two molecules for all reactions, although Artificial Chemistries that have different numbers of molecules in their reactions are conceivable. One possible collection of reactants is two atoms of different elements, and this can be summarised as A + B. The sub-symbolic structure of the reactants used for this example is shown in figure 6.2.1a.

The first step in determining whether two reactants form a bond is to select an empty bonding site in each reactant. This is done at random from all available empty bonding sites with equal probability. For this example reaction, the right-hand bonding site of one atom and the left-hand bonding site of the other were selected. Other combinations are possible, and may include rotating atoms.

Once the bonding sites have been selected, the value of the bonding property for each bRBN, those bonding sites are attached to is calculated; for this example RBN-World Artificial Chemistry cycle length is used for the bonding property and equality of bonding property values is used as the bonding criterion. This results in bonding property values of 1 for A and 1for B. The bonding criterion is tested for the two reactants, and in this case a bonding property value of 1 is equal to a bonding property value of 1. Therefore the bonding criterion is met and the bond formation can proceed.

One the bonding criterion has been initially satisfied, the process can move forward to the next stage in forming a bond. Each of the selected bonding sites is 'filled' and the bonding properties of the associated bRBNs is recalculated. As different bonding sites were selected in the different atoms, it is possible that they will move to different basins of attraction with potentially different bonding properties (cycle lengths). These bRBNs with filled bonding sites can be symbolically represented as A– and –B for the atoms on the left and right

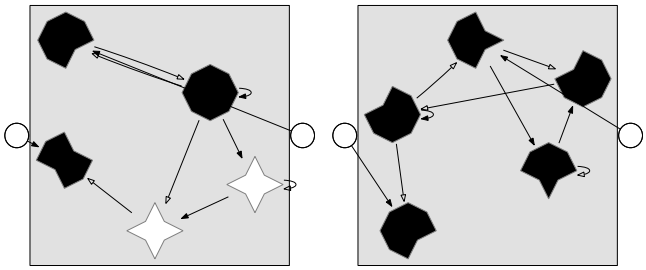
side of the newly forming bond respectively. In this example the atoms have a bonding property values of 1 and 1, which means that the equality bonding criterion is still met. This allows the reaction to proceed to the next stage of bonding formation, composing a molecule.

Once a bond has been formed, a composite bRBN for the new molecular species can be calculated. As described previously (§5.3.3), this is done by replacing inputs from the pair of filled bonding nodes with direct reciprocal inputs and the result is shown in figure 6.2.1c. The product is both a composite bRBN and the underlying atomic bRBNs and although the equivalent nodes in both representation have the same functions and inputs[a] the state of a node may be different in its molecular bRBN compared to its atomic bRBN. This is indicated in the sub-symbolic representation by subdividing the colour of the nodes vertically into a number of regions corresponding to composing the composed bRBN node state at the top and underlying atomic bRBN node state at the bottom. In this example there is no difference in the state of nodes between the molecular and atomic bRBNs, but this will not be so for other examples.

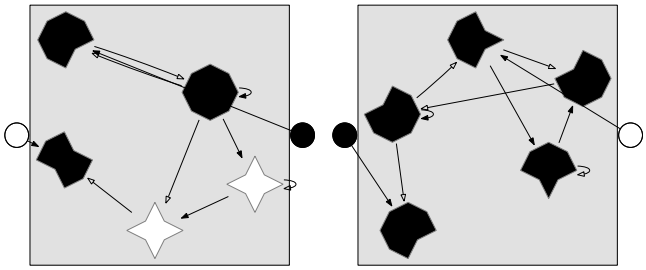
This reaction can be summarised symbolically as A + B → (A–B), though it has a richer sub-symbolic representation underlying both the individual atoms and the overall molecule.

Reactions of this type are a key foundation for interesting emergent properties in Artificial Chemistries in RBN-World. However, those involving only single atom reactants are a relatively specialised subset. Therefore, further examples will be presented extending into larger and more complicated molecular structures.

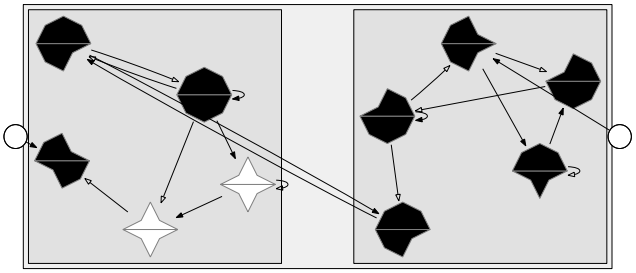
[a]In the case of nodes that were from a filled bonding node, this is easily determined by reciprocated inputs spanning atomic bRBN boundaries



(a) Initial reactants A + B



(b) Filling of the selected bonding sites in each reactant A– and –B



(c) Composition of the product (A–B)

Figure 6.2.1: Stages involved in the reaction A + B → (A–B).

### 6.2.2 Example Reaction: (A–B) + C → (A–B–C)

Once molecules have been formed, they can undergo further reactions with other atoms or with other molecules. The collection of reactants in this example reaction can be symbolically represented as (A–B) + C; note that the sub-symbolic bRBN for each elemental name (A, B etc. ) do not necessarily match the elements from either previous or subsequent example reactions. This is so that a range of examples can be presented covering all of the various features of Artificial Chemistries in RBN-World.

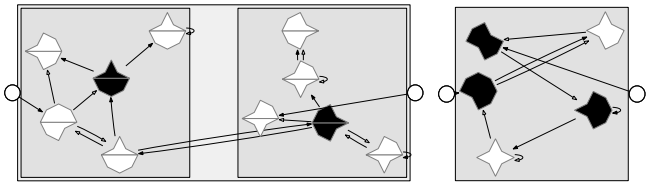
As in example 1, the first step in the reaction process is to select a bonding node in each reactant at random. For this example, the right-hand bonding site of (A–B) and the left-hand bonding site of C will be used, which can be seen in figure 6.2.2a.

Because one of the reactants is a molecule, the other reactant can react with either one of the atoms in the molecule or with the overall molecular structure. However, in order to reduce the combinatorial explosion of molecular structures in RBN-World all bonds are initially attempted at the smallest scale and if that does not interact then progressively larger scales are attempted to bond until both complete molecular structures are attempted.

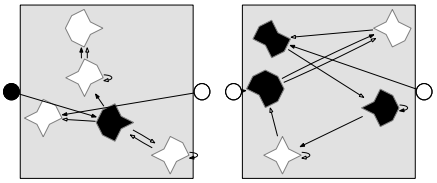
Therefore, the initial interactions are between the atoms –B and C, which can be seen in figure 6.2.2b. As the bonding property values of these meet the bonding criterion (1 = 1), the reaction can proceed to the next state of filling the selected bonding sites, as shown in figure 6.2.2c.

This leads to the recalculation of the bonding property values of –B– and –C. As the bonding criterion is still met (1 = 1) the bond forms. As one of the bRBNs participating in the new bond is already part of a larger composite structure (the bRBN –B is part of (A–B) ) rather than forming a new composite structure, the existing composite structure is extended to include the additional bRBN. This is symbolically represented as (A–B–C) and its sub-symbolic structure is shown in figure 6.2.2d.

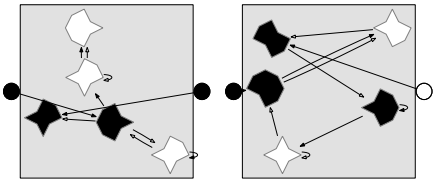
The final part of the reaction process is to ensure that none of the existing bonds were broken. In example 1, this was not an issue because there were no existing bonds. In this example however, the bond A–B existed and this may have been broken by the change from –B to –B–. The bonding property value of –B– is 1 and the bonding property value of A– is 1; therefore the



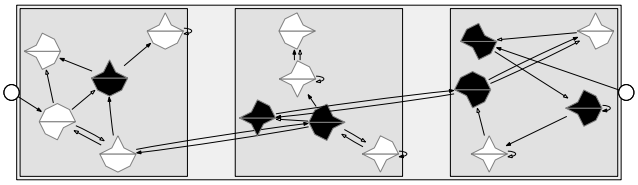
(a) Reactants (A–B) + C



(b) Initial interaction –B) + C



(c) Filling of the selected bonding sites in each reactant represented as –B)– and –C



(d) Composition of the product (A–B–C)

Figure 6.2.2: Stages involved in the reaction (A–B) + C → (A–B–C).

bonding criterion is valid and the bond is not broken. This means that the final product is (A–B–C) and the entire reaction can be overall represented as (A–B) + C → (A–B–C).

As with example reaction 1, this type of reaction is a key building block of rich Artificial Chemistries within RBN-World. The formation of larger structures from building blocks that are themselves the composed products of previous reactions is essential to the construction of macro-molecular structures such as enzymes or genomes.

### 6.2.3 Example Reaction: (A–B) + C → ((A–B)–C)

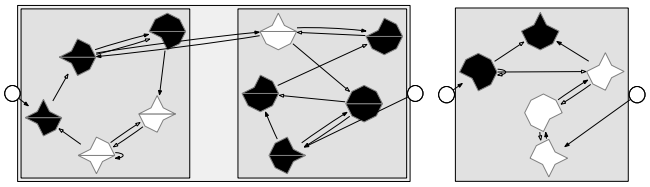
This example is similar to the previous example, but has some important differences. Although the reactants can still be symbolically expressed as (A–B) + C, the sub-symbolic representations are different (compare 6.2.2a with 6.2.3a).

As before, the initial interaction is between –B and C (see figure 6.2.3b). However, unlike the previous example, in this case the bonding properties of these bRBNs do not meet the bonding criterion. Because –B is part of the larger structure (A–B), a second interaction is attempted between (A–B) and C (see figure 6.2.3c). This second interactions does meet the bonding criterion, and therefore the reaction proceeds. As described in previous examples, bonding sites are filled and the bonding criterion retested between (A–B)– and –C which still meet the bonding criterion.

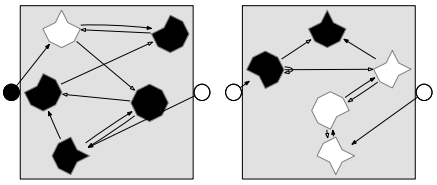
The composition of the product of this reaction is different from previous examples. Unlike in the previous example, neither reacting bRBN is part of a larger structure. However, one of the reacting bRBNs is itself composed of subunits. This means that the product is the result of bonding between the composed bRBN (A–B) rather than any of its individual components. Multi-level structures such as this are represented symbolically as ((A–B)–C) to differentiate from the simpler (A–B–C). The sub-symbolic representation of ((A–B)–C) is shown in figure 6.2.3e.

Testing that none of the pre-existing bonds have been broken is the final stage of the reaction. For this example, the only pre-existing bond was A–B. Because the composite (A–B) is now bonded, the right-hand bonding site of the B atom has been filled. This may have lead to changes in the bonding property such that the bonding criterion between A– and –B– is not longer valid. However, this is not the case here and therefore no bonds break.

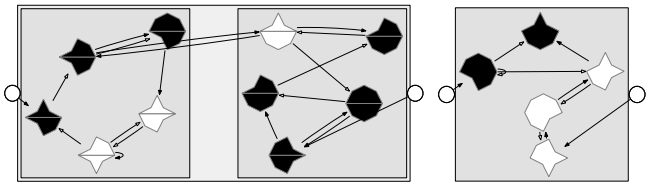
This example demonstrates a key feature of RBN-World, that of the emergent formation of functional groups of atoms within the molecular structure. Although the (A–B) group within the ((A–B)–C) molecule is a minimal example, it is nonetheless a functional group.



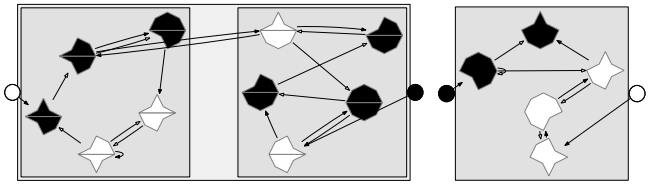
(a) Reactants (A–B) + C



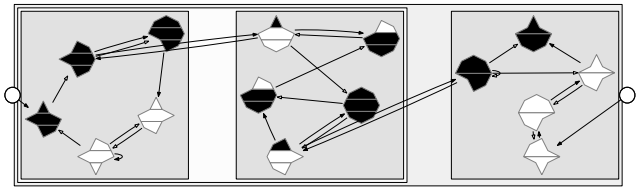
(b) Initial interaction –B) + C



(c) Second interaction (A–B) + C



(d) Filling of the selected bonding sites in each reactant (A–B)– and –C



(e) Composition of the product ((A–B)–C)

Figure 6.2.3: Stages involved in the reaction (A–B) + C → ((A–B)–C).

### 6.2.4 Example Reaction: (A–B) + (C–D) → ((A–B)–(C–D))

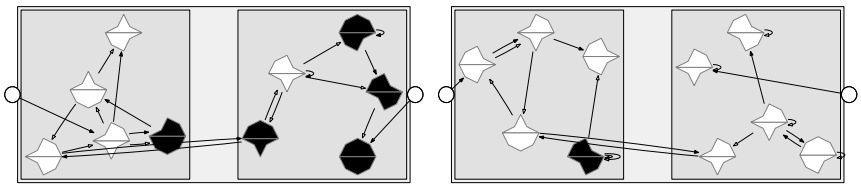
All of the examples so far have involved a single atom molecule as one of the reactants. However, reactions can occur between reactants that are of any size. To demonstrate this, this example involves two molecules that are each composed of two atoms, (A–B) and (C–D). The empty bonding nodes that have been randomly selected for this example are the right-hand of –B and the left-hand of C–.

As before, the next stage is to test the bonding criterion between the selected bRBNs. In this case, the bonding properties of –B and C– (see figure 6.2.4b) do not meet the bonding criterion. As at least one of the selected bRBNs is part of a larger composite structure, the reaction does not end here. However, there are multiple composite structures that could be used. In the default variant of RBN-World, the larger composite structure is used. However, both (A–B) and (C–D) are the same size (two atoms or 10 bRBN nodes depending on which measurement of size is used). In this case, the default variant of RBN-World selects both composite structures. Note that there are several points where alternative Artificial Chemistries could use different approaches; it is unknown if any of these alternatives would produce significantly different emergent high-level properties. The selection made for the default variant was an arbitrary choice.

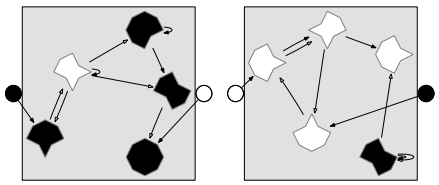
The second interaction is between (A–B) and (C–D) (see figure 6.2.4c). These do meet the bonding criterion, and therefore the bonding nodes are filled (see figure 6.2.4d). As both of the selected bRBNs are the top-level, a new composing bRBN is constructed rather than extending an existing one. This product, ((A–B)–(C–D)), is shown in figure 6.2.4e.

This example has shown two key points about RBN-World. Firstly, that molecular species of varying sizes are capable of being reactants. Although only molecules of two atoms are shown here as reactants, molecules of tens, hundreds, or even thousands of atoms are conceivably capable of being reactants once they are formed. The second key point about RBN-World that this example highlights is the potential for alternative Artificial Chemistries within RBN-World to have different ingredients. For example, rather than selecting the larger composite structure for subsequent interactions, an alternative could select the smaller composite structure. It is not obvious what effects this might

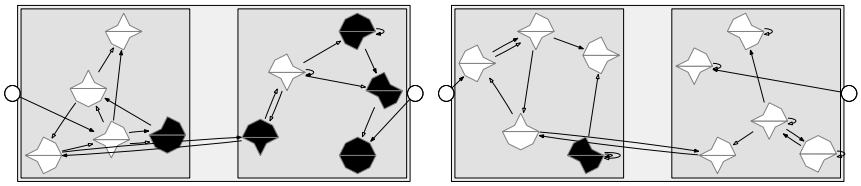
have on the overall properties of the Artificial Chemistry but, as discussed previously, these alternatives are important to distinguish which ingredients lead to desirable properties and which do not, or which have no effect. This is important in turn to understand how emergent life-like properties appear and which aspects of their components drive them.



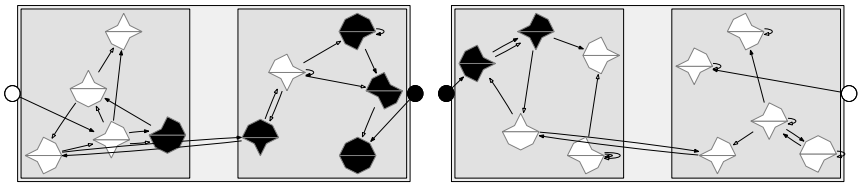
(a) Reactants (A–B) + (C–D)



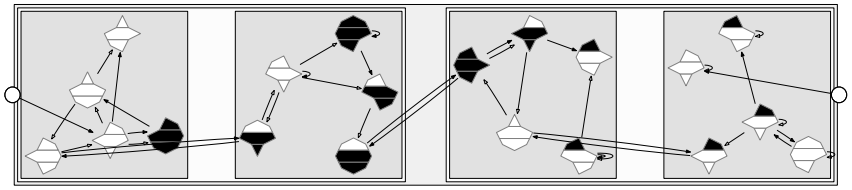
(b) Initial interaction –B) + (C–



(c) Second interaction (A–B) + (C–D)



(d) Filling of the selected bonding sites in each reactant (A–B)– and –(C–D)



(e) Composition of the product ((A–B)–(C–D))

Figure 6.2.4: Stages involved in the reaction (A–B) + (C–D) → ((A–B)–(C–D)).

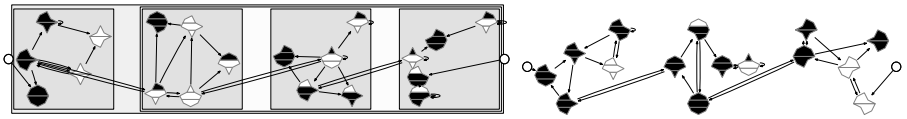
### 6.2.5 Example Reaction: (A–(B–C–D)) + ((E–F)–G) → ((A–(B–C–D)–E–F)–G)

Some of the examples shown previous have demonstrated the extension of a composing bRBN on formation of a bond. Other examples have shown composition of a new RBN from subunits. This example demonstrates the extension of an bRBN composed of composite bRBNs and extension of a composing bRBN, neither of which has been seen in any of the other examples.

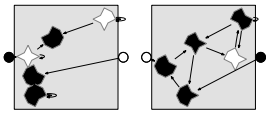
The reactants of this example are (A–(B–C–D)) and ((E–F)–G) (see figure 6.2.5a) and the right-hand bonding site of D and left-hand bonding site of E have been selected at random. The initial interaction is between –D and E– (see figure 6.2.5b). However these do not meet the bonding criterion. As D is part of the largest composite, the next interaction is between –(B–C–D) and E– (see figure 6.2.5c). These do meet the bonding criterion, and therefore the bonding sites become filled. After recalculating the bonding properties, –(B–C–D)– and –E– still meet the bonding criterion (see figure 6.2.5d). This means that the product ((A–(B–C–D)–E–F)–G) is formed.

As –E– is part of the composite –(E–F)– and –(B–C–D)– is part of the composite (A–(B–C–D))–, these composites merge to become (A–(B–C–D)–E–F)–. Note that although –(B–C–D)– is itself a composite, it is treated as a single bRBN for the formation of the (A–(B–C–D)–E–F)– composite. However, –(E–F)– is itself part of the composite –((E–F)–G) which expands to include the (A–(B–C–D))– structure resulting in the final composite bRBN of ((A–(B–C–D)–E–F)–G).

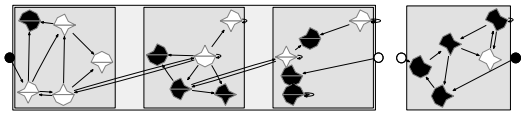
Reactions such as the one shown in this example allow the formation of progressively more complicated structures within larger molecules, though not all larger molecules have to be so complicated. It would also be possible, with different elemental bRBNs to have reactions from similar reactants but with different products. For example, (A–(B–C–D–E–F)–G) would be the product if D and E could form a stable bond. Another possible alternative product would be ((A–(B–C–D))–((E–F)–G)) if a bond was formed between the composite bRBNs of the reactants (and none of the other interactions occurred). This shows that RBN-World is not only complex in terms of which collections of reactants will form a product, but also complex in terms of the structure and composition of the product formed; a feature which is obviously present in



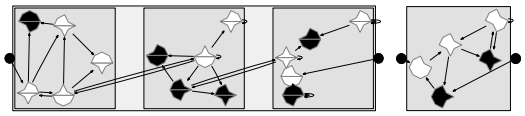
(a) Reactants (A–(B–C–D)) + ((E–F)–G)



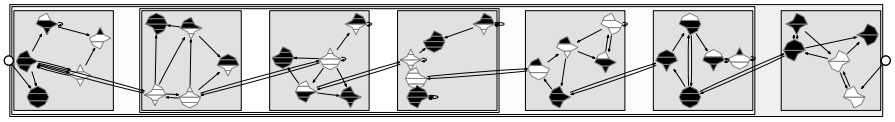
(b) Initial interaction –D + E–



(c) Second interaction –(B–C–D) + E–



(d) Filling of the selected bonding sites in each reactant –(B–C–D)– and –E–



(e) Composition of the product ((A–(B–C–D)–E–F)–G)

Figure 6.2.5: Stages involved in the reaction (A–(B–C–D)) + ((E–F)–G) → ((A–(B–C–D)–E–F)–G).

real-world biomolecules.

### 6.2.6 Example Reaction: $\mathbf{A}^1 + \mathbf{B} \rightarrow \mathbf{A}^2 + \mathbf{B}$

As well as forming composite structures, all bRBNs in RBN-World have a state associated with them. This state may be changed by a reaction as bonding sites are filled and emptied. So far, all of the bRBNs have remained in the same state and therefore this has not been included in the symbolic representation. However, to do so the different states are represented as superscript numerals; bRBNs without such are always seen in the same state and therefore the states do not need to be differentiated. The use of a number is merely for convenience; the state of a bRBN does not have any mathematical properties such as cardinality etc. but is easier to conceptualise and discuss than reusing letters or introducing abstract symbols. By convention, the first observed state of a bRBN is represented by 1 and further different states assigned higher integers as they are observed.

The reactants in this example are $A^1$ and B (see figure 6.2.6a). The symbolic representation $A^1$ is the first observed state of the atom A; B could also be represented as $B^1$, but as there is only one state of B in this example it is omitted for clarity.

The first step is to randomly select a free bonding site on each reactant (with a uniform probability distribution). In this example, the right-hand side of $A^1$ and the left-hand side of B (see figure 6.2.6a) are selected. The bonding property values of $A^1$ and B meet the bonding property criterion and therefore the reaction can proceed to the next stage.

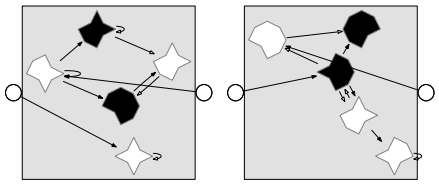
Filling the right-hand bonding node of the $A^1$ bRBN changes the basin of attraction and the cycles within it. This is shown in figure 6.2.7 where the basins of attraction for all possible bonding node states are shown. In figure 6.2.7, each point represents a state of the bRBN and the edges indicate which will be the next state; in each of the four quadrants, the same state is in the same position. As both bonding sites are initially empty, the bottom left quadrant is the basin of attraction at the start of the reaction. When created, the bRBN state is the circled state in that quadrant, but as it is iterated to determine the bonding property, the state transitions until the state indicated by a square is reached when the cycle is determined. This is the state that is represented as $A^1$ and depicted in figure 6.2.6a. When the right-hand bonding site is filled, the basin of attraction for the bRBN becomes the one shown in

the bottom right of figure 6.2.7. The state remains the same initially, and is indicated by the circled state in the bottom right quadrant. However, it is again iterated in order to determine the bonding property until it reaches the square state in the bottom right quadrant.

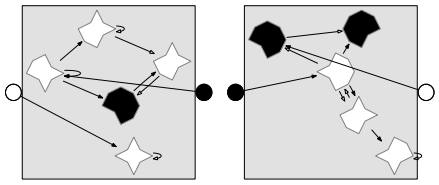
At this point the bonding properties of the reactants are retested to see if the bonding criterion is still met; it is not. As both reactants are single atoms, there is no composing structure to use and therefore the bonding sites are emptied. In the case of the B bRBN, the final state is the same as the initial state. However, for the $A^1$ bRBN, the basin of attraction it is currently in is different which causes the cycle to be different. Therefore the final state of the bRBN is different to the state it was in at the start of the reaction; this state is represented by $A^2$ (as it is the second state of A in that bonding state that has been observed). This state is indicated by a triangle in the bottom right quadrant of figure 6.2.7. The difference of state between $A^1$ and $A^2$ is important for several reasons; not only could it have an effect on the bonding property value of the bRBN for future reactions, but also it could have an effect on how the state changes in future reactions. The overall change of state of A means that the reaction is not a simple elastic $A + B \rightarrow A + B$, but rather it is a richer $A^1 + B \rightarrow A^2 + B$.

This example demonstrates a key feature of RBN-World that is often missing from other Artificial Chemistries, that of multiple states. Unlike real chemistry however, the states in RBN-World cannot be ordered (despite the numeric labels used here). These different states also do not necessarily represent a difference in the bonding property values of the bRBNs, nor of their reactions or other properties. However, without multiple possible states such differences would not be possible because all molecules of a particular structure would be equivalent. Although this example had single-atom bRBNs, multiple states can also apply to both composing and composed bRBNs and this is explored in the next example.

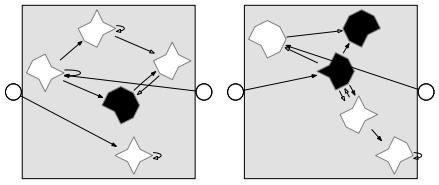
The example presented here also demonstrates a simple form of catalysis; B is unchanged by the reaction yet the reaction overall changes $A^1$ into $A^2$. This type of catalysis was not designed into any part of the system, but rather is an emergent property of the sub-symbolic representations in RBN-World. Reactions that involve a catalyst can be represented in the form $A^1 \xrightarrow{B} A^2$ to indicate that although the catalyst (B in this example) participates in the reaction, it is not changed by the reaction.



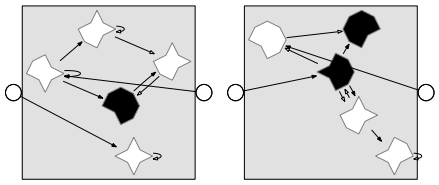
(a) Reactants $A^1$ and B



(b) Filling of the selected bonding sites in each reactant $A^1$– and –B



(c) Emptying of the selected bonding sites in each reactant $A^2$ and B



(d) Products $A^2$ and B

Figure 6.2.6: Stages involved in the reaction $A^1 + B \rightarrow A^2 + B$.

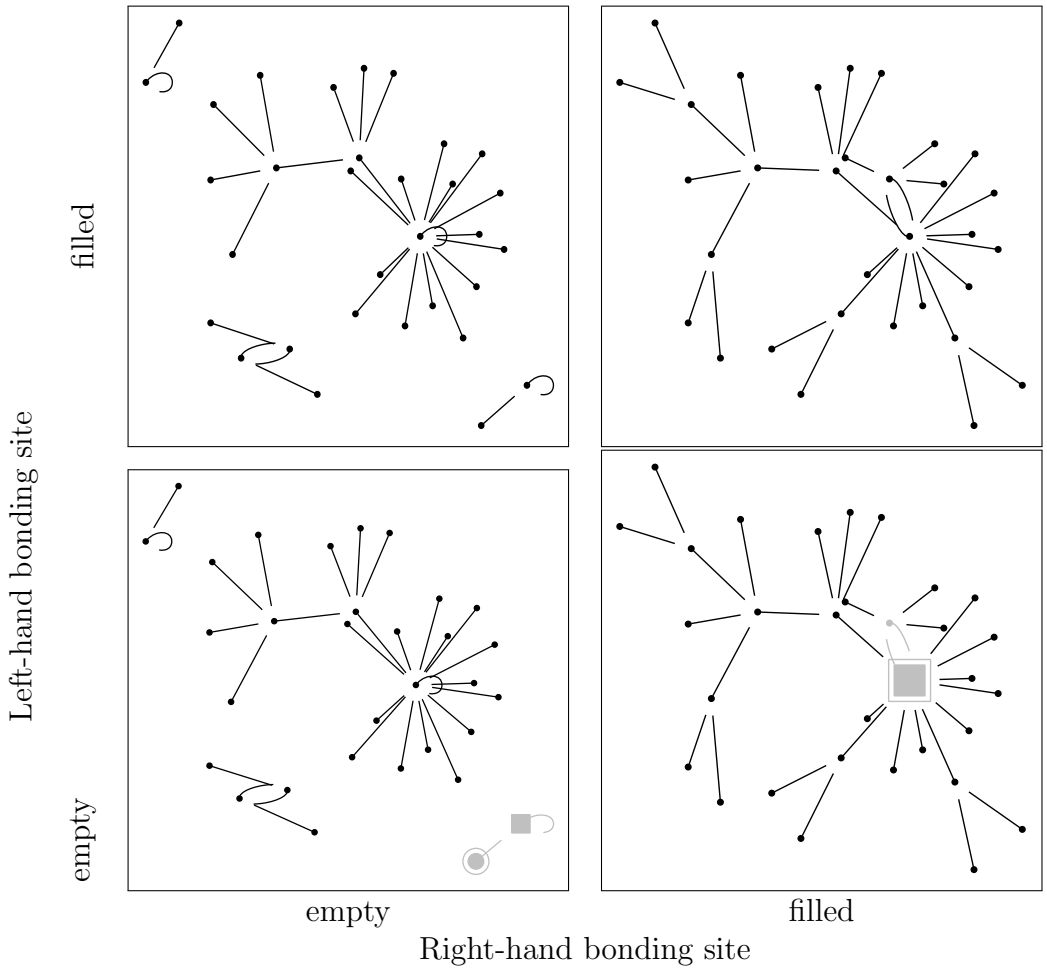



Figure 6.2.7: Basins of attraction of bRBN of atom A when bonding sites are filled and/or empty. Dots represent states and the same state is in the same position in each quartile. Grey parts indicate the state changes on iteration from the initial position (circled) to the point where a cycle is formed (square). Note that the different combinations of filled and empty bonding sites result in different basins of attraction, and that in each basiin there are multiple cycles which each correspond to one or more states.

### 6.2.7 Example Reaction: $(\mathbf{A–(B–C)^1}) + \mathbf{D} \rightarrow (\mathbf{A–(B–C)^2}) + \mathbf{D}$

The previous example demonstrated multiple states of bRBN atoms in RBN-World. However, multiple states can apply to any bRBN not just atomic ones. This example demonstrates multiple states of bRBNs that are composing and/or composed of other bRBNs.

The reactants involved in this reaction are (A–(B–C)$^1$) and D (see figure 6.2.8a). The symbolic representation of (A–(B–C)$^1$) indicates that the B–C bRBN has multiple states in the Artificial Chemistry covered by this example; (B–C)$^1$ and (B–C)$^2$. Note that in the composite bRBN A–(B–C) the state of any composing bRBNs is not relevant to the state of the composed bRBN. For example, the bRBN (A–(B–C)$^1$) is the same as the bRBN (A–(B–C)$^2$) because $^1$ and $^2$ refer to the state of the composing bRBN (B–C) by itself, not as part of the composed bRBN A–B–C. When referring to a single bRBN, only zero or one state indicators are necessary. When referring to a molecule, it is possible for each composite and atomic bRBN to have their own state indicator.

As before, the bonding sites for the reaction are randomly selected and the right-hand bonding site of (A–(B–C)$^1$) and the left-hand bonding site of D are selected. This means that the first interaction is between –C and D (see figure 6.2.8b) whose bonding properties do not meet the bonding criterion. The second interaction is between –(B–C)$^1$ and D (see figure 6.2.8c), which does meet the bonding criterion and therefore the bonding sites become filled (see figure 6.2.8d).

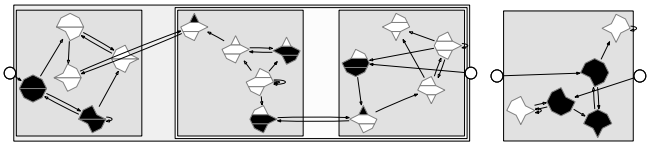
Once the bonding sites are filled, the criterion is retested between –(B–C)– and –D (see figure 6.2.8d)[b]. In this case however, the bonding criterion is no longer met. Therefore the filled bonding sites are emptied. The bRBN D returns to the same state it was at the start of the reaction, however, the –(B–C) bRBN does not.

Although this example here presented two states, in a larger system with multiple reactions more states are possible. Further, bRBNs with multiple states do not have to be collected together within the molecule in any pre-supposed structure as the property of multiple states is an complex emergent property that emerges from the dynamics - not imposed in a top-down *a priori*

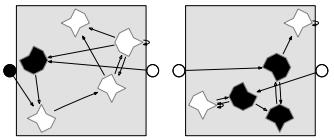
[b]Note that because –(B–C)– is not the same as –(B–C) the state indicators used for –(B–C) are not needed with –(B–C)– because it is only seen in one state in this reaction.

manner. These can be symbolically represented as described and could include (A–(B[1]–C)), (A–(B–C)[1]–D[1]), as well as others.

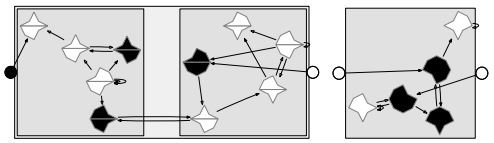
The importance of multiple states in an Artificial Chemistry for Artificial Life is unclear. By analogy with known living creatures it can be hypothesised that different states of proteins are critical, however those 'states' are often changes in molecular structure rather than states as described here. The metal ions that form the catalytic centres of many enzymes are a closer analogy to the states described here, but it is not known if they are a requirement of life or a consequence of it. RBN-World could be used in the future to investigate this issue; one possible alternative Artificial Chemistries would not maintain the state of nodes after determining bonding properties.



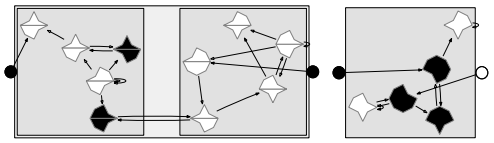
(a) Reactants (A–(B–C)$^1$) and D



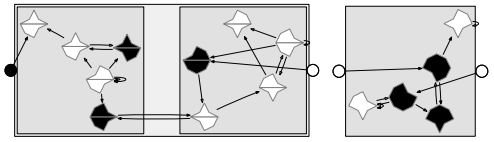
(b) Initial interaction between –C and D



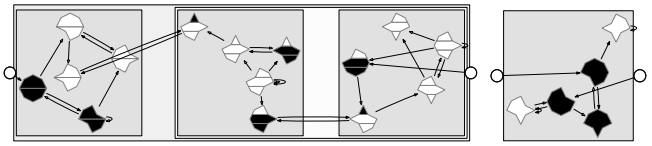
(c) Second interaction between –(B–C)$^1$ and D



(d) Filling of the selected bonding sites –(B–C)– and –D



(e) Emptying of the selected bonding sites in –(B–C)$^2$ and D



(f) Products (A–(B–C)$^2$) and D

Figure 6.2.8: Stages involved in the reaction (A–(B–C)$^1$) + D $\rightarrow$ (A–(B–C)$^2$) + D.

### 6.2.8 Example Reaction: (A–B) + C → A + B + C

In all of the reactions so far, there have been equal or fewer products than reactants. However, some reactions in RBN-World result in the decomposition of one or more reactants which can result in more products than reactants. This example shows one such reaction, where the intermediate molecular structure formed by the reaction is unstable and decays into several separate products.

In this example, the reaction initially proceeds the same as previous examples; reactants (A–B) and C (see figure 6.2.9a) have the right-hand and left-hand bonding sites selected respectively, the bonding properties of –B) and C (see figure 6.2.9b) meet the bonding criterion and the bonding sites are filled resulting in –B–) and –C (see figure 6.2.9c). The bonding properties of –B– and –C still meet the bonding criterion and therefore a bond forms. This forms the intermediate molecular species (A–B–C) shown in figure 6.2.9d.

At this point, the reaction deviates from previous examples. The process of filling the right-hand bonding site of –B) causes the bonding property to change. This means that the A–B bond no longer meets the bonding criterion, and therefore the A–B bond breaks.

Breaking the A–B bond involves three steps. Firstly, the bonding sites used in that bond are emptied in the bRBNs that participated in the broken bond and any smaller scale composing bRBNs. Secondly, any larger bRBNs composed of the bRBNs that participated in the broken bond are split at the point of the bond with the directly reciprocated inputs replaced with new empty bonding sites. The final step in breaking the bond is to remove any composing bRBNs that are now composed of only one bRBN. In this example, the (A–B–) bRBN has been split into one bRBN A and one bRBN (B–) which are both composed of only one bRBN; A and B– respectively. The result after removing the bRBNs composed of only one bRBN is shown in figure 6.2.9e.

At this point, the bRBNs involved in the reaction can be represented as A, B– and –C with the B– and –C bRBNs interacting (see figure 6.2.9e). Because the left-hand side bonding site of B– has been emptied, the bonding property must be recalculated and the bonding criterion re-tested between B– and –C. In this case, the bonding criterion is no longer met between B– and –C, and therefore the B–C bond is broken by following the same procedure used when breaking the A–B bond. This means that the final products of this reaction

are A, B and C as shown in figure 6.2.9g.

This example demonstrates a feature of RBN-World that has not been shown by previous examples, decomposition. Although a bond was not formed by the reaction, the reaction did result in the breakage of a bond which lead to the fragmentation of one of the reactants. This process of breaking apart is crucial for life-like properties; in real-world biology it is present at scales ranging from enzyme-substrate interactions, through processes such as DNA replication, all the way up to cell division. As with the formation of bonds, the breakage of bonds can occur at different levels within the molecular structure.

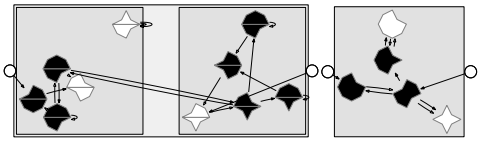
The example shown here also demonstrates another important feature of RBN-World, the removal of bRBNs composed of only one bRBN. Much like decomposition, this feature balances the constructive aspects of the Artificial Chemistry by allowing composed components of composed structures to return to their prior form. Without the removal of bRBNs composed of only one bRBN, over time there would be an ever-increasing number of composing levels. This would in turn lead to an ever increasing combinatoric explosion in the number of molecular species ever observed as well as an inability for the system to return to previous states, preventing cyclic behaviours of any kind.

As with the other examples, it should be noted that there are many alternatives within RBN-World including different processes for decomposition. These different decomposition processess may result in different emergent properties. However, as it is unknown whether those alternatives are better or worse for emergent life-like properties they are not considered in-depth here.

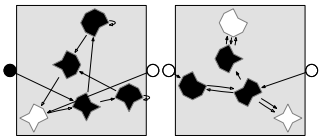
## 6.3 Conclusion

In this chapter, the implicit reaction scheme within RBN-World has been described. The examples shown here give insight into some of the low-level properties of reactions that this core gives rise to, as well as demonstrating how the core of RBN-World functions to exhibit those properties.

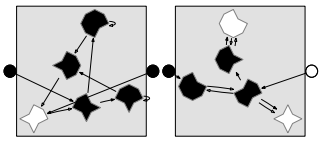
Also, the core demonstrated in this chapter highlights some of the alternative chemistries that exist in RBN-World. However, it is not known if any of these alternatives chemistries would offer a better environment for rich and dynamic emergence than the basic core chemistry. Rather than ignoring those alternatives, the next step is to investigate some of them in order to determine how variable those alternative Artificial Chemistries are and which, if any, are



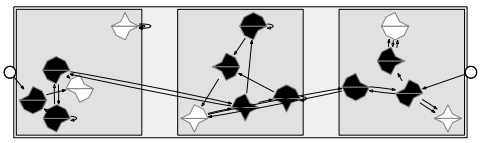
(a) Reactants (A–B) and C



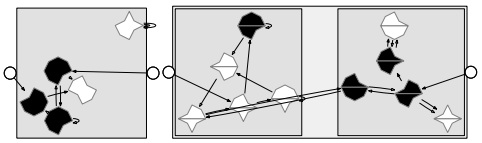
(b) Initial Interaction between –B and C



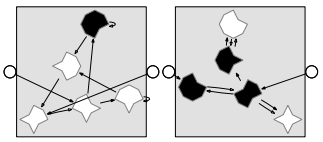
(c) Filling of the selected bonding sites –B– and –C



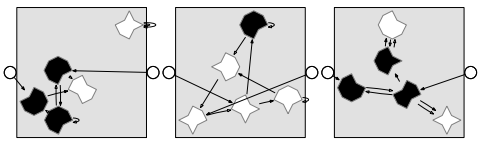
(d) Intermediate molecule (A–B–C)



(e) Breaking A–B bond



(f) Emptying of the selected bonding sites B and C



(g) Products A, B, and C

Figure 6.2.9: Stages involved in the reaction (A–B) + C → A + B + C

suitable for further investigation.

The examples given in this chapter focus on individual reactions and do not address the higher-level properties previously suggested to be important for life-like environments that emerge from multiple reactions, such as hypercycles and autocatalytic sets. These will be addressed at a later stage once the alternative Artificial Chemistries have been investigated to determine if any are more appropriate than the core described here for further study.

# Chapter 7

# RBN-World: The Hunt For A Rich Artificial Chemistry

The examples in chapter 6 are from one Artificial Chemistry in RBN-World that used cycle length as the bonding property and equality as the bonding criterion. However, other variations are possible such as the number of nodes that change state during the cycle. In addition, there are probably further variations that have not been identified. The choice of which variant is used may have impact on the emergent properties of the Artificial Chemistry, in some cases the variant may be better than the core Artificial Chemistry described previously.

To investigate the alternative chemistries, some of the variant choices have been explicitly defined and their effect upon the resulting Artificial Chemistry examined. It is worth noting that the decisions around which alternatives to study have been made based on limited information from preliminary experiments and exploratory ideas and represent neither an exhaustive list of alternatives nor even a list of alternatives that are thought to be potentially useful.

## 7.1 Alternative Chemistries

Four different categories of alternatives have been selected - size of bRBNs, connectivity of bRBNs, bonding property and bonding criterion. In each

| | Min | Max | Description |
|---|---|---|---|
| cycle length | 1 | $2^n$ | Count of steps on cycle |
| Flashing | 0 | $n$ | Count of nodes that change state |
| Flashes | 0 | $n \times c$ | Count of changes of node states over |
| Total | $-n \times c$ | $n \times c$ | Sum of node states over cycle |
| Proportion | 0 | 1 | Proportion node steps on cycle that are True |
| Magnitude | 1 | $n \times c$ | Maximum of node states with False/True on cycle |

Table 7.1.1: Alternative bRBN bonding properties. $n$ is the number of nodes within the bRBN, $c$ is the cycle length of the bRBN.

of these categories there are multiple options possible, and combinations of alternatives from different categories are investigated.

### 7.1.1 Bonding Property

One of the novel aspects of RBN-World compared to other Artificial Chemistries is the use of properties of an underlying sub-symbolic representation to determine bonding. However, it is not clear which property would be most

| | bRBN node | | | |
|---|---|---|---|---|
| | A | B | C | D |
| Cycle steps | F | T | F | F |
| | F | F | F | F |
| | T | T | F | F |
| | T | T | T | F |
| | T | F | T | F |
| | F | T | T | F |

$$
\begin{aligned}
c &= 6 \\
N_{\text{flashing}} &= 1+1+1+0 \\
&= 3 \\
N_{\text{flashes}} &= \tfrac{4+8+4+0}{2} \\
&= 8 \\
N_{\text{tot}} &= 0+2+0+-6 \\
&= -4
\end{aligned}
\qquad
\begin{aligned}
N_{\text{mag}} &= 14 \\
N_{\text{mag}_{\text{T}}} &= 3+4+3+0 \\
&= 10 \\
N_{\text{mag}_{\text{F}}} &= 3+2+3+6 \\
&= 14 \\
N_{\text{prop}} &= \tfrac{10}{4\times 6} \\
&= 0.417
\end{aligned}
$$

Table 7.1.2: Example bonding properties for a $n = 4$ bRBN. Although a specific Artificial Chemistry would only use one they are all displayed here for completeness. The table indicates the states of the bRBN nodes at each sequential step on the cycle.

suitable and what effect different properties might have. Several alternatives are considered here, each with distinct distributions. See tables 7.1.1 and 7.1.2 for summary and example.

#### 7.1.1.1 cycle length

cycle length ($c$) is the number of different states the bRBN passes through between repeats. cycle length has a large but bounded asymmetric discrete distribution of values, with a median of approximately $\sqrt{n}$ for small values of $k$ (Kauffman, 1969).

#### 7.1.1.2 Flashing

Flashing counts how many Boolean nodes change state during the cycle. RBNs typically have a 'frozen core' of static Boolean nodes, and flashing is the inverse of this. This can be expressed as follows; let a state of 'true' have a value of 1 and a state of 'false' have a value of $-1$; $N$ be the set of nodes in the bRBN; $s_{i,j}$ be the state of the $i^{\text{th}}$ node at the $j^{\text{th}}$ state of the repeating cycle. Then:

$$N_{i_{\text{flashing}}} = \begin{cases} 1 & \text{if} \quad \left| \sum_{j=1}^{c} s_{i,j} \right| \neq c \\ 0 & \text{otherwise} \end{cases} \tag{7.1.1}$$

$$N_{\text{flashing}} = \sum_{i \in N} N_{i_{\text{flashing}}} \tag{7.1.2}$$

#### 7.1.1.3 Flashes

Flashes is the total number of Boolean node state changes over the cycle. As at least one node must change state at each step around the cycle, this is related to the cycle length and the flashing property. This can be expressed as:

$$N_{\text{flashes}} = \frac{1}{2} \sum_{i \in N} \sum_{j=1}^{c} \left| s_{i,j} - s_{i,j-1} \right| \tag{7.1.3}$$

#### 7.1.1.4 Total

Total is the sum of all Boolean node values at all time steps over the cycle. This is a property of the states of the bRBN rather than its dynamics and is

related to the cycle length property and the number of Boolean nodes.

$$N_{\text{tot}} = \sum_{i \in N} \sum_{j=1}^{c} s_{i,j} \tag{7.1.4}$$

#### 7.1.1.5 Magnitude

Magnitude is the larger out of the total number of Boolean nodes at all time steps over the cycle that are in the 'true' state compared with the number that are in the 'false' state.

$$N_{\text{mag}_{\text{T}}} = \frac{1}{2} \sum_{i \in N} \sum_{j=1}^{c} (1 + s_{i,j}) \tag{7.1.5}$$

$$N_{\text{mag}_{\text{F}}} = \frac{1}{2} \sum_{i \in N} \sum_{j=1}^{c} (1 - s_{i,j}) \tag{7.1.6}$$

$$N_{\text{mag}} = \max\{N_{\text{mag}_{\text{T}}}, N_{\text{mag}_{\text{F}}}\} \tag{7.1.7}$$

#### 7.1.1.6 Proportion

Proportion is the proportion of nodes in state 'true' averaged over both cycle length and number of Boolean nodes.

$$N_{\text{prop}} = \frac{N_{\text{mag}_{\text{T}}}}{n \times c} \tag{7.1.8}$$

### 7.1.2 Bonding Criterion

In addition to the bonding property, the bonding rule requires a comparison between the properties of two bRBNs for some criterion to be met. There are multiple possibilities to conduct this comparison, and this is another area for exploration.

#### 7.1.2.1 Equal

Equal is the simplest bonding criterion; form a bond where the value of bonding property is *equal* (within 0.1% of the maximum possible range of values to allow

for numerical error). This can be expressed as:

$$\frac{p(N_i) - p_{\min}}{p_{\max} - p_{\min}} - \frac{p(N_j) - p_{\min}}{p_{\max} - p_{\min}} = 0 \pm 0.001 \tag{7.1.9}$$

where $N_i$ and $N_j$ are the bRBNs involved in the bond, $p(x)$ is a function to calculate the bonding property of bRBN $x$, and $p_{\min}$ & $p_{\max}$ are the minimum and maximum possible bonding property values.

#### 7.1.2.2 Similar

Similar is a relaxation of the equal criterion — i.e. within 5% of the maximum possible range of values.

$$\frac{p(N_i) - p_{\min}}{p_{\max} - p_{\min}} - \frac{p(N_j) - p_{\min}}{p_{\max} - p_{\min}} \leq 0.05 \tag{7.1.10}$$

#### 7.1.2.3 Different

Different is the inversion of similar.

$$\frac{p(N_i) - p_{\min}}{p_{\max} - p_{\min}} - \frac{p(N_j) - p_{\min}}{p_{\max} - p_{\min}} \geq 0.05 \tag{7.1.11}$$

#### 7.1.2.4 Sum One

Sum One allows the formation of bonds where the proportion property of the interacting molecules total to one ($\pm 0.001$ allowing for numerical error). This is only applicable to proportion bonding criterion as all the other ones that have been selected are integers, and therefore cannot sum to one.

$$p(N_i) + p(N_j) = 1 \pm 0.001 \tag{7.1.12}$$

#### 7.1.2.5 Sum Zero

Sum Zero requires that the total property of the bRBNs sum to a value of zero ($\pm 0.001$). This is only applicable to total bonding criterion as all the other ones that have been selected are positive numbers and cannot sum to zero.

$$p(N_i) + p(N_j) = 0 \pm 0.001 \tag{7.1.13}$$

Sum One and Sum Zero are applicable only to proportion and total bonding properties respectively as these are the only bonding properties that can meet these bonding criterion.

| n | k | Bonding Property | Bonding Criterion |
|---|---|---|---|
| 5 | 2 | Equal | cycle length |
| 10 | 3 | Similar | Flashing |
| 15 | | Difference | Flashes |
| 20 | | | Total |
| 25 | | | Magnitude |
| | | | Proportion |
| | | Sum One | Proportion |
| | | Sum Zero | Total |

Table 7.1.3: Features of the 200 alternative Artificial Chemistries tested. Every chemistry must have one feature from each column. Horizontal lines cannot be crossed within the table when moving from one column to the next. For example, 5 – 2 – Equal – cycle length is valid, 20 – 2 – Sum One – Proportion is valid, but 5 – 2 – Sum One – Flashes is not valid.

### 7.1.3 Sizes of bRBNs

The number of nodes ($n$) within each bRBN-atom must be chosen. A range of values at intervals was investigated ($n \in \{5, 10, 15, 20, 25\}$ with the potential to expand this range if there appears to be a directional trend).

The size of a bRBN does not have much impact on the chemistry directly. However, it does alter the distribution of the bonding properties, and their responses to bond formation, which in turn affects the propensity for different types of reactions.

### 7.1.4 Connectivity of bRBNs

Previous work on RBNs (Kauffman, 1993) has shown that the number of inputs ($k$) each node has can have an impact on their properties. There is also an interplay with the Boolean function assigned to each node; certain functions can result in one or more inputs having no affect on the state of the node (canalisation) and more different Boolean functions are possible with more inputs.

As an initial assessment, alternatives of two- and three-input bRBNs are considered ($k \in \{2, 3\}$). In theory, any positive integer value equal to or less than the total number of nodes could be used. However, these are values known to be on the 'edge-of-chaos' — higher values are chaotic and lower values are highly ordered.

### 7.1.5 Combinations of Alternatives

The alternatives discussed above each change different, but potentially interlinked, aspects of the Artificial Chemistry. Different combinations of alternatives can be used, though some are mutually exclusive. Table 7.1.3 shows the possible combinations; in total there are 200 different Artificial Chemistries to be considered, each of which may have potentially different and interesting features.

Previous work (Faulconbridge et al., 2011) used $n = 10$ $k = 2$ with 'cycle length' as the bonding property and 'equal' for the bonding criterion as an arbitrary initial choice from the 200 alternatives

## 7.2 Method

As discussed previously, there are a large number of potential alternative chemistries, and each of those has a very large number of potential elemental bRBNs.

Due to the vast number of possible bRBNs, exhaustively testing multiple chemistries is not feasible. Therefore, a random sampling approach is taken. In order for a chemistry to be have the potential for sufficiently rich properties, it is important that at the desired low-level behaviours are seen at least once. However, it is also important that the behaviours are not omnipresent — consider the synthesis test for example (described below); if every interaction resulted in the formation of a stable bond, it would rapidly coalesce into a single molecule and would therefore not exhibit sufficiently rich properties.

The optimal subset of bRBNs in the optimal Artificial Chemistry is not the goal; the goal is to remove those alternative Artificial Chemistries unlikely to exhibit sufficiently rich emergent properties.

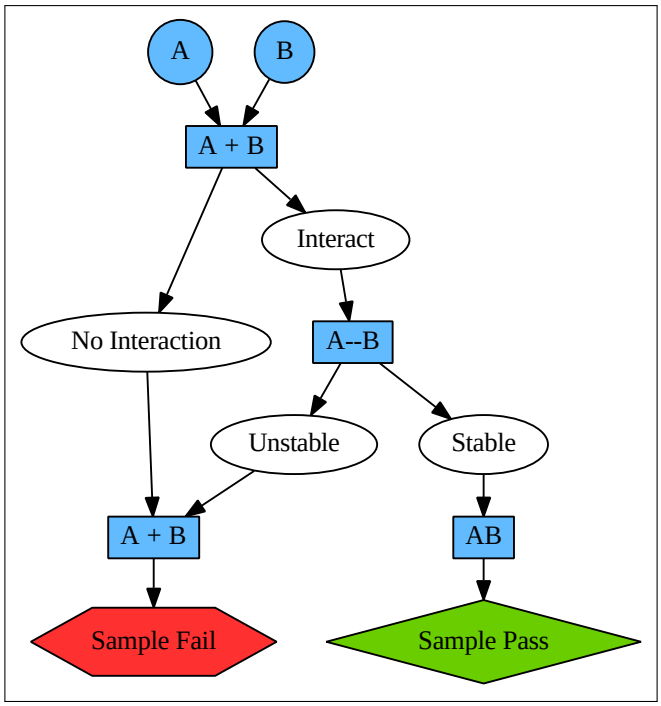



Figure 7.2.1: Schematic depiction of the 'synthesis' test. A and B indicate the sampled atoms.

### 7.2.1 Desired Behaviours

As well as the alternative chemistries, the tests for required low-level behaviours must also be defined. There is a natural structuring of prerequisites within the behaviours – decomposition can only occur if synthesis occurs for example. This can be used to increase the efficiency of the sampling.

#### 7.2.1.1 Synthesis

Synthesis is the lowest-level behaviour possible in an atom-based Artificial Chemistry. A pair of atoms is randomly sampled, the two atoms interact, and the outcome is recorded. RBN-World has a two-stage bonding process, and the bonding criterion must be met both at the start of the interaction and after bonding. If a stable bond can be formed, then the sample passes; if not, the sample fails (figure 7.2.1).

#### 7.2.1.2 Self-synthesis

The self-synthesis test is a synthesis test between two copies of the same element. If a stable bond can be formed, then the sample passes; if not,

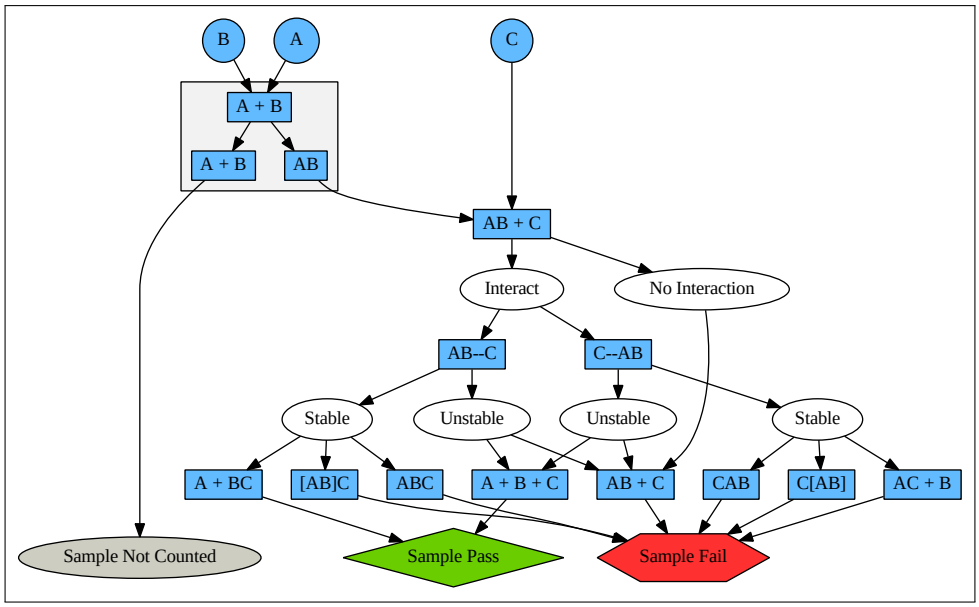



Figure 7.2.2: Schematic depiction of the 'decomposition' test. A and B indicate the sampled atoms.

the sample fails.

#### 7.2.1.3 Decomposition

This is the breaking of bonds, potentially leading to a molecule separating into two (or more) smaller molecules. In RBN-world this is triggered by an interaction between an bRBN molecule and another bRBN. In the decomposition test, samples of three atoms are taken and the first two attempt to form a stable bond. If they cannot form a stable bond, then that sample is ignored for determining pass/fail; this is a test for decomposition, not for synthesis. Once a stable molecule has been formed, it interacts with the third sample. This can have several possible outcomes; no interaction, formation of a larger molecule, or breakdown into two or three separate molecules. If it results in the bond between the first two sampled bRBNs breaking, then it is recorded as a pass; other outcomes are classed as fail (figure 7.2.2).

#### 7.2.1.4 Substitution

Similar to decomposition, the substitution test involves an interaction with a molecule that leads to replacement of part of the molecule with the reacting bRBN. The process is the same as the decomposition test, but the only valid

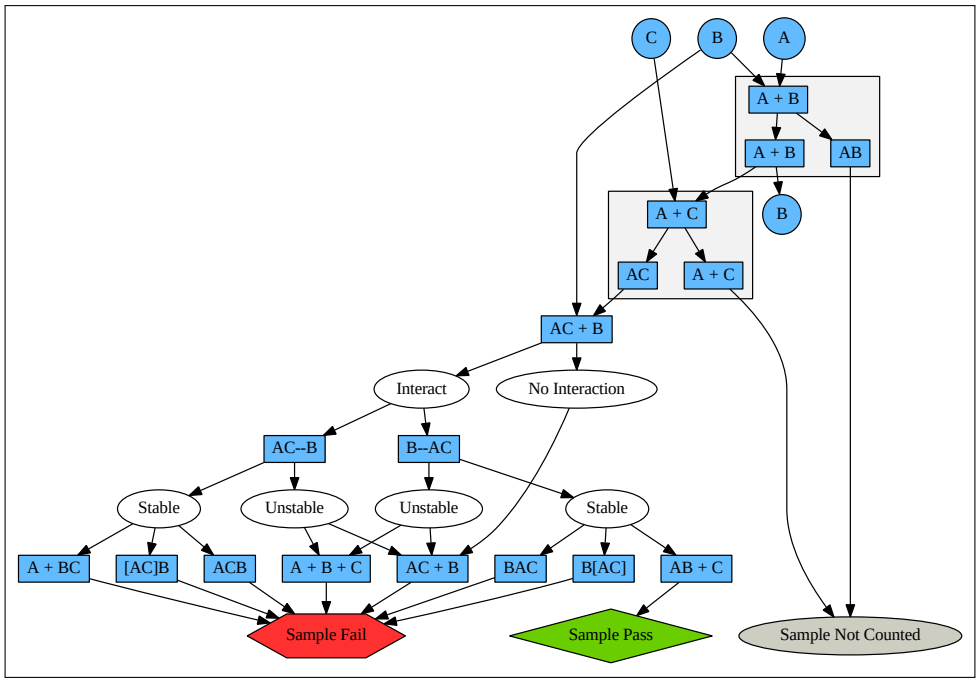



Figure 7.2.3: Schematic depiction of the 'catalysis' test. Requirements of the catalysis test are that A and B must *not* synthesise, and that C and A must synthesise; these are indicated by the highlighed subgraphs (details removed for brevity).

outcome is a direct replacement of the second sampled bRBN with the third sampled bRBN, i.e. AC+B in figure 7.2.2.

#### 7.2.1.5 Catalysis

This is the highest-level property investigated here. Unlike the other desired properties, catalysis can take many forms. Any of the other tests could be repeated requiring the presence of a catalyst. For simplicity, catalysis of synthesis reactions is focused on.

The test proceeds as follows: as before, a sample of three bRBNs is taken and the first two attempt to form a stable bond. However, unlike decomposition or substitution tests, this time it is important that a stable bond does not form. If a bond does form, then the sample is not counted for pass/fail.

After that initial bond formation stage, the third bRBN in the sample attempts to form a bond with the first; this is analogues to interacting with a catalyst to form a temporary intermediate. If this does not form a stable bond, then again the sample is not counted for pass/fail.

| Test | # of Artificial Chemistries tested | # of Artificial Chemistries where all samples Passed or Not Counted | # of Artificial Chemistries where all samples Failed or Not Counted | # of Artificial Chemistries remaining |
|---|---|---|---|---|
| Synthesis | 200 | 10 | 7 | 183 |
| Self-Synthesis | 183 | 110 | 53 | 20 |
| Decomposition | 183 | 0 | 6 | 177 |
| Substitution | 177 | 0 | 18 | 159 |
| Catalysis | 177 | 0 | 39 | 138 |

Table 7.3.1: Results from testing 10,000 samples from each of 200 alternative chemistries for low-level emergent behaviours. The prerequisite for decomposition and self-synthesis tests is synthesis. The prerequisite for substitution and catalysis tests is decomposition. See text for details.

The final step is to test that the second bRBN from the sample can substitute for the third bRBN. If this is the case, then the third bRBN has acted as a catalyst for the formation of the bond between the first and the second bRBN that would not occur directly (figure 7.2.3).

## 7.3 Results

The outcomes of testing the described alternative chemistries with 10,000 randomly generated samples of bRBNs is summarized in table 7.3.1 (testing took approx. 2 days on a 24 quad-core CPU cluster). With each test a number of alternative Artificial Chemistries are ruled out; the chemistries that pass all tests are listed table 7.3.2.

Less than 5% of alternative chemistries pass all the tests. The $n$ & $k$ categories of alternatives have little or no influence on the low-level properties of the chemistry. The anomaly is $n = 25$, $k = 3$ with bonding property 'total' and a comparison of 'sum zero'; however, this may be due to sample

| n | k | Measurement | Comparison |
|---|---|---|---|
| 5 | 2 | Proportion | Sum One |
| 10 | 2 | Proportion | Sum One |
| 15 | 2 | Proportion | Sum One |
| 20 | 2 | Proportion | Sum One |
| 25 | 2 | Proportion | Sum One |
| 5 | 3 | Proportion | Sum One |
| 10 | 3 | Proportion | Sum One |
| 15 | 3 | Proportion | Sum One |
| 20 | 3 | Proportion | Sum One |
| 25 | 3 | Proportion | Sum One |
| 5 | 2 | Total | Sum Zero |
| 10 | 2 | Total | Sum Zero |
| 15 | 2 | Total | Sum Zero |
| 20 | 2 | Total | Sum Zero |
| 25 | 2 | Total | Sum Zero |
| 5 | 3 | Total | Sum Zero |
| 10 | 3 | Total | Sum Zero |
| 15 | 3 | Total | Sum Zero |
| 20 | 3 | Total | Sum Zero |

Table 7.3.2: The 19 alternative Artificial Chemistries that exhibit variation across all 5 low-level emergent behaviours tested.

size. Closer examination of this case shows that of $10,000$ samples in the decomposition test, $9,677$ were not counted (as they did not form a molecule that could break down) and none of the remaining 323 samples passed. In comparison, the $n = 20$ equivalent Artificial Chemistry where $9,382$ were not counted and 43 of the remaining 618 samples passed.

For the property and comparison alternatives, only those using 'proportion' as property and 'sum one' as the criterion or those using 'total' as the property and 'sum zero' as the criterion pass all tests. Whilst alternatives should be kept in mind, there is now evidence that these options are more likely to be capable of rich emergent properties. As various different values of $n$ and $k$ were tested and did not affect which chemistries passed the tests, these values can be chosen based on other concerns, such as computational tractability. One potential issue is that this work has only sampled from atomic constituents; it is not guaranteed that molecular structures will also exhibit these behaviours. While various values of $n$ were tested, molecular bRBNs of many atoms may not behave as an equivalent large bRBN atom due to the constrictions from

reciprocal bonding sites between atoms.

## 7.4 Conclusions

Described here are a number of simple tests of an Artificial Chemistry that can be used to restrict the design space to non-trivial chemistries. This is important as for many Artificial Chemistries there are a large number of alternatives that should be considered – for RBN-World only a small fraction of possible alternatives have been examined. It has also been shown that the initial choice of parameters did not pass these tests (Faulconbridge et al., 2011, 2010).

A filtering metric provides a useful testing approach that does not require computationally expensive and/or exhaustive testing of molecules and/or reactions. It is also interesting to see that some Artificial Chemistries fail because all tested sample interactions failed, but some chemistries fail because all tested sample interactions passed; the presence of variation is a requirement for emergent properties.

The work described here could be extended in several directions. Firstly, the conclusions drawn from the tests could be made more robust by looking for the properties in a more comprehensive simulation, rather than the specific test set up described here. However, this has the drawback that it does not have the simplicity and transparency of the atom-based testing.

Secondly, more alternatives could be considered – in particular alternatives relating to reaction algorithm. In contrast to the alternatives investigated here, these are harder to implement and execute but may show more dramatic changes than the alternatives investigated here.

Finally, more properties could be investigated. The properties examined are all low-level and easily expressed as specific tests. However, it is the higher-level properties such as autocatalytic sets and hypercycles that are of greatest interest in the context of this work and therefore will be investigated in the next chapter. If it is found that the alternatives investigated so far do not demonstrate such higher-level properties, further alternatives can be considered.

# Chapter 8

# RBN-World: Searching Elemental subsets

Following on from the previous chapters, two out of many possible alternative Artificial Chemistries within RBN-World have been identified to have interesting properties — including synthesis, decomposition and catalysis. This was done by looking at large numbers of possible elements, which lacks the elegance found in natural biochemistry, in which almost all important molecules are mostly composed of a small number of elements; carbon, hydrogen, oxygen, nitrogen, phosphorus, etc. A possible consequence of this is that any hypothetical metabolic pathways involving large numbers of elements are more complicated; large molecules are expected be more specific in their metabolic requirements because elements cannot be changed.

To address this, it is desirable to discover a small subset of elements (approximately 5) that could form a life-like system. It is expected that these elements could be identified by interesting emergent properties in their molecular interactions. Due to the very large number of possible elemental bRBNs, it is impractical to do a systematic investigation even for small bRBNs, and therefore a search-based approach is more appropriate. This chapter describes the method used and the result of this investigation.

## 8.1 Desirable properties

Although the properties examined in chapter 7 are interesting, they are not suitable for use in a search because they are all-or-nothing properties - either

there is variation in synthesis in the Artificial Chemistry, or there is not. Though they could be converted to a numeric value, for example the proportion of reactions that lead to synthesis, it is unclear what values would be more desirable; values of 0% or 100% are clearly not desirable but, for example, how should elements with 34% and 35% synthesis reactions be ordered? Because of this other properties need to be selected for a productive search.

Hypercycles and autocatalytic sets are some of the properties thought to be useful for Artificial Life, as discussed in §4.3.9 and §4.3.8. Unlike the properties used in chapter 7, autocatalytic sets can be ordered by increasing size.

There is a significant problem with using the size of hypercycles and autocatalytic sets as search criterion however; the size is an integer value and most elemental subsets would be expected to have no hypercycles or autocatalytic sets. This means that much of the fitness landscape would be a large plateau at size zero with further plateaus at integer values. This gives no information for the search to across each plateau efficiently. Due to these issues, the size of autocatalytic sets would not be useful for a computationally tractable search.

Hypercycles were inspired by catalysed metabolic cycles, such as the Krebs (a.k.a TCA) cycle. In natural chemistry, every catalysed reaction can occur without the presence of the catalyst, though it may be several orders of magnitude slower. Furthermore, it is reasonable that there were fewer catalysed biological reactions in early forms of life on Earth than in modern cells. Therefore, rather than looking for catalysed cycles (or hypercycles thereof), metabolic loops without catalysis of every step can be searched for. It is expected that metabolic loops are be more common and therefore the initial search plateau would be smaller. While this is an improvement, it is still not ideal for use in a search.

Searching for metabolic loops would be expected to have significant plateau effects, not only for discovering an initial metabolic loop, but also for discovering larger metabolic loops afterwards. This problem could be reduced by using additional properties as tie-breakers when deciding between sets of elements that have the same size metabolic loop. The size of metabolic pathways are a property that could be used in this manner. In this case, metabolic pathways are similar to metabolic loops but not closed. While this is also an integer value and therefore has the same plateau effects, it is expected that the combination of loops and pathways will minimize these issues.

## 8.2 Generating Reaction Networks

Searching for interesting elemental subsets within RBN-World requires the generation of a reaction network from an initial set of molecular species. The reaction network can then be analysed for the properties of interest. There are several ways a reaction network could be generated and these are addressed below.

### 8.2.1 Systematically

One method for generating a reaction network from a set of initial molecular species is to exhaustively enumerate all possible reactions from all possible collections of reactants. When previously unseen molecular species are generated, new reactions become possible which may lead to more previously unseen molecular species.

For example, consider a set of molecules $A^1$, $B^1$. Initially, there are 4 possible collections of reactants — $A^1 + A^1$, $A^1 + B^1$, $B^1 + A^1$ and $B^1 + B^1$. For this example assume that the following reactions occur:

$$A^1 + A^1 \rightarrow (A^1\text{–}A^1)^1$$
$$A^1 + B^1 \rightarrow A^1 + B^1$$
$$B^1 + A^1 \rightarrow B^1 + A^1$$
$$B^1 + B^1 \rightarrow B^2 + B^2$$

Although two of the collections of reactants did not interact, and can therefore be ignored, the other two collections of reactants did interact. These reactions produce two previously unseen molecular species $(A^1\text{–}A^1)^1$ and $B^2$. From these new molecular species, 12 new collections of reactants are possible, which are likely to produces at least one novel molecular species. This process leads to a combinatorial explosion of molecular species and possible interactions between them.

Although the number of molecular species increases, these may not reflect the species present in dynamic behaviour of the system over time. This is because there are no reaction rates, overall conservation of mass, divergent reactions, convergent pathways or other effects. In particular, a significant amount of computational time may be spent on the consequences of interactions between molecules that rarely coexist. This effect is exaggerated if there are divergent reactions as all of the outcomes must be calculated for a specific collection of reactants.

To address this, a reaction network can be generated from a simulation of the Artificial Chemistry over time. By recording each reaction that occurs rather than each reaction that is possible, more effort is concentrated on areas of the Artificial Chemistry that are dynamically reachable rather than theoretical. The drawback of such an approach for RBN-World is that a mixing algorithm has to be selected, and this could have a significant impact on the system as well as the computation requirements.

### 8.2.2 From Simulation

In order to generate a reaction network from a simulation of RBN-World, a mixing algorithm must be selected in order to complete the Artificial Chemistry. There are several approaches that have been used in other Artificial Chemistries and have been described previously (§4.1.3).

An explicit 2D or 3D mixing algorithm could be used; e.g. Hutton (2004b, 2007). However, this would be very computationally intensive and therefore not appropriate for searching many reaction networks.

An aspatial mixing algorithm is less computationally intensive. As discussed previously, many desirable properties of Artificial Chemistry for Artificial Life are not possible with an aspatial mixing algorithm (cells, membranes, *etc*). However, as the goal here is to find an interesting subset of atoms rather than a complete Artificial Chemistry, this is not a problem. Once an interesting elemental subset has been discovered, properties of a spatial mixing algorithm can be investigation.

#### 8.2.2.1 Iterative

A simple mixing algorithm is as shown as pseudocode in figure 8.2.1 (Banzhaf, 1993a,b,c).

The advantage of this mixing method is that it is simple to understand and implement, and that it is computationally fast. The disadvantage of this method is that it is strictly linear algorithm and therefore does not scale with the number of molecules. The running time is proportional to both the number of molecules and the number of reactions evaluated.

```
Require: b := a bucket of N molecules
 1: repeat
 2:    mola, molb := randomly pick and remove reactants (2 molecules)
       from b
 3:    products := products of a reaction between mola and molb
 4:    add products to b
 5: until sufficient reactions have been attempted
```

Figure 8.2.1: Pseudocode of an iterative mixing algorithm.

```
Require: b := a bucket of N molecules
 1: repeat
 2:    bnew := an empty bucket
 3:    while b contains at least 2 molecules do
 4:       mola, molb := randomly pick and remove reactants (2
          molecules) from b
 5:       products := products of a reaction between mola and molb
 6:       add products to bnew
 7:    end while
 8:    add remaining molecules in b to bnew
 9:    b := bnew
10: until sufficient reactions have been attempted
```

Figure 8.2.2: Pseudocode for a stepped mixing algorithm.

#### 8.2.2.2 Stepped

A possible alternative to the iterative mixing algorithm is to use a stepped mixing algorithm. This is shown in figure 8.2.2.

Unlike the iterative mixing algorithm, the stepped mixing algorithm directly incorporates simulated time that takes the number of molecules into account. The running time of this algorithm is also $p(A^1–A^1)^1$, but for RBN-World the products of each reaction can be determined independent of any other reaction. This means that the stepped algorithm is embarrassingly parallel with respect to each collection of reactants, and therefore can harness more computational resources simultaneously.

#### 8.2.2.3 Gillespie-like

The final aspatial mixing algorithm considered here is based on the Gillespie algorithm (Gillespie, 1977). A pseudocode representation of this is shown in figure 8.2.3. Similar to the iterative algorithm, it determines what the next

**Require:** $b$ := a bucket of $N$ molecules
1: $t := 0$
2: **repeat**
3: *mola*, *molb* := randomly pick and remove reactants (2 molecules) from $b$
4: $t$ += randomly determined interval until reaction occurs.
5: *products* := products of a reaction between *mola* and *molb*
6: add *products* to $b$
7: **until** $t >$ desired value

Figure 8.2.3: Pseudocode for a Gillespie-like mixing algorithm

reaction will be and then uses that to generate the interval until that reaction occurs. Unlike the iterative algorithm, in the Gillespie algorithm these are both stochastic processes that approximate a well-mixed three-dimensional space. It is commonly used for simulating known reaction networks, but can easily be adapted for Artificial Chemistries where the reaction network is not known in advance.

A Gillespie-like mixing algorithm is based on the collision of particles in a continuous three-dimensional space, and therefore is a better approximation of natural chemistry than either the iterative or stepwise approaches. The incorporation of continuous time directly into the algorithm also allows for more appropriate comparisons between simulations. It does share a drawback with the iterative mixing algorithm that it cannot be parallelized to take advantage of additional computational resources; however, in a search-based approach this may be mitigated by parallelizing the search algorithm.

## 8.3 Search

The goal of the search is to identify an elemental subset that has interesting emergent behaviours. The simplest approach would be to examine all possible elemental subsets, but there are so many possible elemental subsets that this is not feasible. An iterative search, such as hill-climbing or simulated annealing, could be used but this has several potential problems.

The process of simulating an Artificial Chemistry to produce a reaction network is stochastic. This means that the fitness of an elemental subset can vary between repeated simulations. The standard approach to account for this is to estimate an average fitness based on several repeats, but this can

drastically increase the computational requirements.

Another problem with iterative search approaches for identifying elemental subsets in an Artificial Chemistry is that of local maxima[a]. These are points in the fitness landscape that are the best value possible within their local neighbourhood, but are significantly less than the overall best-possible global maximum fitness. Although it is not know whether or not the fitness landscape contains these, local maxima are often found in rich complex systems such as Artificial Chemistries. This problem can be solved by restarting the search when a local maximum is found, or by adjusting the cooling of the simulated annealing. However, both of these significantly increase the computational demand of the search process.

### 8.3.1 Genetic Algorithm

An alternative form of search to iterative approaches is that of a genetic algorithm (Barricelli, 1957). Inspired from evolution by natural selection a genetic algorithm looks at many individual points on the fitness landscape, discards the least fit ones, and then looks at multiple new points similar to the individuals that remain. This process repeats for multiple generations.

Compared to iterative approaches, a genetic algorithm is more computationally intensive for simple fitness landscapes. Calculating the fitness of many individuals each generation requires more effort than looking at a few neighbouring points. However, for rich, complex and multi-dimensional fitness landscapes, such as those to be expected in searching for an interesting subset of elements in an Artificial Chemistry, a genetic algorithm is a more appropriate solution. For example, because it uses many individuals at once, if some of them become stuck on local optima they will be overcome once more globally optimum locations are found. The use of many individuals also allow for diversification over plateaus in the fitness landscape until higher fitness regions are located. The computational cost of a genetic algorithms can be mitigated by processing each individual in parallel. This enables a genetic algorithm to take advantage of more hardware resources.

A drawback of genetic algorithms is the amount of complication and complexity. There are several different parameters and options that can be changed, and often require fine-tuning to get optimum performance.

[a] also known as a rugged fitness landscape

## 8.4 Searching RBN-World

In light of the points discussed above, the search for interesting subsets of elements within RBN-World uses a genetic algorithm. A Gillespie-like simulation method is used in the fitness calculation in order to better approximate a well-mixed three-dimensional space and enable comparison between simulations. Because the genetic algorithm can be parallelized across multiple computing resources, the linear time complexity of a Gillespie-like algorithm is mitigated.

In order to implement this, several components must be described. Firstly, the genome representation used in the genetic algorithm must be chosen, as well as how it is interpreted to generate the initial bRBNs of the simulation. Secondly, the mutation processes that can apply to the genome representation must be specified. The third component is the fitness function itself; given a reaction network produced from a simulation, it must be scored in some way. Finally, the various parameters of the genetic algorithm must be defined; population size, number of generations, simulation duration, number of elements, number of bRBN nodes, bonding score and criterion, etc.

### 8.4.1 Genome representation

The basic unit of an elemental subset is a bRBN (§5.2). There are multiple bRBNs present in an elemental subset, but each one is a distinct unit. Therefore, the most obvious representation is that each element is a separate chromosome and there are multiple chromosomes in each individual. It is assumed that each individual is haploid with only one copy of each chromosome, rather than more complicated genomes inspired by sexual diploid organisms.

Within each chromosome, several features must be defined. As described previously (§5.2) a bRBN is composed of $n$ nodes, each of which has $k$ inputs, a Boolean function, and an initial state. By fixing the values of $n$ and $k$ for the genetic algorithm overall, the structure of each chromosome is simplified. On each chromosome are $n$ repeated structures, each composed of $k$ Boolean function, $k$ inputs, and an initial state. The Boolean function consists of $2^k$ genes that can have binary values, corresponding to the outputs from the Boolean function for all possible inputs. The $k$ inputs are represented as $k$ integers where each integer has a value between 1 and $n$ and represents the position of the input node on the genome i.e. the first node is represented by

| | | | |
|---|---|---|---|
| node 1 | inputs | 1 | 1 |
| | | 2 | 3 |
| | function | false, false | true |
| | | false, true | false |
| | | true, false | false |
| | | true, true | true |
| | initial state | | true |
| node 2 | inputs | 1 | 2 |
| | | 2 | 2 |
| | function | false, false | false |
| | | false, true | false |
| | | true, false | true |
| | | true, true | false |
| | initial state | | false |
| node 3 | inputs | 1 | 1 |
| | | 2 | 2 |
| | function | false, false | true |
| | | false, true | true |
| | | true, false | false |
| | | true, true | true |
| | initial state | | false |

Figure 8.4.1: Genome representation of a $n = 3, k = 2$ bRBN (left) and an example chromosome (right).

1 and the last by $n$. The final component of each repeated structure is the initial Boolean state, represented as another binary value.

The structure of an $n = 3, k = 2$ genome representation is shown in figure 8.4.1. This includes an example chromosome, the bRBN of which is shown in figure 8.4.2.

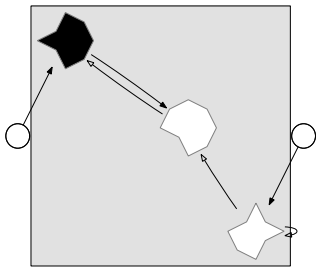


Figure 8.4.2: The bRBN of the example genome shown in figure 8.4.1

### 8.4.2 Mutation

In addition to the genome representation, the mutation operation is an important component of a genetic algorithm. Point mutations, where only a single value is changed in each mutation event, are the simplest mutation system to implement. It was selected that for each gene on each chromosome, there it would change to a different value within the possible values of that gene with probability $p = (k^n + k \times n + n)^{-1}$. This is an average of 1 point mutation per chromosome per generation. However, these point mutations may be silent or otherwise not affect the fitness of the individual e.g. rewiring an ignored input, mutations in an element that does not form any metabolic loops.

For individuals where there were issues with the fitness evaluation — for example, where no reactions occurred or the evaluation consumed excessive computational capacity — these were replaced with new randomly generated individuals rather than offspring. These newly generated individuals are replaced by offspring in the next generation as normal. This was done in order to initially explore the fitness landscape and escape regions of the fitness landscape that are computationally hard to evaluate

### 8.4.3 Fitness function and selection

The fitness function is the most complicated aspect of a genetic algorithm, and has the most impact on the outcomes. It is also the aspect that is most specific to Artificial Chemistries.

There are two components to the fitness function used for searching RBN-World. The first component generates a reaction network from the elements encoded in the genome representation. The second component calculates a numerical fitness from that reaction network that can be used for comparisons between the reaction networks generated from the elemental subsets encoded on the genome.

The fitness function used here to search RBN-World for interesting elemental subsets is shown in figure 8.4.3.

#### 8.4.3.1 Generating a Reaction Network

To generate a reaction network for evaluation, the Gillespie-like simulation approach outlined above (§8.2.2) is used. This was selected as its inherent

**Require:** $e$ := elemental bRBNs
1: *initial* := make many copies of each bRBN in $e$
2: *net* := generate reaction network from *initial* for $t$ seconds
3: replace each reaction in *net* with direct edge from each reactant to each product
4: remove edges in *net* where ends are not sufficiently similar
5: *loops* := loops in *net*
6: remove loops in *loops* where all members are not sufficiently similar
7: **return** length of longest loop in *loops*

Figure 8.4.3: Pseudocode of fitness function used to identify interesting subsets of elements in RBN-World.

stochasticity enables the genetic algorithm to progress by chance, in contrast to deterministic methods that may be exploitable by the fitness function, leading to genomes that score high on fitness yet are not as intended. The use of a stochastic evaluation process reduces this because although it is possible for a genome to score highly by chance, it is unlikely that such artificially high fitnesses is reproducible. Therefore, the danger of premature convergence on points of artificially high fitness is reduced, but not eliminated. Another solution to this problem would be to determine the fitness based on an average of several replicates, however this would require increased computational effort to accurately determine the fitness of points that are not of interest due to low fitness.

#### 8.4.3.2 Evaluating a Reaction Network

Once a reaction network has been generated, it must be evaluated. Unlike the properties used in chapter 7, the fitness score of the reaction networks must be sortable in order to be selected from.

One of the most interesting properties of reaction networks is hypercycles. However, hypercycles are rare and therefore it is likely that fitness landscape based on hypercycles would have a very large plateau where no hypercycles are present. This would 'trap' a search as there is no information available to progress and therefore it is reduced to a random walk. Furthermore, comparing hypercycles is not a straightforward process; should it be by number of members, by number of layers, or some other criterion? Because of these reasons, hypercycles are not suitable for evaluating reaction networks in a

genetic algorithm.

Autocatalytic sets — and therefore catalytic cycles — are another desired property of reaction networks. As with hypercycles, however, autocatalytic sets are too rare to be useful in a genetic algorithm.

The non-catalysed 'loops' discussed earlier (§4.3.9.4) are expected to be more widely present. Unlike hypercycles, an obvious initial comparison is to order by maximum length. Therefore, these may form a good basis for fitness evaluation of reaction networks in a genetic algorithm.

#### 8.4.3.3 Identifying loops

There are several well-established algorithms to identify the shortest route(s) between two nodes on a graph. These can be used to identify cyclic structures by considering each edge and finding the shortest route(s) to go from the end node of that edge to the start node.

However, applying these directly to a reaction network does not produce very satisfactory results. For example, consider the reaction network shown in figure 8.4.4. At first glance, this contains a loop: A $\rightarrow$ ABCD $\rightarrow$ BDC $\rightarrow$ ADB $\rightarrow$ A. However, the molecules that compose this loop do not have much in common — the molecule A is a key link in the loop, yet the element A is not present in one out of the four molecules.

To address this problem, a four-stage solution is used. The first step is to remove the reaction nodes and associated edges, being replaced by direct edges between each reactant and product of each reaction (see figure 8.4.5). The second step is to remove those edges where the joined molecular species are not sufficiently similar. At this point, the loops are generated using shortest route algorithms. The third step is to discard any loops where the molecular species that compose the loop are not sufficiently similar (see below). The final stage is to remove those loops which are too similar; without this stage the genetic algorithm quickly converges on elemental bRBN that exhibit a high number of alternative states, but few molecular structures. In this step, a loop which is "too similar" is a loop where the proportion unique molecular structures out of all molecular species composing the loop is below a tunable threshold — typically 50%.

The use of the term "sufficiently similar" in the paragraph above is an important consideration. If there is no limitation on similarity of molecular

species when detecting loops then the problem described above arises — small molecules and many uninteresting loops. If the requirement for similarity is too strict, then the opposite issue will arise; too few loops detected, resulting in unsatisfactory genetic algorithm progression. A simple similarity measure between molecular species for RBN-World is the size of the largest common subgraph between the molecules, where the molecules are represented as undirected graphs of their elements.

This can be demonstrated by examples. In the simplest case, given the molecular species ABCD and BC, the molecular species BC is entirely contained within ABCD and therefore the shared subgraph is BC. For larger molecules, such as ABCD and BCDE, the shared subgraph will not be an entire molecular species — in this case it is the subgraph BCD. As molecular species are reduced to the undirected graphs of their elements for this comparison, state and functional groups are ignored; e.g. A-B-(C-D) and (B-C)D-E share the subgraph B-C-D; A-$B^1$ and A-$B^2$ share the subgraph A-B; ((A-$B^1$)-C)$^2$D and (($B^2$-C)$^1$-(D-E))$^3$ share the subgraph B-C. Although this is a severe approximation of the richness of RBN-World, at this early stage of the investigation it is useful to make this assumption in order to progress.

Using this similarity measure, a cut-off for "sufficiently similar" must be defined. The most naïve approach is to use a fixed value; however, this would have different effects for different sizes of molecules. To take this into account, a proportion of the largest molecule is used instead. For example, for A-B-C-D and B-C-D-E the shared subgraph is B-C-D which is 75% of the largest molecule.

This process can be seen completely in figures 8.4.4, 8.4.5 and 8.4.6. The first (figure 8.4.4) shows the reaction network. In figure 8.4.5, the reactions of figure 8.4.4 are removed and replaced with pairwise edges between all the reactants and all the products of each reaction. The third figure 8.4.6 only shows those edges that are between "sufficiently similar" molecular species —- those that share 50% or more of their elemental graphs.

### 8.4.4 Genetic algorithm parameters

Several parameters need to be specified before the search can be executed. Some of these apply to the RBN-World Artificial Chemistry, and some of these apply to the genetic search algorithm.

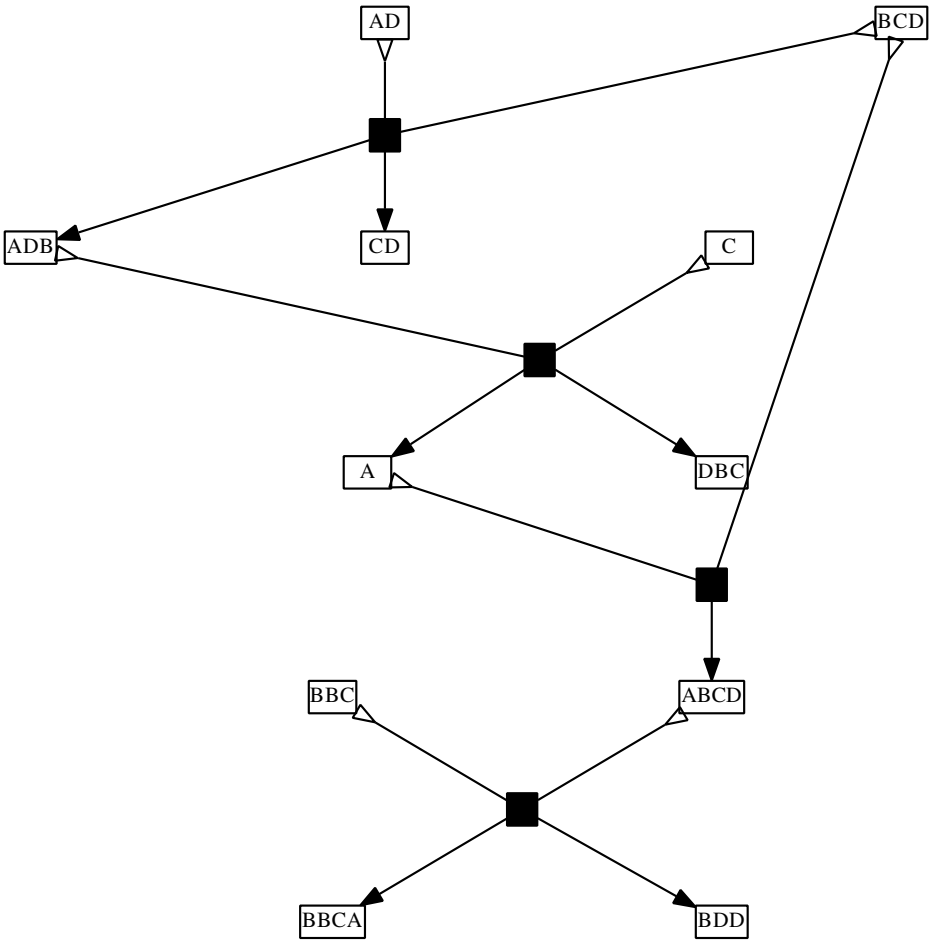



Figure 8.4.4: An example reaction network, manually created. This should be compared with figure 8.4.5 and figure 8.4.6 for an example of how loops are identified.

#### 8.4.4.1 Artificial Chemistry parameters

For RBN-World, the most important properties to specify are the bonding score and bonding criterion. Based on the work in chapter 7, there were two promising options; 'proportion sum one' & 'total sum zero'. There is little evidence to inform a choice between these alternatives, therefore it was arbitrarily chosen to use the 'proportion' as the bonding score and 'sum one' as the bonding criterion. The alternative of 'total sum zero' could be used instead for future work. Other RBN-World properties that need to be specified are $n$ and $k$. Chapter 7 indicated that these parameters do not significantly alter the behaviour of the system (within the limits tested) and therefore values of $n = 10$ and $k = 2$ were selected for computational tractability.

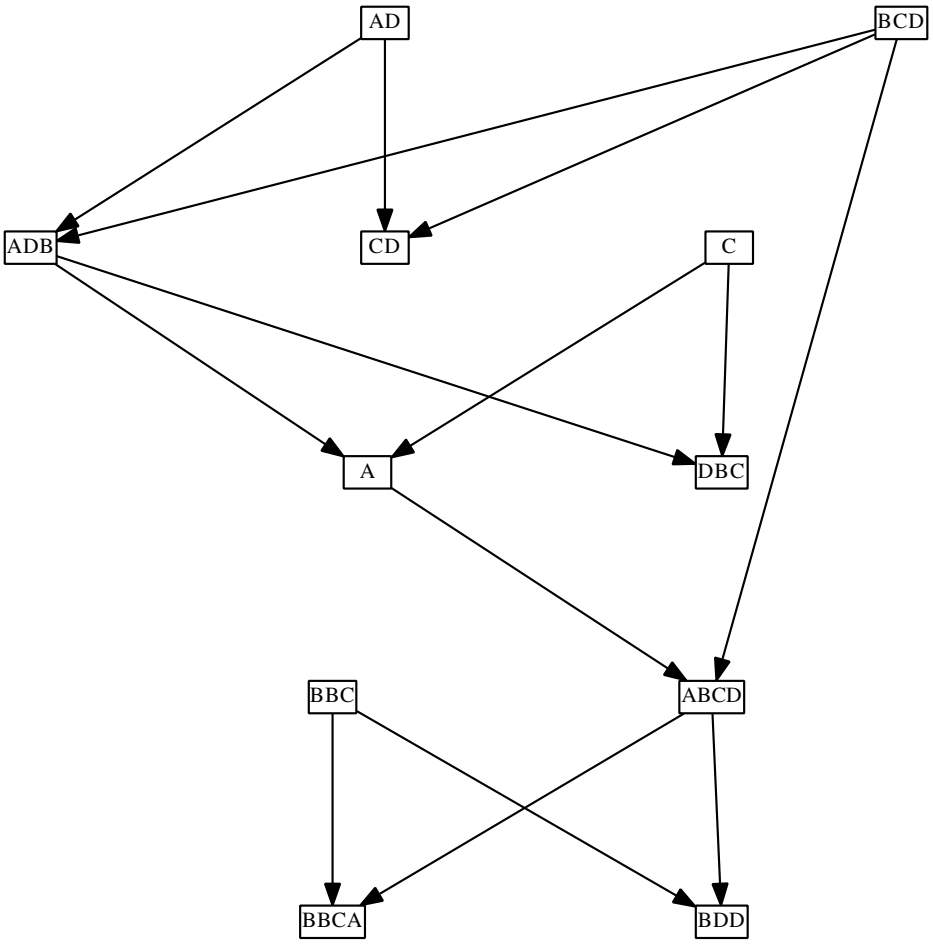



Figure 8.4.5: An example reaction network with reactions replaced by direct connections between each reactant and each product. This should be compared with figure 8.4.4 and figure 8.4.6 for an example of how loops are identified.

#### 8.4.4.2 Simulation parameters

The simulation of RBN-World from the initial elemental molecules to generate the reaction network also has some parameters associated with it that need to be specified. Firstly, the number of initial copies of each element has to be selected. If this value is too small, then there will be insufficient numbers of molecules to reliably exhibit interesting properties within the simulation. If this value is too large, then the simulation will take additional computation time to perform and the fitness evaluation may encounter computational issues such as insufficient memory. A maximum simulated duration also needs to be specified. This determines how much will be simulated, and should be the minimum to capture interesting properties without requiring excessive

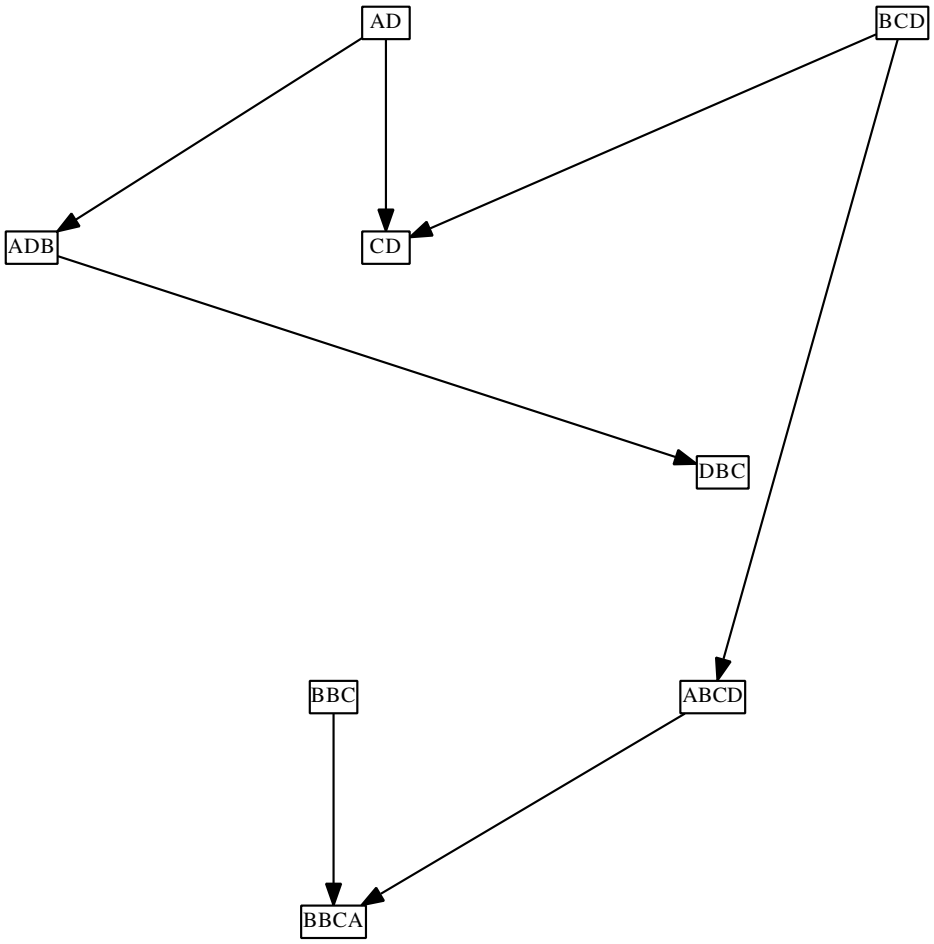



Figure 8.4.6: An example reaction network with reactions replaced by direct connections between each reactant and each product then filtered by 50% shared elemental graphs. This should be compared with figure 8.4.4 and figure 8.4.5 for an example of how loops are identified.

computational time. Furthermore, the simulated time is also related to the number of copies as the Gillespie-like mixing algorithm uses a fixed volume. As there is no prior evidence what these values should be they were estimated based on limited preliminary trials.

It was decided that 1,000 copies of each element would be used with a duration of 5 simulated time units. This was selected to balance several factors; CPU time, size (depth and width) of reaction network, reliability of reaction network.

The CPU time was the most important factor; because the simulation would be used in a genetic algorithm, it would be performed many times and therefore the CPU time for each simulation should be small. This means that

the simulation should run for a small number of time units and use a small number of copies of each element.

However, the size and reliability of the reaction network are also important. The size of the reaction network places an upper limit on the size of loops within it. The reliability of the reaction network influences how quickly the overall genetic algorithm converges; if reliability is too low then the genetic algorithm becomes too stochastic. Both size and reliability can be increased by increasing the number of copies and number of time units, but this then increases the average CPU time of a simulation. The relationships between number of copies / time units and distribution of CPU times is not straightforward; in particular there is a long tail of CPU time that has a disproportional large effect on the time required to evaluate a generation of the genetic algorithm. Rather than spend a large effort on investigating this relationship, 1,000 copies and 5 simulated time units were selected so that the search for interesting elemental subsets could proceed.

#### 8.4.4.3 Genetic algorithm parameters

For the genetic algorithm, several parameters are required. Most importantly, the size of the population must be selected is a balance between exploring the fitness landscape and searching regions in more detail. It was decided that 100 individuals in each generation would be an appropriate compromise, and would scale effectively over available computational hardware. The other key parameter needed for the genetic algorithm is the proportion of each generation to survive. A value of 0.10 was selected; higher than typical genetic algorithms which may only allow 0.01 or less to survive. This was chosen to enable resilience to stochastic fluctuations of the simulation; to reduce events where a "good" genome was removed from the population by bad luck. The genetic algorithm was run for 300 generations.

The fitness function also needs to be parametrized. Based on limited preliminary testing, a value of 50% similarity of elemental networks was selected, both for measures of "sufficiently similar" and for "too similar".

## 8.5 Results

There are two main components to the results of this genetic algorithm. Firstly, there is the analysis of the progress of the genetic algorithm to ensure that it identified more interesting elemental subsets as it progressed. Secondly, there is an in-depth analysis of one of the most elemental subsets.

### 8.5.1 Result of Genetic Algorithm

The results of the genetic algorithm showing the longest loop length present in the simulation of each elemental subset are shown in figure 8.5.1.

The progress of the genetic algorithm can be separated into two phases - an initially improving phase over the first 50 generations (approx) and then a plateau afterwards. One possible explanation for this is that the number of time units that the simulations ran for (which was estimated based on a random sampling) is too short to allow for longer loops to reliably form in the area of the search space that the genetic algorithm converged upon. This suggests that the search should be repeated with each simulation lasting for more time units, if more resources were available. This does not affect the validity of individual simulation for discovering interesting elemental subsets however, even if longer loop lengths might be possible within the wider system.

Although there are a few individuals with particularly high longest loop lengths (10 or more), these appear to be unstable as these results are not reproduced in future generations.

There are noticeably fewer individuals with a longest loop length of 5 molecules compared to longest loop lengths of 4 or 6. It is unknown why this should be the case; one hypothesis is that this phenomena is due to the non-factorial nature of prime numbers. This suggests that the emergent property of loop length is itself a rich and complex property, although further work would be required to confirm this.

#### 8.5.1.1 Analysis of a reaction network from an elemental collection

A reaction network generated by an individual elemental subset was selected for deeper analysis. This was not the elemental subset that generated the longest loop, rather it was selected as being representative of the group of elemental subsets that generate moderately long loops (8 molecules) yet having

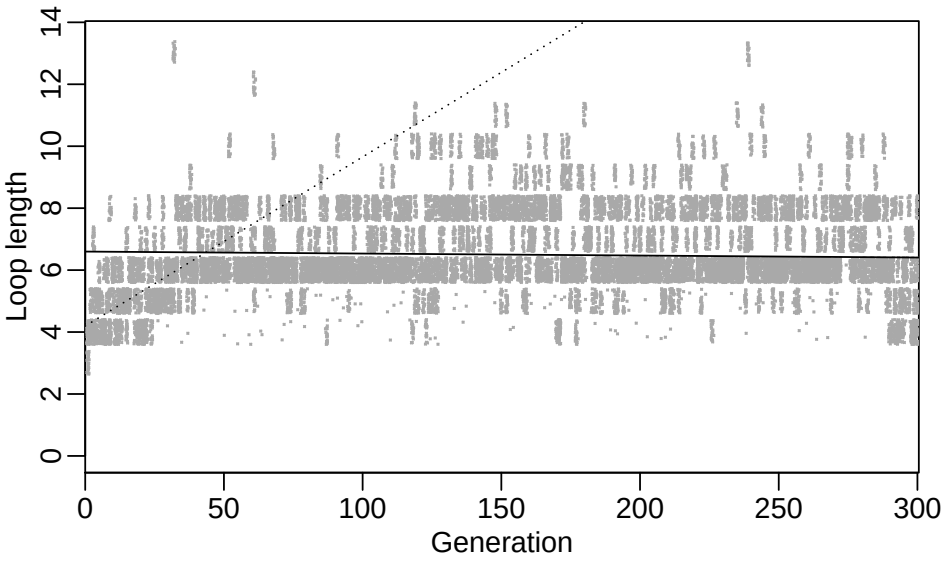



Figure 8.5.1: Plot of the longest loop length against generation of the genetic algorithm for parent individuals in each generation. Points are randomly perturbed $\pm 0.3$ on both axis to reduce complete overlapping of points with the same values. Dotted line indicates a linear regression fitted to the first 50 generations ($r^2 = 0.324$, $p << 0.001$). Solid line indicates a linear regression fitted to generations 50-300 ($r^2 = 0.000347$, $p << 0.001$).

few enough molecular species and reactions to be suitable for analysis. As the process of generating a reaction network from an elemental subset used in the genetic algorithm is stochastic, this reaction network may not contain all possible reactions. In particular, as RBN-World is (potentially) an open-ended Artificial Chemistry, it is highly likely that there are additional reactions that did not occur within the simulation. As such additional reactions may be unbounded, prolonging or repeating simulation will not completely address this problem. Alternative methods to generate the reaction network from the elemental subset, such as exhaustive enumeration, may address some of these issues, but they would raise further issues, e.g. dynamic properties. Due to these factors, this analysis is indicative rather than attempting to analyse the true reaction network.

In the reaction network generated from the elemental subset there were 1,286 reactions and 645 molecular species with 63 unique molecular structures (ignoring states and functional groups). The longest loop length was 8, and there were 5 different loops of that length that form 3 distinct groups; these are shown in figures 8.5.2, 8.5.3, and 8.5.4. For two pairs of loops there is a

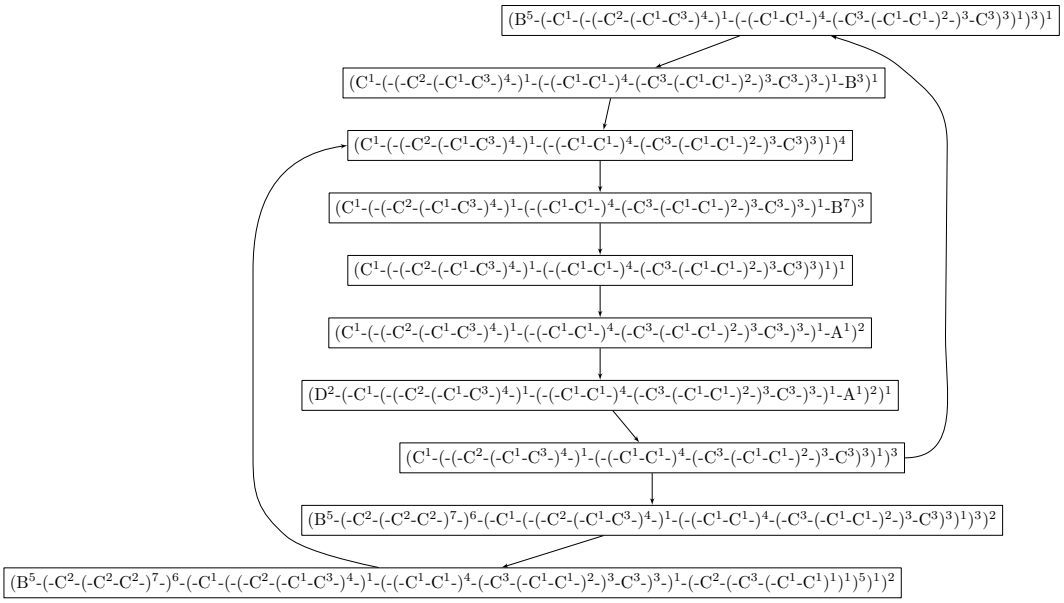



Figure 8.5.2: The first two of the five loops present in reaction network selected for in-depth analysis. Letters indicate element, numbers indicate state, brackets indicate functional groups.

shared subset of molecules.

More generally, the reaction network has 363 synthesis reactions and 515 decomposition reactions, as well as 517 directly catalysed — including 86 autocatalytic reactions. There are 142 divergent reactions, and there is also a single reversible reaction in the reaction network: $B^{10} + (B^2\text{–}A^1)^1 \leftrightarrow B^{12} + (B^2\text{–}A^1)^4$. This reversible reaction is not interesting in this particular reaction network, but its presence indicates that RBN-World is capable of such reactions in a wider context.

Even though this elemental subset was selected in part due to its relatively small number of molecules and reactions, there are still far too many to be visualized for manual examination. One approach is to simplify the reaction network to a sufficiently small subset. It can be seen from the loops shown in figures 8.5.2, 8.5.3 abd 8.5.4 that the elements are mostly C, with small amounts of A, B, and D. Therefore, examining C by itself may explain part of the wider reaction network.

Looking at the reactions that can occur starting from the molecular species $C^1$, there are only 3 reactions:

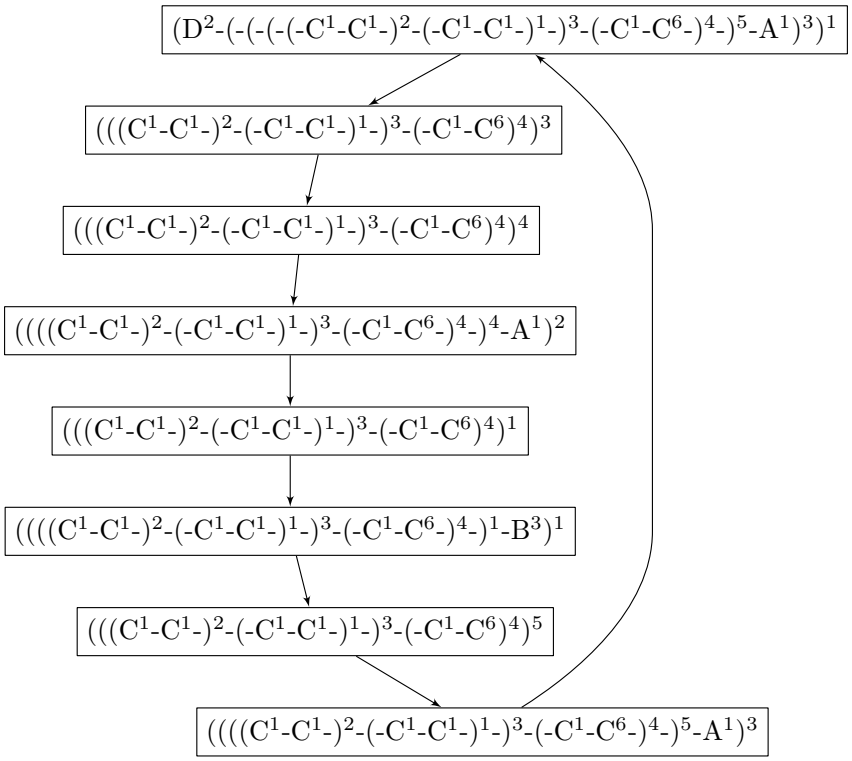



Figure 8.5.3: The third of the five loops present in reaction network selected for in-depth analysis. Letters indicate element, numbers indicate state, brackets indicate functional groups.

$$
\begin{array}{rcll}
C^1 & + & C^1 & \rightarrow (C^1\text{–}C^1)^1 \\
(C^1\text{–}C^1)^1 & + & C^1 & \rightarrow ((C^1\text{–}C^1\text{-})^1\text{–}C^1)^1 \\
C^1 & + & (C^1\text{–}C^1)^1 & \rightarrow (C^1\text{–}(\text{-}C^1\text{–}C^1)^1)^1
\end{array}
$$

After this point, no novel reactions occur and $((C^1\text{–}C^1\text{-})^1\text{–}C^1)^1$ and $(C^1\text{–}(\text{-}C^1\text{–}C^1)^1)^1$ become the most common molecules.

This lack of further development is an interesting feature. Although C is the main component of all the molecules in the loops, those molecules are considerably larger than the maximum size that molecules composed of C can reach in the absence of any other materials. This implies that although the other elements may not be permanently incorporated into the larger molecules, they do fill a critical role in their formation.

This interdependence of elements for higher-order properties can be seen if the reactions starting from $A^1$ and $C^1$ are examined. In addition to the reactions described above from $C^1$ alone, a number of other reactions exist where C and A interact, but unlike C atoms, a stable bond between two A atoms is never seen. There are two states of single atom A molecules — $A^1$

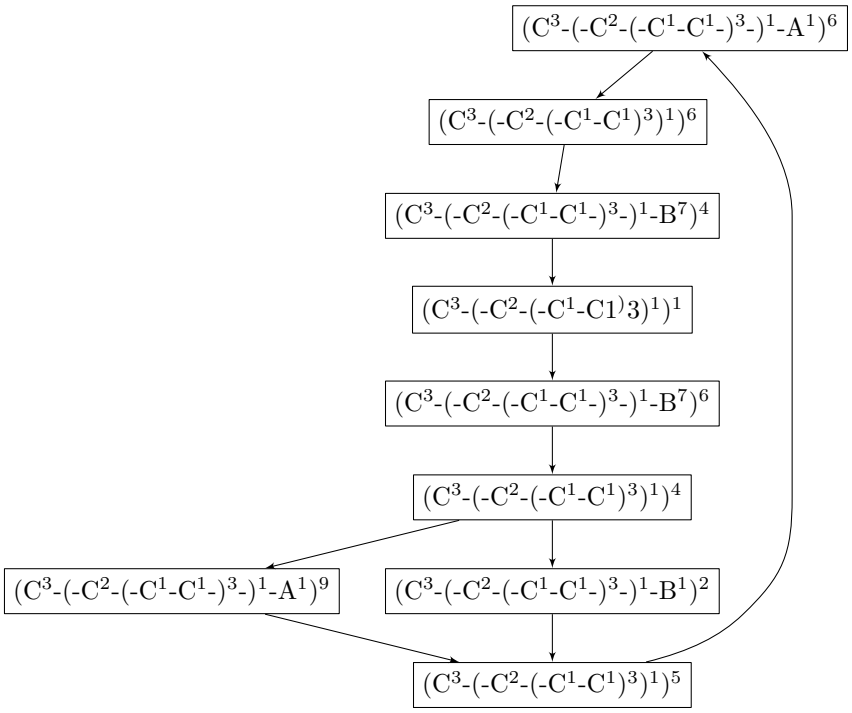



Figure 8.5.4: The fourth and fifth of the five loops present in reaction network selected for in-depth analysis. Letters indicate element, numbers indicate state, brackets indicate functional groups.

and $A^2$ — and each of these catalyses the change of itself into the other, as shown in the following reactions:

$$A^1 + A^1 \rightarrow A^1 + A^2$$
$$A^2 + A^2 \rightarrow A^1 + A^2$$

This type of reaction is highly unusual, and may not exist in natural chemistry. It is an opposite to autocatalysis, as rather than catalysing the production of itself both $A^1$ and $A^2$ catalyse their own change into another molecular species. Additionally, this is similar to a reversible reaction; any imbalance in the relative abundance of $A^1$ and $A^2$ will equalise over time.

In total, molecular species formed from the starting single atom molecules $A^1$ and $C^1$ account for 35 of 645 molecular species involved in 65 of 1,286 reactions including synthesis, decomposition, direct catalysis and direct autocatalysis. The only loop formed by elements of type A and C is the loop between $A^1$ and $A^2$ described above — no loops that are longer or contain larger molecules exist when the starting material is only $A^1$ and $C^1$.

If $B^1$ and $C^1$ are the initial starting materials, then 134 of 645 molecular species are involved in 173 out of 1,286 reactions. Again, this includes syn-

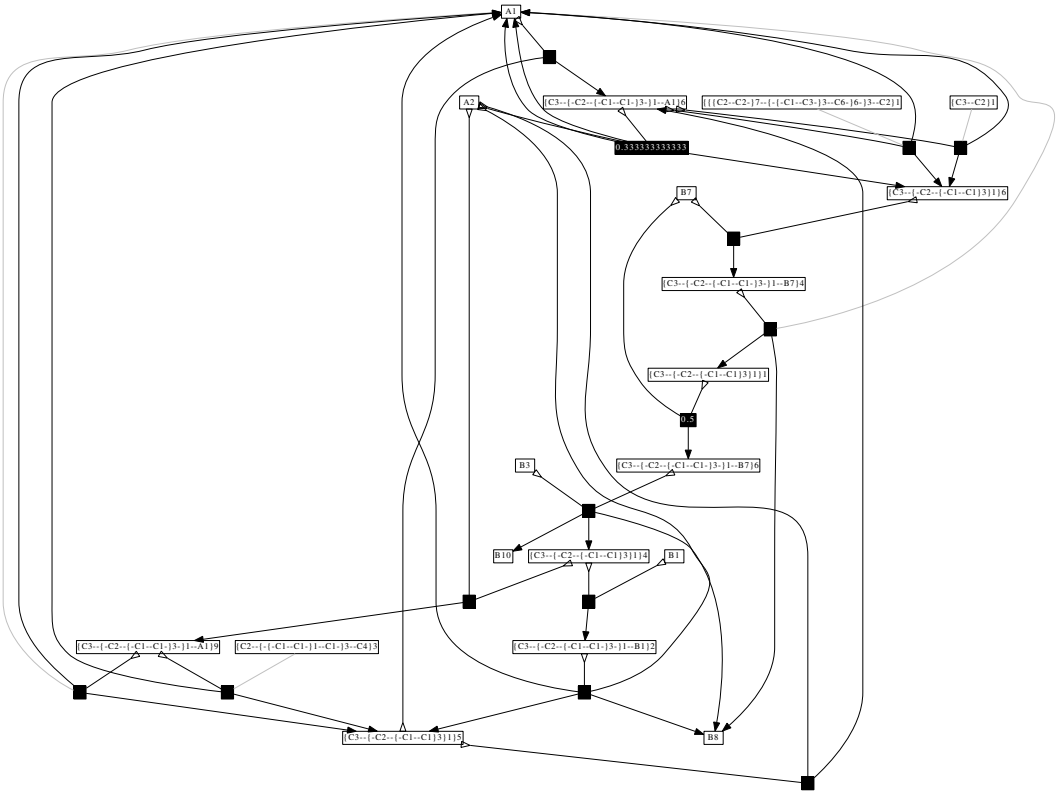



Figure 8.5.5: Reactions involved in fourth and fifth of the five loops present in the reaction network selected for in-depth analysis, as shown in figure 8.5.4.

thesis, decomposition, direct catalysis but not direct autocatalysis. There are loops formed from initial materials of $B^1$ and $C^1$, but these are much smaller than the loops in the full reaction network shown in figure 8.5.5. The longest involves only 4 molecular species — $(B^2–B^1)^1$, $(B^5–B^1)^2$, $(B^2–C^1)^2$, and $(B^4–B^7)^3$.

For the two-molecule initial mixture $D^1$ and $C^1$, $D^1$ is inert and undergoes no reactions which results in the same properties as if $C^1$ was alone.

If $C^1$ is omitted from the mixture and the reactions from an initial mixture of $A^1$ and $B^2$ are investigated, another different property is seen. In this subset of the reaction network, there are 133 reactions involving 26 molecular species which is a higher average number of reactions per molecular species than the other initial mixtures of only two elements. Furthermore, the largest molecular species consists of only two atoms. These features combined mean that there is a large number of states and transitions between them; single atom molecules of B have 7 states and single atom molecules of either $A^1$ or $A^2$ are involved in 67 of the 133 reactions. There are only 9 reactions where both reactants are

two-atom molecules, and 6 of these result in decay to 4 single-atom molecules; for example $(B^4–B^3)^3 + (B^5–A^1)^2 \rightarrow A^1 + B^3 + B^6 + B^7$. This suggests that both A and B are highly reactive and unstable elements with significant catalytic potential, as evidenced by the 8 direct autocatalytic and 41 direct catalytic reactions.

If $A^1$, $B^1$ and $C^1$ are all present in the initial mixture then 849 reactions (out of 1,286) involving 409 (out of 645) molecular species can occur. As well as synthesis, decomposition, direct catalysis and direct autocatalysis, this includes two of the five loops described in figures 8.5.2, 8.5.3 and 8.5.4. It is clear that this initial mixture is more than simply the sum of the initial mixtures $\{A^1,C^1\}$, $\{B^1\},C^1\}$ and $\{A^1,B^1\}$. Furthermore the initial mixture of $\{A^1,B^1,C^1\}$ appears to support complex structures with catalytic activity. This is indicated by a total of 291 directly catalytic reactions from 110 catalysts. Many of these catalysts appear to derive their catalytic activity from the structure of the molecule, rather than simply their elemental composition. For example, there are 43 direct catalysts composed only from element C (e.g. $(((C^1–C^1\text{-})^2–(\text{-}C^1–C^1\text{-})^1\text{-})^3–(\text{-}C^1–C^6)^4)^4$), an element that had not shown catalytic behaviour in the smaller initial mixture subsets.

Although reducing the reaction network to an initial mixture of two out of $A^1$, $B^1$ and $C^1$ has highlighted several interesting features, in each combination the longer loops are not present. It is only in an initial mixture of all three of $A^1$, $B^1$ and $C^1$ that two of the five loops of length 8 are present.

The reactions involved in these loops are shown in figure 8.5.5. From this, it can be seen that, overall, each complete repeat of the loop consumes and produces several molecular species, shown in table 8.5.1. It is interesting to note that $A^2$ is consumed by the loop and $A^1$ is produced; these molecular species react to equalize each others concentrations as discussed above. In figure 8.5.5 it can also be seen that for some steps in the loop there are multiple reactions that can occur; for example, there are three reactions with $(C^3–(\text{-}C^2–(\text{-}C^1–C^1\text{-})^3\text{-})^1–A^1)^6$ as a reactant and $(C^3–(\text{-}C^2–(\text{-}C^1–C^1\text{-})^3\text{-})^1)^6$ as a product. Such alternative pathways are interesting as it suggests that the loops may be robust to some perturbations, and that there is variation not only in members of a loop but also between the reactions which compose a loop.

Further interpretation and analysis is possible, but is beyond the available resources and therefore could be performed as future work.

| Consumed | Produced |
|---|---|
| $A^2$ ×4 | $A^1$ ×6 |
| $B^1$ | $B^8$ ×3 |
| $B^3$ | $B^{10}$ |
| $B^7$ ×2 | |

Table 8.5.1: Table of the molecular species consumed and produced by the loops shown in figure 8.5.4 and 8.5.5.

## 8.6 Conclusion

In this chapter, a genetic algorithm for discovering elemental subsets of bRBNs exhibiting interesting emergent structure — in this case extended reaction loops — has been described. This included a representation of elemental bRBN sets suitable for use in a search algorithm, as well as methods for generating and evaluating reaction networks derived from those elemental subsets. A genetic algorithm lead to the identification of many elemental subsets with emergent loops of a range of lengths. Further examination of one of these elemental subsets revealed rich, emergent, and complex interactions; it is a reasonable assumption that other elemental subsets would also have such properties.

A number of choices were made during the design and specification of the search. These include:

- use of 'proportion sum one' rather than 'total sum zero' as the bonding property and comparison.
- use of simulation rather than exhaustive enumeration.
- selection of a Gillespie-like mixing algorithm
- number of elemental copies and duration in simulations.
- the "sufficiently similar" comparison ignores state and functional groups.
- stochastic fluctuations in the fitness function rather than being averaged into an approximately deterministic metric.

All of these may affect the results of the search, and given sufficient time should be investigated as further work.

Furthermore, only one reaction network was investigated in more depth. As a result of this investigation, several additional questions present themselves -

how many molecular species and reactions are typically present? how large are molecules in a typical reaction network? how stable are the dynamics of these reaction networks? — and many other questions. It is tempting to try to use the reaction networks generated here to answer some of these questions, but it should be noted that the reaction networks are not a random sample. Therefore it is not an appropriate data source to answer many of these questions.

It should also be noted that the reaction network analysed in depth here is only one of many reaction networks — not only of the reaction networks identified in the genetic algorithm, but also from other possible elemental subsets or from other methods of generating reaction networks from elemental subsets. However, the reaction network examined in depth has shown that RBN-World can contain interesting reaction network features and many of those features would not have been expected *a priori*. This encourages the use of RBN-World as a base for future work.

# Chapter 9

# Conclusions

This thesis has covered a range of inter-related topics around Artificial Chemistries for Artificial Life. The first stage was to describe the wider contexts of both Artificial Chemistry and Artificial Life. This included addressing the separation of simulations of existing living systems from virtual environments containing novel living systems, as well as addressing some of the prejudices and assumptions these topics are often approached with.

The re-implementation of an existing Artificial Chemistry was an early investigation to explore previous work on the topic. However, this was an unexpectedly challenging proposition due to the combination of a fragile Artificial Chemistry and a non-specific specification that omitted details such as the value of constants. Re-implementation of an existing Artificial Chemistry for independent verification is an important step for scientific reproducibility of results, yet is rarely done in practice. In some cases, the source code of a program that implements an Artificial Chemistry may be available; however, this can only be supplementary to other published materials and does not address the issue of independent verification.

The issue of re-implementation of Artificial Chemistries combined with the lack of independent verification also impacts how newcomers to the topic approach it. Because there are no community standards for quality control and no software libraries or other re-usable assets, newcomers are forced to start by writing their own software from scratch. This is reinforced by the multitude of disciplines Artificial Chemistries may be approached from — for Artificial Life, for simulation of chemical systems, or for more abstract applications. Solving these issues is challenging, but one approach would be an easy-to-use software

tool-kit and application suite for Artificial Chemistries supported by an online archive and repository. By providing a resource for creating and distributing Artificial Chemistries, this could anchor a disparate community and encourage informal peer discussion and review. Such issues are not unique to Artificial Chemistry, affecting other fields seeking to use computational resources such as biological simulations.

A major new type of Artificial Chemistry has been described in this thesis; sub-symbolic representation (§4.1.1.3). Although this was identified from comparisons between Artificial Chemistry and natural Chemistry, sub-symbolic representations have never been used previously; all other Artificial Chemistries have been symbolic or structured symbolic. Sub-symbolic representation shows considerable promise for Artificial Chemistry with rich emergent properties, as demonstrated in chapters 7 & 8. Furthermore, other researchers in the field have already begin to use sub-symbolic representations in other Artificial Chemistries (Hatcher et al., 2011).

Compared to Hutton's Artificial Chemistry (Hutton, 2004b, 2007), RBN-World has overcome and addressed both the issues of emergence and of computational tractability. RBN-World has been shown in chapter 8 to be capable of showing many of the interesting emergent properties of Artificial Chemistries identified in chapter 4, such as autocatalytic sets. Because of the emergent nature of RBN-World, during a simulation novel molecules can be introduced and will interact with the existing system. This was shown indirectly by chapters 7 and 8. Although calculating the reactions of novel material increases computational requirements for the reactions, RBN-World compensates for this by separating intra- and inter- molecular space rather than Hutton's two-dimensional combined molecular space mixing algorithm.

This work has also clarified many of the terms used in relation to Artificial Chemistry for Artificial Life, such as catalysis and hyper-cycles. In previous works, these terms have been interpreted in multiple different ways which has lead to confusion and miscommunication (e.g. hypercycles §4.3.9). Additionally, this revised terminology was used to investigate Artificial Chemistries based on the presence or absence of properties. This approach was demonstrated in chapter 7 by application to variants of RBN-World, but it could also be used to compare more radically different Artificial Chemistries. While other tools exist to compare Artificial Chemistries, such as Chemical Organization Theory (Dittrich and di Fenizio, 2007), these have not been applied multiple

Artificial Chemistries simultaneously.

The initial motivation for using an Artificial Chemistry approach to Artificial Life was the aim of rich and complex emergent properties from relatively simple components, with the ultimate goal of emergent living systems. Although emergent living systems were not investigated, it has been clearly demonstrated that RBN-World can lead to rich and complex emergent properties such as reaction loops. However, it is not clear which of the features of RBN-World resulted in these emergent properties. Are they a result of a specific component of the Artificial Chemistry such as the bonding rule, bonding property or reaction algorithm? Or are these higher emergent properties a consequence of lower level ones such as functional groups or multiple states? These properties could also be a product of the stochastic simulation used to generate the reaction networks these properties were measured from.

There are some limitations of the work described in this thesis. Although RBNs as a sub-symbolic representation have shown emergence of interesting properties, no alternatives to RBNs have been considered here. These alternatives may be better suited to artificial life investigations, but are outside of the scope of this thesis. Furthermore, RBN-World has a number of features that has not been specifically investigated previously — such as multiple states, functional groups — and it is not know to what extent each of these features contributes to the emergent properties. It is possible that some of these are not required, and therefore that RBN-World can be simplified accordingly.

Another issue with these investigations of RBN-World is that they make a number of general assumptions. For example, the search in chapter 8 assumes that the reaction network generated from a simulation is representative of the underlying, potentially unbounded, reaction network. Furthermore, it is also assumed that a potential reaction loop identified in a reaction network is an interesting property that is indicative of potential for life-like systems.

## 9.1 Future Work

There are several directions that future would could proceed in. One direction would be to develop RBN-World Artificial Chemistries with the aim of emergence of life-like systems. For example, in this thesis only aspatial mixing algorithms were used for RBN-World. This prevents properties such as membranes and cells, and therefore the emergence of systems that could

be intuitively recognizable as "living". Using other mixing algorithms, such as a discrete grid of locally aspatial regions or combined intra- inter-molecular space, may enable higher-order emergence.

Related to this, larger and longer simulations could be performed in order to give more opportunity for life-like properties to emerge. Some of the techniques developed in chapter 8 could be extended to identify interesting features in long-running simulations while the are still on-going, building towards an interactive experimental tool-kit for manipulation and feedback. This could include the use of advanced computational techniques and optimizations to the simulation software to improve performance.

An alternative direction for future work would be to investigate additional alternatives within RBN-World. Although the Artificial Chemistries examined in this thesis showed that RBN-World has interesting properties, it is not clear which aspects of RBN-World these properties depend upon. For example, it would be interesting to know if those properties still emerged if molecules and functional groups only had a single possible state rather than the multiple states described here. The existence of functional groups is also an interesting possibility to investigate, and one that could be done simply by always forming bonds at the lowest level.

In the wider context of Artificial Chemistries other tha RBN-World, it would be interesting to develop sub-symbolic representations using systems other tha RBNs. This could include many different representations, some of which could include mathematical functions (e.g. similar to AlChemy (Fontana, 1991)) or image comparison (e.g. generated from Mandlebrot sets (Bentley, 2003)). Such wider investigations could establish what features are intrinsic to a sub-symbolic representation and which depend upon the specific sub-symbolic representation selected. As the choice to use RBNs as the sub-symbolic representation in RBN-World was based on limited information. As a discrete dynamical system that is computationally tractable yet also spans a wide range of behaviours, RBNs met the appropriate criteria. It is not expected that RBNs are the best representation however; others may be more suitable for particular emergent properties.

All of these directions for future work depend upon comparisons between Artificial Chemistries. However, there are few 'null models' of Artificial Chemistry to use as a reference. Generating reaction networks at random is one possibility, such as the system used by Kauffman to investigate autocatalytic

sets (Kauffman and Farmer, 1986). However, in graph theory the algorithm used can have significant impact on the resulting properties, and it is expected that this is also true for Artificial Chemistry reaction networks. For example, some methods of generating random reaction networks would have intrinsic properties such as conservation of mass, multiple states, or functional groups. Determining the extent to which these intrinsic properties give rise to higher-order properties without the need for more complicated Artificial Chemistries would be an interesting challenge.

Overall, it is clear that Artificial Chemistry in general and RBN-world in particular is a young and vibrant research area with great potential for future discoveries.

# Bibliography

W. Banzhaf. Self-replicating sequences of binary numbers. Foundations I: General. *Biological cybernetics*, 69(4):269–274, 1993a.

W. Banzhaf. Self-replicating sequences of binary numbers. Foundations II: Strings of length N= 4. *Biological Cybernetics*, 69(4):275–281, 1993b.

W. Banzhaf. Self-replicating sequences of binary numbers. *Computers and Mathematics with Applications*, 26(7):109–118, 1993c.

W. Banzhaf. Self-replicating sequences of binary numbers: The build-up of complexity. *Complex Systems*, 8(3):215–225, 1994.

N. Barricelli. Symbiogenetic evolution processes realized by artificial methods. *Methodos*, 9(35-36):143–182, 1957.

P. Bentley. Evolving fractal proteins. In *Proceedings of the fifth international conference on evolvable systems: from biology to hardware*, pages 81–92. Springer-Verlag, 2003.

P. Dittrich and P. S. di Fenizio. Chemical organisation theory. *Bulletin of Mathematical Biology*, 69(4):1199–1231, 2007.

P. Dittrich, J. Ziegler, and W. Banzhaf. Artificial chemistries — a review. *Artificial Life*, 7(3):225–275, 2001.

M. Eigen and P. Schuster. The Hypercycle: Part A. *Naturwissenschaften*, 64 (11):541–565, 1977.

M. Eigen and P. Schuster. The Hypercycle: Part B. *Naturwissenschaften*, 65 (1):7–41, 1978a.

M. Eigen and P. Schuster. The Hypercycle: Part C. *Naturwissenschaften*, 65 (7):341–369, 1978b.

A. Faulconbridge, S. Stepney, J. F. Miller, and L. Caves. RBN-World: The hunt for a rich AChem. In *Proceedings of twelfth international conference on the synthesis and simulation of living systems (ALife XII)*, pages 261–268, 2010.

A. Faulconbridge, S. Stepney, J. F. Miller, and L. S. D. Caves. RBN-World: a sub-symbolic artificial chemistry. In *Advances in Artificial Life*, volume 5777 of *Lecture Notes in Computer Science*, pages 377–384. Springer, 2011.

W. Fontana. Algorithmic chemistry. *Artificial life II*, pages 159–209, 1991.

M. Gardner. The fantastic combinations of John Conway's new solitaire game "life". *Scientific American*, 223(4):6–10, 1970.

M. Gibson and J. Bruck. Efficient exact stochastic simulation of chemical systems with many species and many channels. *Journal of Physical Chemistry*, 104(9):1876–1889, 2000.

D. T. Gillespie. Exact stochastic simulation of coupled chemical reactions. *The Journal of Physical Chemistry*, 81(25):2340–2361, 1977.

G. H. Hardy. Mendelian proportions in a mixed population. *Science*, 28(706): 49–50, 1908.

M. Hatcher, W. Banzhaf, and T. Yu. Bondable cellular automata. In *Proceedings of the eleventh european conference on advances in artificial life (ECAL '11)*, pages 326–333, 2011.

S. Hickinbotham, A. Faulconbridge, and A. Nellis. The Blind Watchmaker's Workshop: three Artificial Chemistries around Eigen's Paradox. In *Proceedings of twelfth international conference on the synthesis and simulation of living systems (ALife XII)*, pages 82–89. MIT Press, 2010.

S. Hickinbotham, E. Clark, S. Stepney, T. Clarke, A. Nellis, M. Pay, and P. Young. Molecular microprograms. In *Advances in Artificial Life*, volume 5777 of *Lecture Notes in Computer Science*, pages 297–304. Springer, 2011.

W. Hordijk and M. Steel. Detecting autocatalytic, self-sustaining sets in chemical reaction systems. *Journal of Theoretical Biology*, 227(4):451–461, 2004.

T. Hutton. Making membranes in artificial chemistries. In *Workshop and Tutorial Proceedings, Artificial Life IX*, pages 71–76. The MIT Press, 2004a.

T. Hutton. A functional self-reproducing cell in a two-dimensional artificial chemistry. In *Artificial life IX: proceedings of the ninth international conference on the simulation and synthesis of artificial life*, pages 444–449. The MIT Press, 2004b.

T. Hutton. Evolvable self-reproducing cells in a two-dimensional artificial chemistry. *Artificial Life*, 13(1):11–30, 2007.

M. Jeschke, A. Park, R. Ewald, R. Fujimoto, and A. M. Uhrmacher. Parallel and Distributed Spatial Simulation of Chemical Reactions. In *Proceedings of the 22nd Workshop on Principles of Advanced and Distributed Simulation*, pages 51–59. IEEE, 2008.

S. A. Kauffman. Metabolic stability and epigenesis in randomly constructed genetic nets. *Journal of Theoretical Biology*, 22(3):437–467, 1969.

S. A. Kauffman. *The Origins of Order: Self-Organization and Selection in Evolution*. Oxford University Press, 1993.

S. A. Kauffman and J. D. Farmer. Autocatalytic sets of proteins. *Origins of Life and Evolution of the Biosphere*, 16(3-4):446–447, 1986.

D. E. Koshland. The seven pillars of life. *Science*, 295(5563):2215–22156, 2002.

C. Langton. Self-reproduction in cellular automata. *Physica D: Nonlinear Phenomena*, 10(1-2):135–144, 1984.

C. Langton. Studying artificial life with cellular automata. *Physica D: Nonlinear Phenomena*, 22(3):120–149, 1986.

A. Lindenmayer. Mathematical models for cellular interactions in development I. Filaments with one-sided inputs. *Journal of Theoretical Biology*, 18(3): 280–299, 1968a.

A. Lindenmayer. Mathematical models for cellular interactions in development II. Simple and branching filaments with two-sided inputs. *Journal of Theoretical Biology*, 18(3):300–315, 1968b.

E. N. Lorenz. Deterministic nonperiodic flow. *Journal of the Atmospheric Sciences*, 20(2):130–141, 1963.

A. Lotka. *Elements of physical biology.* Williams and Wilkins company, 1925.

H. R. Maturana and F. J. Varela. *Autopoiesis and cognition: The realization of the living.* Springer, 1980.

C. McKay. What is Life — and how do we search for it in other worlds? *PLoS Biology*, 2(9):e302, 2004.

S. Nishio. *Geminoid: Teleoperated android of an existing person.* I-Tech Education and Publishing, 2007.

J. Peterson. *Petri Net Theory and the Modeling of Systems.* 1981.

P. Prusinkiewicz, A. Lindenmayer, and J. Hanan. *The algorithmic beauty of plants.* Springer-Verlag, 1990.

T. S. Ray. An approach to the synthesis of life. *Artificial Life II*, pages 371–408, 1992.

J. Riskin. The defecating duck, or, the ambiguous origins of artificial life. *Critical Inquiry*, 29:599–633, 2003.

Y. Sakagami and R. Watanabe. The intelligent ASIMO: System overview and integration. In *IEEE/RSJ International Conference on Intelligent Robots and Systems*, volume 3, pages 2478–2483, 2002.

H. Sayama. Swarm chemistry. *Artificial Life*, 15(1):105–14, 2009.

P. Smolensky. On the proper treatment of connectionism. *Behavioral and Brain Sciences*, 11(01):1–23, 1988.

M. Steel. The emergence of a self-catalysing structure in abstract origin-of-life models. *Applied Mathematics Letters*, 13(3):91–95, 2000.

H. Suzuki. Network artificial chemistry — molecular interaction represented by a graph. In *Proceedings of the ninth international conference on the simulation and synthesis of living systems (ALife IX)*, pages 63–70, 2004.

H. Suzuki, N. Ono, and K. Yuta. Several necessary conditions for the evolution of complex forms of life in an artificial environment. *Artificial Life*, 9(2):153–174, 2003.

D. L. Theobald. A formal test of the theory of universal common ancestry. *Nature*, 465(7295):219–222, 2010.

A. Turing. The chemical basis of morphogenesis. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, 237(641):37–72, 1952.

S. S. Vogt, R. P. Butler, E. J. Rivera, N. Haghighipour, G. W. Henry, and M. H. Williamson. The Lick-Carnegie Exoplanet Survey: A 3.1 M⊕ Earth Planet in the Habitable Zone of the Nearby M3V Star Gliese 581. *Astrophysical Journal*, pages 954–965, 2010.

L. Yamamoto and W. Banzhaf. Catalytic search in dynamic environments. In *Proceedings of twelfth international conference on the synthesis and simulation of living systems (ALife XII)*, pages 277–284, 2010.

Y. L. Yung and W. B. DeMore. *Photochemistry of Planetary Atmospheres*. Oxford University Press, 1998.