

# Sub-Symbolic Artificial Chemistries

Penelope Faulkner, Mihail Krastev, Angelika Sebald  
and Susan Stepney

**Abstract** We wish to use Artificial Chemistries to build and investigate open-ended systems. As such, we wish to minimise the number of explicit rules and properties needed. We describe here the concept of *sub-symbolic Artificial Chemistries* (ssAChems), where reaction properties are emergent properties of the internal structure and dynamics of the component particles. We define the components of a ssAChem, and illustrate it with two examples: RBN-world, where the particles are Random Boolean Networks, the emergent properties come from the dynamics on an attractor cycle, and composition is through rewiring the components to form a larger RBN; and SMAC, where the particles are Hermitian matrices, the emergent properties are eigenvalues and eigenvectors, and composition is through the non-associative Jordan product. We conclude with some ssAChem design guidelines.

## 1 Introduction

Artificial Chemistries (AChems) are examined for many reasons. In the context of Artificial Life (ALife) they can form an underpinning technology. In these cases, the systems are often carefully crafted, with atoms and reaction rules hand tuned to produce the desired properties, such as replication and evolution.

Another use of AChems is to study open-ended systems in general [6]. In such cases it is important to ensure that the open-ended properties are not specifically

---

P. Faulkner · A. Sebald  
Department of Chemistry, University of York, York, UK  
e-mail: pf550@york.ac.uk

A. Sebald  
e-mail: angelika.sebald@york.ac.uk

M. Krastev · S. Stepney (✉)  
Department of Computer Science, University of York, York, UK  
e-mail: susan.stepney@york.ac.uk

M. Krastev  
e-mail: mk599@york.ac.uk

designed in, but rather emerge from the properties underlying system. Such behaviour is needed to allow the possibility of multiple levels of emergence.

In 2009, Faulconbridge, Stepney, Miller and Caves presented the first work on a *subsymbolic artificial chemistry* (ssAChem), called bRBN-world [11], with subsequent work described in [9, 10]. The core idea of ssAChems is that reaction properties should be emergent properties of the internal structures of the relevant atoms and molecules, analogous to the way that these are emergent properties of the electronic structures of physical atoms and molecules. The behavioural specifics emerge from these structures, rather than being solely defined by external rules.

We have continued work on ssAChems, enriching and diversifying the original concept. In this chapter we bring together some of the recent work. We start with some background material on AChems in general (Sect. 2). We then give a formal definition of ssAChems (Sect. 3). We use this definition to introduce a new ssAChem, SMAC, based on matrix algebra (Sect. 4). We also use the definition to summarise bRBN-world, and add a new feature: an environmental temperature model (Sect. 5). We finish with some guidelines to follow when designing a new ssAChem (Sect. 6).

## 2 Why AChems?

### 2.1 What Chemistry Can Give ALife/Complexity Science

When we take a good look at the contemporary activities carried out under the broad umbrella of ‘chemistry’ we are confronted with a vast range of fascinating developments and output. An increasing proportion of these research and development activities no longer fits the traditional boxes of the various science disciplines, but instead has moved to cover areas in between classical disciplines, ranging from shared approaches with physics, engineering, materials science, biology, pharmacology, archaeology, forensic sciences all the way to medical sciences.

This contemporary state of affairs can act as an encouragement for our endeavour to design subsymbolic artificial chemistries (we come back to various definitions in a moment): whatever theoretically underpins this wide range of natural activities must surely cater well for the emerging properties of a rich system.

Traditionally, the discipline has been subdivided into the classical domains of organic, inorganic and physical chemistry. This subdivision is still maintained in many areas of undergraduate teaching and in organisational structures of many chemistry departments. However, its mainly output/substance focus cannot help us with finding inspiration for the rational design of artificial chemistries.

If we want to take inspiration from chemistry for the design of algorithms with emergent properties we need to take a step back, and take a dispassionate and perhaps slightly unconventional look at the very core of what forms the basis of the conceptual and descriptive framework that helps us to deal with chemistry. Stripping matters back to the essentials yields the following picture.

Nature provides a set of possibilities: the elements in the periodic table. Their properties in turn provide variability, and their number forms the basis of endless options for combinatorics constrained by reaction properties. In practice much of the rich behaviour of chemical systems relies on a fairly small subset of chemical elements. Biochemistry produces its own larger ‘units’ from this small subset of chemical atoms: the four unit genetic code, and the 20 unit amino acid code. Primary linear combinatoric systems (DNA and proteins) then exploit spatial properties to gain functional secondary and tertiary structures. In each case, the properties of the units emerges from their internal structure. Further exotic units can be added to these small alphabets in certain circumstances. The properties of these further units emerge in a similar manner.

Faced with having to understand, handle, predict or manipulate a massive set of possibilities, the discipline of chemistry has developed empirical rules to come to grips with this richness in nature. The basic set of rules in use is limited and fairly straightforward. The energetics of the systems considered are described by thermodynamics; kinetics describe rates of change (and often reaction mechanisms) of the systems as macroscopic entities. The laws of quantum mechanics describe microscopic properties such as electronic structures of atoms and molecules. Statistics, mainly in the form of statistical thermodynamics, provides the necessary link between macroscopic and microscopic properties.

This small but effective set of rules in operation deals with essentially the entire vast range of ‘chemistry’. Note that (i) none of these rules applied to ‘chemistry’ make any reference to chemical properties, and (ii) all of the many and varied properties of chemical systems emerge from the combination of a set of possibilities with a set of rules. This ‘chemistry rule book’ seems to be quite capable of dealing with the rich natural system.

Let us translate this highly minimalistic view of the working of the foundations of chemistry to our task of recreating similarly rich emergent properties *in silico*. To do this we need recognise that chemistry uses descriptive rules that are a scientific model built from observation of what the system does; such a model may be wrong or partial. AChems, on the other hand, implement underlying rules that govern what the simulation system does; such rules cannot be wrong, cannot be broken. Thus we must choose our simulation rules judiciously, not to rigidly enforce one level of behaviour, but to provide sufficient richness of possibilities that higher level rule-based behaviours can emerge.

A stripped down view of the underpinning rules of chemistry lends itself to being applied in the rational design of algorithms with emergent properties: we need (i) to pick a sufficiently large and versatile set of possibilities (‘atoms’) and (ii) combine this set with a small but powerful set of rules to govern the possibilities and steer the system toward a sweet spot of just enough complexity. Nature does this fabulously well, so there is no immediately obvious reason why we should not be able to recreate such behaviours *in silico*, as long as we are able to do so in a computationally feasible manner.

## 2.2 Quick Definition of AChems

An artificial chemistry (AChem) is a computational system that is analogous to certain aspects of the dynamics of atomic and molecular level interactions of real chemistry. AChems are not typically used as *simulations* of real chemistry, but rather exploit combinatorics, dynamics, and other properties in a computational setting. Often, AChems are used to explore aspects of Artificial Life, and aspects of the origin of life. Here we are interested in exploring how open-ended systems can develop as a consequence of emergent properties.

Dittrich et al. [7] formally define an AChem as a triple  $(\mathbb{S}, \mathbb{R}, \mathbb{A})$ , where  $\mathbb{S}$  is the set of possible molecules,  $\mathbb{R}$  is the set of rules for interacting the molecules, and  $\mathbb{A}$  is an algorithm describing the dynamics of the environment, the reaction vessel, and how the rules are applied to the molecules.  $\mathbb{S}$  and  $\mathbb{R}$  are more fundamental to the AChem, and may be thought of as the underlying ‘physics’ of the system;  $\mathbb{A}$  may then be varied to see how the specified molecules and rules behave in different environments or contexts, for example, spatial *versus* aspatial, closed *versus* chemostat.

$\mathbb{S}$  and  $\mathbb{R}$  can be defined explicitly (by listing all possible molecules, and all possible interactions), or implicitly, using a procedural algorithm or declarative expression. For an open-ended system, both  $\mathbb{S}$  and  $\mathbb{R}$  need to be defined implicitly, as novel molecules and their interactions can continually arise.

## 2.3 Historical Context

AChems were originally created as an addition to the study of artificial life in the hopes that they would illuminate the transition from inanimate to animate matter through complex chemistry. Three properties are thought to be the basis for open-ended evolution and complexification of life [2, 8, 24]

1. *self-replication*: a property that directly or indirectly propagates the creation of copies of itself [14, 15, 23]
2. *metabolism*: a lifeform’s ability to change and maintain itself in its life time, by processing energy to reproduce, repair damage, create or destroy structures [16, 23]
3. *mutability*: the capability for minor changes to occur during replication [8], needed for life capable of evolutionary behaviours

From these requirements of life the need for several high level emergent properties of chemistries have been inferred [11]: autocatalytic sets [17] (a precursor of self-reproduction); hypercycles [8]; and heteropolymers (giving an information-bearing molecule made up from a set of repeating subunits, for example RNA and DNA [22]).

## 2.4 Desirable Properties of an AChem

Complex *emergent* properties such as self-replication and metabolism should not be designed *into* an AChem. Instead, we can define a set of low level requirements of a chemistry that can be tested at the level of small ‘molecules’ comprising only a few atoms [9, 11, 25]. These properties are not sufficient for these and other higher level properties to emerge, but are necessary to allow them to exist.

### 2.4.1 Macroscopic Properties

- *unbounded molecular size*: an open-ended system must have some aspect of its state space unbounded in size; a bounded system could (at least potentially) exhaust its state space.
- *conservation of mass*: some form of conservation law (energy, mass, particle number) is needed to stop the system simply being a white hole; we conserve atoms, in a context where our reaction vessel may have inflows and outflows.

### 2.4.2 Microscopic Properties

- *synthesis*: the forming of a bond between two molecules, producing a new larger molecule:  $A + B \rightarrow D$ . The products of synthesis (and decomposition, below) should be the same kind of objects as the reactants, so that these can then react without the need to introduce new kinds of rules.
- *self-synthesis*: the forming of a bond between two identical molecules:  $A + A \rightarrow D$ . The overall goal is start from a relatively small set of atoms, so we must allow some atoms and molecules to bind with copies of themselves. However, self-synthesis should not be universal, as we do not want every randomly selected set of atoms to form a molecule. If all molecules were possible, the system would be the basic combinatorics of all possible combinations of the atoms. We want the system to have the same complexity as real chemistry, so we need implicit restrictions to the set of possible molecules.
- *decomposition*: the removal of a bond in a molecule (which may affect other bonds in the molecule):  $D \rightarrow A + B$ . Decomposition allows for the potential formation of structures that cannot be formed directly by synthesis, for example by allowing the use of scaffolding. It also allows cycles in reaction networks, for example, metabolic cycles as a combination of anabolic (building up) and catabolic (breaking down) reactions.

### 2.4.3 Environmental and Contextual Properties

Molecules and reactions should be able to be influenced by environmental or contextual conditions, such as space and temperature analogues. This allows the same ‘physics’ of the molecules to exhibit different behaviours in different contexts, and so for that behaviour to potentially be controlled, regulated, or influenced.

Additionally, molecules and reactions should be able to influence their environment, permitting feedback cycles so that the system can influence its own dynamics. For example, endothermic and exothermic reactions can affect the environmental temperature, which in turn can affect reaction rates.

## 2.5 Rationale for Sub-symbolic AChems

In real-world chemistry, whether an atom or molecule bonds with another is a result of a complex set of factors based on the states of the reactants, such as the number of electrons and energy states. These are emergent properties that cannot be inferred from the chemical formula (symbolic form) of a molecule.

In order to build a system with such properties we use a new approach, of defining atoms with internal structure. Rather than assigning an atom a symbol and then defining its bonding behaviour purely in the rule system, we define an atom with its own structure and emergent properties. We then define bonding rules in terms of these emergent properties.

We call AChems that display this property of controlling binding based on emergent properties of the molecule *sub-symbolic AChems* (ssAChems) [9, 11]. These ssAChems move beyond treating atoms as structureless symbols as in traditional AChems.

We use the same ( $\mathbb{S}, \mathbb{R}, \mathbb{A}$ ) formalism to define our ssAChems below, but our systems have more of their semantics and behaviour provided in the molecules in  $\mathbb{S}$  (since it emerges from their internal structure), and less in the rules  $\mathbb{R}$ ; behaviour can emerge in a uniform manner across molecules. This makes the definitions in ssAChems more implicit than in symbolic AChems.

### 2.5.1 Terminology

To help prevent confusion between the properties of real chemical molecules and our AChem objects, and to prevent the abuse of chemistry terminology, we use the following distinct terminology:

- The objects of interest are **particles**; these are either atomic particles (**atoms**) or composite particles (**composites**).
- Particles can be joined together, or composed, with **links**; links can be broken to decompose composite particles; no operations within the system can decompose an atom.

## 2.5.2 Requirements and Design Principles

We place several requirements on the kinds of things that can be used as the basis of an ssAChem.

- *structure and/or dynamics*: the particles should have an underlying structure and/or dynamics, from which linking properties can emerge
- *multiple emergent properties*: in a system with a rich set of emergent properties, different properties can be selected to serve a variety of functions
- *single type of particle*: atomic particles can be linked to form composite particles that are the same type of thing as the atoms, with the same kind of emergent properties, so that these can in turn link to other atoms and composites, without the need to introduce new rules
- *everything emergent*: any design decision (for example, linking probability) is based on an emergent property, not on a property of the underlying representation (for example, a property should not rely on using the first item in a list, but instead the item in the list with, say, the maximum or minimum value of an emergent property)
- *computational tractability*: we wish to build computational systems, and so we need the calculation of emergent properties, and of the composition rules, to be tractable.

## 3 Definition of an ssAChem

In this section we define a generic ssAChem, in terms of its  $(\mathcal{S}, \mathbb{R}, \mathcal{A})$ . This provides a detailed framework that can be instantiated with specific ssAChems. This framework is then instantiated with two example ssAChems, one based on Hermitian matrices (Sect. 4), and one based on Random Boolean Networks (Sect. 5).

### 3.1 The Set of Possible Particles $\mathcal{S}$

The particles are defined in terms of the underlying **structure** of particles, their **behavioural model**, and the **emergent properties** of that model. Particles in an ssAChem have their own emergent behaviours that are exploited by the rules. Some of the ‘physics’ of the system has been moved from  $\mathbb{R}$  into  $\mathcal{S}$ .

#### 3.1.1 Structure

$\mathcal{S}$  defines the underlying **structure** of the members of the particle set.

Example structures are:

- Binary trees:  $S := A \mid S \times S$ , so  $S$  is the set of possible particles comprising the atoms  $A$  and all pairwise linked particles  $S \times S$ . This structure is used in the matrix chemistry of Sect. 4.
- General  $n$ -ary trees:  $S := A \mid S^+$ , so  $S$  is the set of possible particles comprising the atoms  $A$  and all chains of linked particles  $S^+$  (we also require the chain length to be  $\geq 2$ ). This structure is used in the RBN chemistry of Sect. 5.
- General graphs:  $S := (E, V)$ , where the vertices are the constituent atoms  $A$ , and the edges are the relevant links. This structure is a model for real world molecules.

Other structures can be defined. We require structures to have a ‘memory’ of the underlying constituent atoms. That is, any structure defined must conserve atomic constituents:

$$A + B \rightleftharpoons C \Rightarrow \text{bag}(A) \oplus \text{bag}(B) = \text{bag}(C) \quad (1)$$

where  $\text{bag}(P)$  denotes the bag (multiset) of atoms in particle  $P$ , and  $\oplus$  denotes bag addition.

### 3.1.2 Behavioural Model

$B$  is the **behavioural model** of the particles, which provides the basis of the chemical properties of interest of the particles.

Each particle structure  $s \in S$  has a behavioural model instance  $b \in B$ , derived via linking rules from the behavioural model instances of its constituent atoms.

Some models are static mathematical constructs (for example, matrices, Sect. 4). Others may be discrete dynamical systems (for example, RBNs, Sect. 5), which can have an associated current **state**,  $\Sigma$ .

### 3.1.3 Emergent Properties

Behavioural models have one or more **emergent properties**:  $e : B \times \Sigma \rightarrow X$ . The ssAChem-specific type  $X$  is typically the real numbers, but it may be, say, a vector, or any other type of interest.

These emergent properties can be used for a variety of purposes: here we use them mainly for linking probabilities.

It is desirable for these emergent properties to be efficiently computable.

### 3.1.4 Atoms and Composites

Each of the **atoms**,  $A \subset S$ , is assigned a unique base atomic model  $b_a \in B$ . Where relevant, each atom instance is initialised to a particular or random initial state,  $\sigma_0 \in \Sigma$ .



Each atom is assigned a unique symbol as its name, enabling us to write down the structure of specific particles. This name is typically an arbitrary letter, or a structured name reflecting some of its underlying emergent properties. The name may be decorated with a tag to indicate the current state of an atom instance.

Each composite is defined by its structure, behavioural model, and current state:  $C = S \times B \times \Sigma$ .

Composites may have an internal composition mirroring their structure  $S$ ; that is, subcomposites corresponding to substructures may also have behavioural models and state. These submodels and substates are related to the overall model and state in a way defined by the linking algorithms (see later).

One possibility we have yet to explore is associating multiple diverse models with a given structure, with  $C = S \times (B \times \Sigma)^N$ . For example, one model might express micro/particle level behaviour, and another macro/ensemble level behaviour. Or one might express linking properties, another functional properties, allowing composites to have some derived behaviour. Or the models might be combined in some way to jointly express a single property.

## 3.2 *The Rules, $\mathbb{R}$*

The rules define the result of linking two particles, and of decomposing a composite particle. Each of these is defined through a precondition (whether the rule applies), and an operation (the result of applying the rule).

### 3.2.1 Ancillary Information

Ideally, all the properties of a particle emerge from its behavioural model. However, in some cases extra information is needed to make deterministic the definition of how two particles link, or how a composite particle decomposes. This ancillary information is provided to operations by the parameter  $\lambda \in L$ .

The ancillary information  $\lambda \in L$  might provide some internal position within the structure of  $C$  where the link is to be established or decomposed; it might provide details of which linking site is to be used. The specific value of  $\lambda$  is provided by the algorithm.

For example, in bRBN-world (Sect. 5), each particle has two linking sites, and the ancillary information of which site is to be used in the linking attempt is provided by the algorithm (a random choice if both are available). This choice is made emergently in the later Spiky-RBN ssAChem [20].

### 3.2.2 Linking Criterion

The **linking criterion**,  $K : \mathbb{P}((C \times L) \times (C \times L))$ , defines whether a pair of particles can in principle link, based on their emergent properties, plus any ancillary information  $\lambda \in L$ .  $K$  defines the domain of applicability, or precondition, of the linking operation  $\sim$ .

Whether particles that can link in principle then link in practice is given by the *linking algorithm*, which depends on further factors.

### 3.2.3 Linking Operation

The **linking operation**,  $\sim : (C \times L)^2 \rightarrow C$ , takes two particles, plus ancillary information, and gives their linked composition. The linking operation can be applied only if the linking criterion holds.

It is desirable that the linking operation to be either non-commutative ( $a \sim b \neq b \sim a$  in general), or non-associative ( $a \sim (b \sim c) \neq (a \sim b) \sim c$  in general), or both [12]. An operation that is both commutative and associative cannot capture *isomers*, composites comprising the same collection of atoms but with different structures and properties. This is because with associativity we can omit the brackets, so  $a \sim (b \sim c) = a \sim b \sim c$ , and then with commutativity we can swap adjacent atoms:  $= a \sim c \sim b = c \sim a \sim b$ . By this means we can get all permutations of  $a, b, c$ ; the linked system is merely an *unstructured* bag of atoms, not a structured composite, and so loses much of the possible combinatoric power.

### 3.2.4 Decomposition Criterion

The **decomposition criterion**,  $K_d : \mathbb{P}(C \times L)$ , defines whether a composite particle can in principle decompose, based on its emergent properties, plus any ancillary information  $\lambda \in L$ .  $K_d$  defines the domain of applicability, or precondition, of the decomposition operation.

Decomposition acts to break links between subcomposite components: it does not cleave atomic particles. Hence no atomic particle is in  $K_d$ .

Whether composite particles that can decompose in principle then decompose in practice is given by the *decomposition algorithm*, which depends on further factors.

### 3.2.5 Decomposition Operation

The **decomposition operation**,  $D : C \times L \rightarrow C^2$ , takes a composite particle, plus ancillary information, and gives its decomposed products resulting from breaking a single link. The decomposition operation can be applied only if the decomposition criterion holds.

There is no requirement for the products to be able to link to form the original composite. Thus decomposition may allow the indirect formation of composites that cannot form by linking alone.

### 3.3 The Algorithm, $\mathbb{A}$

The overall reactor algorithm describes the *dynamics* of the reaction system, including the effect of the environment. Environmental effects may include such things as spatial structure (well-mixed, grid, etc.), inflows and outflows of particles, and (analogues of) energetics. The environmental state can affect the probability of an attempted linking or decomposition succeeding.

#### 3.3.1 Environment

The **environment**,  $\mathcal{E}$ , captures those properties of the system within which reactions happen that are not captured by particle properties alone.

For example, the environment might be aspatial, or a spatial system, allowing varying concentrations and movement of particles; it might have sources and sinks of particles (as in a chemostat); it might have a temperature analogue.

#### 3.3.2 Linking Probability

The **linking probability**,  $Pr_b : \mathcal{E} \rightarrow (C \times L)^2 \rightarrow C \rightarrow \mathfrak{R}$ , defines the probability that, in a given environment, an attempted linking operation  $(c_1, \lambda_1) - (c_2, \lambda_2) = c'$  will be successful.

If the linking criterion is not satisfied, that is, if  $((c_1, \lambda_1), (c_2, \lambda_2)) \notin K$ , then the linking operation is not applicable, and we say that  $Pr_b(e)((c_1, \lambda_1), (c_2, \lambda_2))(c') = 0$ . The linking algorithm can distinguish the case of not linkable in principle (not in  $K$ ) from linkable but with zero probability in this case (in  $K$ , but nevertheless  $Pr_b = 0$ ).

#### 3.3.3 Linking Algorithm

The **linking algorithm** defines how to use the linking rule and linking probability to perform a linking attempt. The behaviours of a given particle structure and rules can be explored in a variety of algorithms.

### 3.3.4 Decomposition Probability

The **decomposition probability**,  $Pr_d : \mathcal{E} \rightarrow C \times L \rightarrow C^2 \rightarrow \mathfrak{R}$ , defines the probability that, in a given environment, an attempted decomposition operation,  $(c, \lambda) \rightarrow (c_1, c_2)$ , will be successful.

### 3.3.5 Decomposition Algorithm

The **decomposition algorithm** defines how to use the decomposition rule and decomposition probability to perform a decomposition attempt.

### 3.3.6 Reactor Algorithm

The **reactor algorithm** defines how to use the linking and decomposition algorithms, and the environment state, to describe the overall behaviour of the reacting system. It defines which particles are in the initial system, which are chosen for linking and decomposition attempts, and if and how particles are added to or removed from the system during execution.

The algorithm for choosing particles can cover all the possibilities: exhaustive pair-wise search, random collisions via a Gillespie-style algorithm [13] for a well-mixed reaction vessel, proximity in a spatial grid of diffusing particles, and more. There might be a fixed number of atoms (conservation of mass), or a chemostat-style inflow/outflow, or other non-physical simulation possibilities. Each of these choices can use the same ‘physics’ of the system (the same  $\mathbb{S}$  and  $\mathbb{R}$ ), merely changing the reaction vessel setup,  $\mathbb{A}$ , to investigate different behaviours and properties of the AChem.

## 3.4 Summary of the ssAChem Framework

1. the set of possible particles  $\mathbb{S}$ , and their underlying properties, defined by
  - a.  $S$ , the underlying **structure** of the particles
  - b.  $B$ , the **behavioural model** of the particles
  - c.  $e$ , the **emergent properties** of  $B$
2. the rules  $\mathbb{R}$ , defining how particles link and decompose, given by
  - a. the **linking rule**, comprising:
    - i. the **linking criterion**, which defines whether a pair of particles can in principle link, based on their emergent properties
    - ii. the **linking operation**, which takes two particles that satisfy the linking criterion, and gives their linked composition

- b. the **decomposition rule**, comprising:
  - i. the **decomposition criterion**, which defines whether a composite particles can in principle decompose, based on its emergent properties
  - ii. the **decomposition operation**, which takes a composite particle that satisfies the decomposition criterion, and gives its decomposition products
3. the algorithm  $\mathbb{A}$ , using the following components to define the overall behaviour of the reacting system:
  - a. the **environment**, defining properties of the system within which reactions occur (e.g., space, temperature)
  - b. the **linking probability** that, in a given environment, an attempted linking operation will succeed
  - c. the **linking algorithm**, which uses the linking rule under the linking probability to perform a linking attempt
  - d. The **decomposition probability**, that, in a given environment, an attempted decomposition operation will succeed
  - e. The **decomposition algorithm**, which uses the decomposition rule under the decomposition probability to perform a decomposition attempt
  - f. The overall **reactor algorithm**, which uses the linking and decomposition algorithms, and defines how particles are chosen for reaction attempts.

## 4 SMAC: Sub-symbolic Matrix Artificial Chemistry

In our first example we illustrate the basics of the ssAChem concepts in a relatively simple system that has no dynamic state, but a lot of rich mathematical structure.

We consider a mathematical system with rich algebraic structure that can be employed to define a ssAChem. The basic particle is the *matrix*. Given the richness of emergent properties of matrices and the fact that most computers are optimised for matrix arithmetic, the matrix seems to be a natural base for a sub-symbolic ACHEM.

There are various so-called matrix ACHEMs in the literature. For example, the Matrix-multiplication chemistry [3–5] works over binary strings and folds them in to matrices to perform linking through matrix multiplication. However, this makes no use of the many emergent properties or algebraic structure of mathematical matrices. In general, existing matrix ACHEMs make little or no use of the mathematical properties of matrices, and would be better named ‘array-based’ ACHEMs.

## 4.1 SMAC's Set of Possible Particles, $\mathcal{S}$

### 4.1.1 SMAC Structure

The structure of SMAC particles is defined as  $\mathcal{S} ::= A \mid \mathcal{S} \sim \mathcal{S}$ , that is, the structure of the set of possible particles comprises the atoms  $A$  and all pairwise linked particles  $\mathcal{S} \sim \mathcal{S}$ . So SMAC particles have an underlying binary tree structure.

### 4.1.2 SMAC Behavioural Model

The SMAC behavioural model  $B$  is the set of  $d$  dimensional Hermitian matrices (here  $d = 3$ ). Particles are static; they have no current state  $\Sigma$ .

#### Matrix Notation

We define a vector  $\mathbf{v}$  of dimension  $d$  in terms of its components  $v_i$ ,  $1 \leq i \leq d$ ; it has magnitude  $|\mathbf{v}| = v$ . We define a matrix  $\mathbf{M}$  of dimension  $d$  in terms of its components  $M_{ij}$ ,  $1 \leq i, j \leq d$ . Vector dot product  $\mathbf{v} \cdot \mathbf{w}$ , matrix addition  $\mathbf{M} + \mathbf{N}$ , matrix-vector multiplication  $\mathbf{M}\mathbf{v}$ , and matrix multiplication  $\mathbf{M}\mathbf{N}$  have their usual definitions.

A Hermitian matrix is one equal to its conjugate transpose:  $\mathbf{M} = \mathbf{M}^\dagger \iff M_{ij} = \bar{M}_{ji}$  where  $M_{ij} \in \mathbb{C}$ .

### 4.1.3 SMAC Emergent Properties

Matrices have several emergent properties, that is, properties of the matrix as a whole, rather than of its individual components. These include:

- Eigenvalues  $\lambda_i$  and eigenvectors  $\mathbf{v}_i$ , solutions of  $\mathbf{M}\mathbf{v} = \lambda\mathbf{v}$ . A  $d$ -dimension matrix in general has  $d$  eigenvalue-eigenvector pairs. Each of the eigenvectors describes a direction, while the corresponding eigenvalue provides a magnitude to that direction. This provides an inherent geometry to the matrix. In general eigenvalues can be complex numbers (they are solutions of the  $d$ -order characteristic polynomial of the matrix). Here we restrict ourselves to Hermitian matrices, which have real eigenvalues.
- Trace:  $Tr(\mathbf{M}) = \sum_{i=1}^d M_{ii}$ . The sum of the eigenvalues equals the trace:  $\sum_{i=1}^d \lambda_i = Tr(\mathbf{M})$ , hence the trace of a Hermitian matrix is real.
- Rank: the dimension spanned by the matrix's rows (or columns).
- Determinant:  $|\mathbf{M}|$
- Similar properties of submatrices of  $\mathbf{M}$

Here we restrict our investigations to  $d = 3$ , both for tractability (the complexity of eigenvalue calculations is  $>O(d^2)$ ), and for a convenient geometric interpretation of the eigenvectors.

The SMAC emergent properties are the absolute values of the three eigenvalues of the behavioural model matrix, truncated to the nearest integer towards zero:

$$e_i := \lfloor |\lambda_i| \rfloor \quad (2)$$

The trace is not necessarily equal to the sum of these truncated values.

We also use the length-normalised eigenvectors  $\hat{v}_i$  in calculating linking probabilities.

#### 4.1.4 SMAC Atoms and Composites

We create an atomic set of  $3 \times 3$  Hermitian matrices. We restrict the entries in our atomic set to have real and imaginary parts of 0 or  $\pm 1$ , and further restrict the leading diagonal elements to be real (because Hermitian).

We disallow the ‘singleton matrices’ (those with a single non-zero entry) and the traceless matrices (which includes the zero matrix), since these have trivial linking properties (later).

We generate our set of atomic behavioural models as:

$$\begin{aligned} B_a = \{ \mathbf{A} \mid & A_{jj} \in \{0, \pm 1\}; \\ & A_{jk} \in \{0, \pm 1, \pm i, \pm 1 \pm i\}, j < k; \\ & A_{jk} = \bar{A}_{kj}, k < j; \\ & \#(A_{jk} \neq 0) > 1; \\ & \text{Tr}(\mathbf{A}) \neq 0 \} \end{aligned} \quad (3)$$

This gives a set of  $3^9 - 6 - 7 \times 3^6 = 14574$  ‘atoms’, which is rather too large to be investigated fully. So we partition this set into 66 equivalence classes of atoms, where atoms in the same class have the same integer values of their eigenvalues rounded to zero, and same traces. We take a single arbitrarily chosen<sup>1</sup> entry from each class. This provides us with a ‘periodic table’ of 66 elements, classified by eigenvalue set and by trace (Fig. 1).

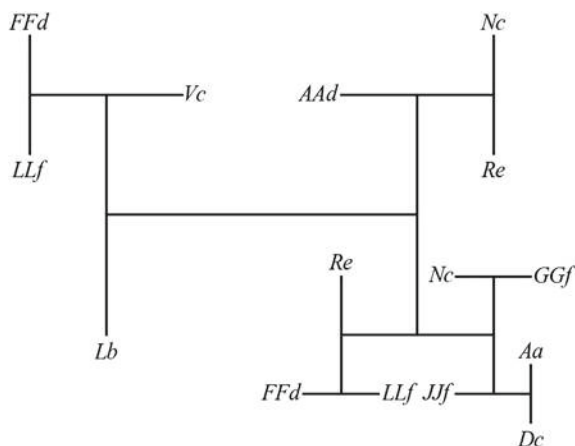
The textual representation of large composites is not very readable. For example, one of the composites generated by our system (see later) is:

$$\begin{aligned} M1000 = & (((((FFd \sim LLf) \sim Vc) \sim Lb) \sim ((AAd \sim (Nc \sim Re)) \\ & \sim ((Re \sim (FFd \sim LLf)) \sim ((Nc \sim GGf) \sim (JJf \sim (Aa \sim Dc))))))) \end{aligned} \quad (4)$$

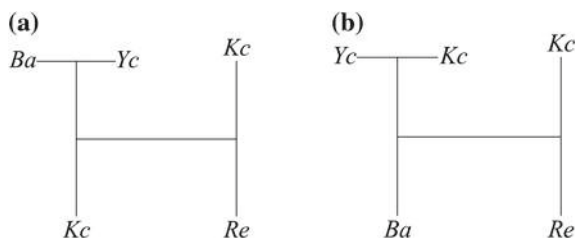
<sup>1</sup>To follow our design criteria, we should choose a non-arbitrary element. Future work includes developing a choice criterion for this.







**Fig. 2** Graph representation of *M1000*. This demonstrates a property of SMAC that differs from real chemistry: links can form between links, as well as between atoms

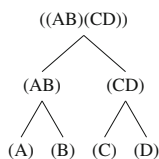


**Fig. 3** SMAC graphs showing the atomic structure of two composites containing the same atoms  
**a**  $((Ba \sim Yc) \sim Kc) \sim (Kc \sim Re)$  **b**  $((Yc \sim Kc) \sim Ba) \sim (Kc \sim Re)$

This is difficult to unpick, so we introduce a graph representation of composites. The graph of the above composite is shown in Fig. 2.

Figure 3 shows two isomers (comprising the same atoms, but linked in different structures) in the graphical representation.

Since the behavioural model has no state, a composite is completely defined by its binary tree structure and its matrix;  $C = S \times B$ . The matrix at the root of the structure is the only one involved in linking; sub-structures have their own associated component matrices which need to be remembered, however, as they may become the root matrix as a result of decomposition. See Fig. 4.



**Fig. 4** The underlying structure of the SMAC compound  $((A-B)-(C-D))$ . The structure  $S$  is given by the binary tree with  $A, B, C, D$  as the atomic leaves. The behavioural model at each sub-tree is given by the relevant matrix product defined in Sect. 4.2.2. The matrix at the root of the structure is the only one involved in the linking algorithm. Sub-structure matrices may become the root of decomposed composites

## 4.2 SMAC Rules

### 4.2.1 SMAC Linking Criterion

The SMAC emergent properties used in linking are the set of  $d$  (truncated) eigenvalue and eigenvector pairs. A SMAC particle can be considered to have  $d$  potential binding sites, labelled  $1, \dots, d$ . For SMAC, ancillary information is provided by the linking criterion, which determines which (if any) specific site of these  $d$  is linkable in each particle:  $L = 1, \dots, d$ . Here  $d = 3$ .

Consider two particles  $A$  and  $B$ , with corresponding emergent properties of sets of truncated eigenvectors  $\{b_i^A\}$  and  $\{b_j^B\}$ . For the particles to link, we require them to have

- a non-zero eigenvalue in common:  $\{b_i^A\} \cap \{b_j^B\} \notin \{\emptyset, \{0\}\}$
- a non-zero trace of their Jordan product,  $Tr(A \circ B) \neq 0$  (in order to apply trace scaling; see Sect. 4.2.2).

### 4.2.2 SMAC Linking Operation

For SMAC, no ancillary information is needed for the linking operation, so we have  $\sim: C^2 \rightarrow C$ .

We impose the following algebraic requirements on our linking operation as applied to our behavioural model of Hermitian matrices:

- It should preserve the property of being Hermitian:  $\mathbf{M} = \mathbf{M}^\dagger$  and  $\mathbf{N} = \mathbf{N}^\dagger \implies \mathbf{M} \sim \mathbf{N} = (\mathbf{M} \sim \mathbf{N})^\dagger$ .
- It should be commutative:  $\mathbf{M} \sim \mathbf{N} = \mathbf{N} \sim \mathbf{M}$ . This is inspired by real chemistry; for example, carbon monoxide could be written as  $C-O$  or  $O-C$ , but is still the same molecule. Note that matrix addition is commutative and matrix multiplication is not commutative.
- It should therefore be non-associative:  $\mathbf{M} \sim (\mathbf{N} \sim \mathbf{P}) \neq (\mathbf{M} \sim \mathbf{N}) \sim \mathbf{P}$  in general (see Sect. 3.2.3). Note that both matrix addition and matrix multiplication are associative.

To meet these requirements, we use a *Jordan algebra* [21] based on Hermitian matrices, and define the linking operation using the Jordan product:

$$\mathbf{M} \circ \mathbf{N} = \frac{1}{2} (\mathbf{MN} + \mathbf{NM}) \quad (5)$$

This is clearly commutative (since matrix addition is commutative). A little algebra demonstrates that it is non-associative, and that the product preserves the Hermitian property.

The Jordan product itself is not suitable for a linking operation that uses eigenvalue matching as its linking property. The eigenvalues of Jordan product matrices tend to grow exponentially fast with the number of Jordan products used. This would result in composites predominantly linking only with composites containing similar numbers of atoms: we would not, for example, be able to see a single atom linking to a large composite.

So our linking operation incorporates trace-scaling, which exhibits matching eigenvalues over a wide range of particle sizes.

$$A \sim B := \frac{|Tr(A)| + |Tr(B)|}{|Tr(A \circ B)|} A \circ B \quad (6)$$

Consequently, we require composites to have a non-zero trace.

### 4.2.3 SMAC Decomposition Criterion

Currently, we have not defined a SMAC decomposition criterion. We are investigating the products of decomposition given that a link breaks, but not yet the criterion for allowing that link breaking.

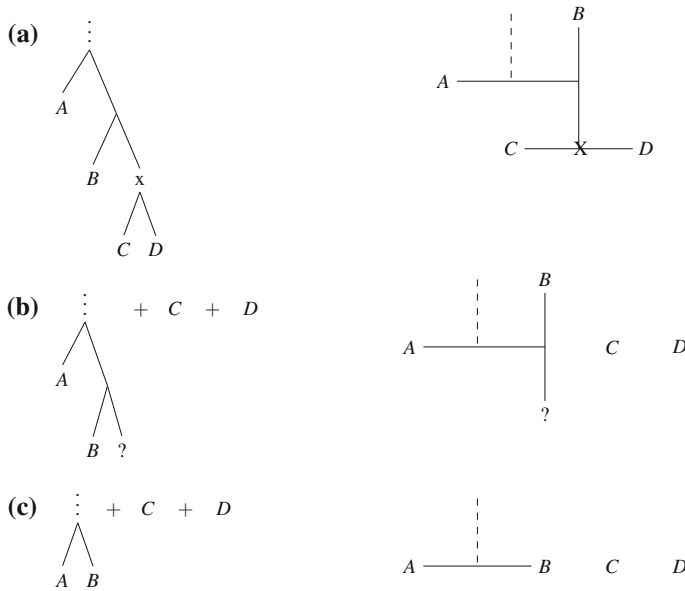
### 4.2.4 SMAC Decomposition Operation

Although the SMAC structure is a binary tree, we wish to have a decomposition operation that allows ‘internal’ links to break. This potentially allows the formation of structures that cannot form directly.

The simplest case is breaking an outermost link.  $A \sim B \rightarrow A + B$ . Both  $A$  and  $B$  may be complex structures.

Breaking a link one deep in the tree is also simple to define:  $A \sim (B \sim C) \rightarrow A + B + C$ .

When we break a link two deep, we need to ‘fix up’ the resulting mal-formed binary tree:  $A \sim (B \sim (C \sim D)) \rightarrow A \sim B + C + D$ . This forms composite  $A \sim B$  indirectly, potentially allowing new composites that could not form directly, or have only a low probability of forming directly.



**Fig. 5** Breaking the link between  $C$  and  $D$  in the composite subcomponent  $A-(B-(C-D))$ . **a** the link to be broken (left shows the binary tree representation; right shows the SMAC notation); **b** the link broken, and the resulting components:  $B$  has no partner; **c**  $B$  migrates up the tree, resulting in  $A-B$ ; in the SMAC notation, we refer to this as ‘link straightening’

This break and fix-up pattern also holds for more deeply nested links:  $A-(B-(C-(D-E))) \rightarrow A-(B-C) + D + E$ . The underlying operation is illustrated in Fig. 5.

### 4.3 SMAC Algorithm

#### 4.3.1 SMAC Environment

The version of SMAC described here has no environmental input.

#### 4.3.2 SMAC Linking Probability

Consider two particles  $A$  and  $B$ , with corresponding sets of truncated eigenvalues  $\{b_k^A\}$  and  $\{b_k^B\}$ , and length-normalised eigenvectors  $\{\hat{v}_k^A\}$  and  $\{\hat{v}_k^B\}$ .

If the linking criterion holds, then we have a matching eigenvalue  $b_i^A = b_j^B = b$ . Let the eigenvectors associated with this matching eigenvalue be  $\hat{v}_i^A$  and  $\hat{v}_j^B$ .

We construct a linking probability from their dot product, such that anti-parallel eigenvectors have the highest linking probability of 1, and parallel eigenvectors have the lowest linking probability of 0:

$$Pr(A \sim B) = \frac{1}{2} (1 - \hat{\mathbf{v}}_i^A \cdot \hat{\mathbf{v}}_i^B) \quad (7)$$

If there is more than one pair of matching eigenvalues, we choose the pair whose eigenvectors are most nearly anti-parallel ( $\min\{\hat{\mathbf{v}}_i^A \cdot \hat{\mathbf{v}}_i^B\}$ ), and so produce the highest linking probability.

This choice of parallel eigenvectors resulting in zero linking probability ensures that not every atom can undergo self synthesis. In order to get self-synthesis we must have a repeated eigenvalue.

### 4.3.3 SMAC Linking Algorithm

For particles to be able to link, they must have non-zero eigenvalues in common, and the trace of the resulting composite must be non-zero. If these criteria are met, then linking can occur on this attempt with probability as given by Eq. 7.

### 4.3.4 SMAC Decomposition Probability

Currently, we have not specified a decomposition probability.

Our decomposition operation can produce composites that cannot form directly by synthesis, because certain links do not have matching eigenvalues. This implies the decomposition probability needs to be different from the linking probability, to ensure such composites do not immediately disintegrate.

### 4.3.5 SMAC Decomposition Algorithm

Currently, we have not specified a decomposition algorithm.

### 4.3.6 SMAC Reactor Algorithm

Here we are in the early stages of investigation, and use a simple reactor algorithm involving linking only, and no decomposition. We generate 10,000 composites using Algorithm 1. We then analyse the resulting composites (see Sect. 4.5).

---

**Algorithm 1** SMAC 10,000 composite synthesis
 

---

```

1: mols := list of the 66 atoms
2: tries := 0
3: repeat
4:    $A, B \in$  mols
5:    $C := A \sim B$ 
6:   if  $C$  then
7:     append  $C$  to mols list
8:   end if
9:   ++tries
10: until # mols = 10,066

```

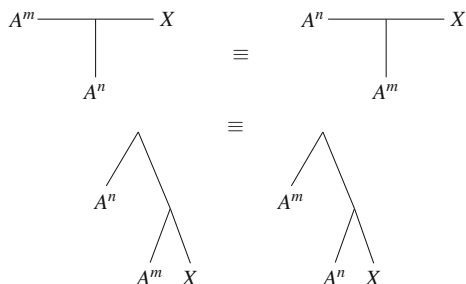
---

## 4.4 Summary of the SMAC ssAChem

### 4.4.1 Definition

- structure  $S$ 
  - structure: binary tree:  $S ::= A \mid S \sim S$
  - behavioural model:  $3 \times 3$  Hermitian matrices,  $Tr \neq 0$ ; no state
  - emergent properties: eigenvalues and eigenvectors; trace
  - atoms: 66 specific matrices
- linking rules  $R$ 
  - linking criterion: matching (truncated, absolute-value) eigenvalues and  $Tr(A \circ B) \neq 0$
  - linking operation:  $A \sim B = \frac{|Tr(A)| + |Tr(B)|}{|Tr(A \circ B)|} (A \circ B)$ ;  $A \circ B = \frac{1}{2} (AB + BA)$
  - decomposition criterion: not defined
  - decomposition operation: ‘link straightening’, see Fig. 5
- algorithm  $A$ 
  - environment: none, for now
  - linking probability:  $\frac{1}{2} (1 - \hat{v}_i^A \cdot \hat{v}_i^B)$
  - linking algorithm: the linking operation is applied with the linking probability
  - decomposition probability: not defined
  - decomposition algorithm: not defined
  - reactor algorithm: see Algorithm 1; used for our initial investigations

**Fig. 6** The isomers  $(A^m \sim X) \sim A^n$  and  $A^m \sim (X \sim A^n)$ , where  $X$  is an atom or a composite, have identical properties in this ssAChem



#### 4.4.2 Structure $\nu$ Model

The commutative but non-associative linking operation used in SMAC allows isomers exist (Fig. 3). However, the linking operation is associative in some special cases. For example, it is clear from Eq. (5) that  $A \circ A = AA = A^2$ , and that therefore

$$\text{Power associative: } A^{\circ m} A^{\circ n} = A^{\circ(m+n)} = A^{m+n} \quad \forall m, n \geq 0 \quad (8)$$

where  $A^{\circ n} = A \circ A \circ \dots \circ A$  ( $n$  times)  $= A^n$ . Hence all composites containing  $n$  copies of just a single atom type  $A$  have identical properties in this ssAChem.

It can readily be checked that the following associativity condition also holds:

$$\text{Jordan identity: } (A^m \circ X) \circ A^n = A^m \circ (X \circ A^n) \quad \forall m, n \geq 0 \quad (9)$$

Hence the two isomers (different structures) in Fig. 6 have identical properties (model values) in this ssAChem.

This demonstrates the difference between structure and properties. Two particles with the same properties can nevertheless be different particles, in that they have a different internal structure. In particular, they may decompose differently.

### 4.5 SMAC Results

We have performed a variety of experiments to investigate SMAC's low-level chemical properties of synthesis and decomposition, to evaluate its promise as the basis for a sub-symbolic AChem. We summarise the results here.

#### 4.5.1 Synthesis and Self-synthesis

The linking probability is zero for parallel eigenvectors, so in order to get self-synthesis we must have a repeated eigenvalue. In such cases there will be (at least

two) eigenvectors orthogonal to one another (because Hermitian), and hence  $\cos \theta = 0$ ,  $p = 0.5$ . Hence self-synthesis probabilities are either be 0 (no repeated eigenvalues: 10 atom classes) or 0.5 (repeated eigenvalues, using the orthogonal eigenvectors: 28 atom classes). Hence we do have atoms capable of self-synthesis, and atoms not capable of self-synthesis.

The probability distribution for general atomic synthesis across all possible atom pairs shows a broad range of probabilities from zero to one.

To investigate the synthesis properties of composites we use the reactor Algorithm 1 to generate 10,000 composites. This process generates a wide range of composite sizes, including example composite  $M1000$  (Eq. 4, Fig. 2), and large composites comprising over 200 atoms.

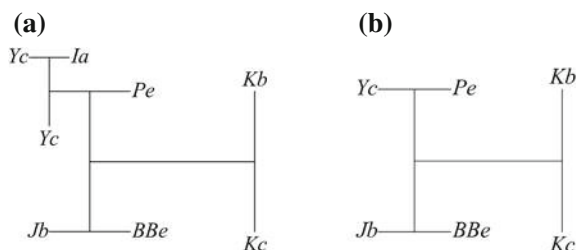
The composite generator took 119,757 tries at forming links to generate 10,000 composites. This implies an 8.35% probability of successful linking in this setup. The probability distribution for this form of composite synthesis also shows a broad range of linking probabilities from zero to one. Additionally, there does not seem to be a strong effect of trace size (and hence composite size) on these linking probabilities. This indicates that large composites can still link effectively.

We observe that large and small composites react with each other, as do similarly sized composites. For any particular size of composite  $s$  the set of reactions creating composites of that size have reactants ranging in size evenly from 1 to  $s - 1$ . Thus there is no bias to linking either similarly or differently sized composites.

## 4.5.2 Decomposition

We have examined the larger composites produced in the synthesis experiment, and have found several cases where the decomposition operation could form composites that cannot be created directly (because of no matching eigenvalues), or only with a very low probability (because of near-parallel eigenvectors).

A particular composite formation by decomposition is shown in Fig. 7; the product composite cannot be formed by synthesis alone.



**Fig. 7** **a** Initial composite before ( $Yc-Ia$ ) link decomposed; **b** The two atoms  $Yc$  and  $Ia$ , plus the composite shown, result from link breaking and link straightening. The new composite cannot form directly because ( $Yc-Pe$ ) and ( $Jb-BBe$ ) do not have matching eigenvalues



## 4.6 *SMAC History Versus Presentation*

The definition of SMAC here following the ssAChem framework starts with matrices as particles, then moves to Jordan products for linking rules. The invention of SMAC followed a different route [12]: the requirement for non-associativity led to the investigation of Jordan algebras, and then the Hermitian matrices as a suitably rich model of these algebras. The overall aim of the SMAC work is to use mathematical structures (starting with the algebraic structure of the particles and their links) as a rich source of inspiration and results for exploitation by ssAChems.

## 5 The bRBN-World ssAChem

Our second example to illustrate the ssAChem concepts exploits a different approach from the algebraic matrix structures in SMAC. Here, we use Random Boolean Networks as our particles and network manipulations as our linking operation: these systems have rich dynamical structure but less overt mathematical foundations.

This work is based mainly on the bRBN models discussed in [9–11], augmented with a first investigation of a temperature model. Current work is exploring more emergent linking sites, to produce ‘spiky RBNs’ [20].

### 5.1 *bRBN-World Particles*

#### 5.1.1 bRBN-World Structure

We define this structure inductively as  $S ::= A \mid S^+$ , with  $n > 1$ . That is, the set of possible particles comprises the atoms  $A$  and all (finite) lists of linked particles  $S \wedge \dots \wedge S$  ( $n > 1$  terms). This results in bRBN-world particles having an underlying  $n$ -ary tree structure.

#### 5.1.2 bRBN-World Behavioural Model

The behavioural model  $B$  is the set of  $K = 2$  ‘linking’ Random Boolean Networks (bRBNs).

#### Random Boolean Network Definition

A Random Boolean Network (RBN) is a discrete dynamical system with the following structure. It has  $N$  nodes. Each node  $i$  has:

- a binary valued state, which at time  $t$  is  $s_{i,t} \in \mathbb{B}$
- $K$  inputs, assigned randomly from  $K$  of the  $N$  nodes (including possibly itself), defining its neighbourhood, a  $K$ -tuple of node labels with no duplicates
- a randomly assigned state transition rule, a boolean function from its neighbourhood state to its own next state:  $\phi_i \in \mathbb{B}^K \rightarrow \mathbb{B}$

The state of node  $i$ 's neighbourhood at time  $t$  is  $\chi_{i,t} \in \mathbb{B}^K$ , a  $K$ -tuple of the binary states of the neighbourhood nodes. At each timestep, the state of all the nodes updates in parallel:  $s_{i,t+1} = \phi_i(\chi_{i,t})$ .

As with any finite discrete dynamical system, the state of an RBN eventually falls on an attractor: a repeating cycle of states. Kauffman [18, 19] discovered that  $K = 2$  RBNs rapidly converge to relatively few relatively short attractors: they are complex, but not chaotic.

### bRBNs: RBNs Modified for Linking

$K = 2$  RBNs are a computationally tractable system with rich microdynamics (of the entire microstate) and complex macrodynamics (the attractor space). We take such  $K = 2$  RBNs as the basis for an ssAChem.

bRBN-world [9–11] uses a modification of basic RBNs, called bRBNs, as the behavioural model. A bRBN atom is constructed from a plain RBN as follows: add  $b$  linking nodes to the RBN (for this work,  $b = 2$ ). For each linking node, select an RBN node at random, and change one of its  $K$  inputs to come from the linking node. Linking nodes do not have inputs, so have no need for an associated random boolean function; instead they have a fixed boolean state (0 or 'cleared' if unlinked, 1 or 'set' if linked; see later for linking details). See Fig. 8.

The linking algorithm can access the microstate of the RBN (to calculate the emergent linking property), each of the  $b$  linking nodes and their state, and the 'wiring' between nodes (to perform the link).

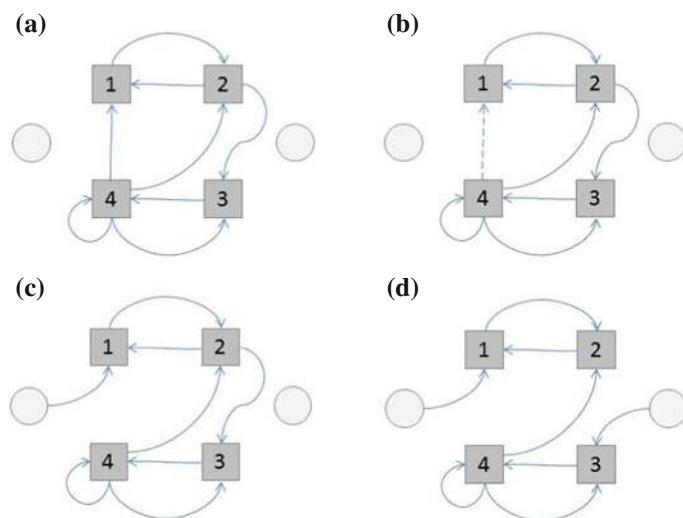
A given particle can exist in different states (the microstate of the underlying RBN at time  $t$ ), dependent on its initial condition.

### 5.1.3 bRBN-World Emergent Properties

An RBN is a discrete dynamical system which exhibits many rich emergent properties. We can look at properties at different levels of resolution. There are properties that merely count the number of timesteps needed to traverse between two states:

- transient length  $e_t$ : the number of timesteps to move from some initial state (for example, all zeros or all ones) to an attractor
- attractor cycle length  $e_c$

Then there are detailed properties of the microstates during these macro-transitions, such as:



**Fig. 8** Constructing a bRBN: **a** a  $K = 2, N = 4$  RBN, plus  $b = 2$  unattached linking nodes (circles); **b** connecting the first bnode: randomly select node 1; randomly select south input (dashed); **c** connecting the first bnode: replace south input with input from b1; **d** connecting the second bnode: randomly select node 3; randomly select north input; replace with input from b2

- flashing: how many nodes change state during the macro-transition
- flashes: total number of state changes during the macro-transition
- total: the sum of the state values during the macro-transition
- proportion: the proportion of nodes that are ‘on’, averaged over the macro-transition

We have investigated each of these properties for attractor cycles to determine their suitability as emergent properties in bRBN-world [9, 10]. The suitable choices, which also depend on choice of linking criterion, are discussed in Sect. 5.5.1.

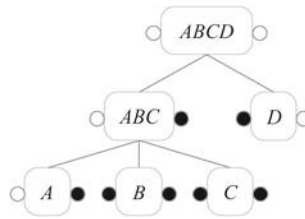
#### 5.1.4 bRBN-World Particles

The atoms are chosen from the set of size  $N$  RBNs. We discuss the choice of  $N$  in Sect. 5.5.1.

Initially [10, 11] we chose small atomic RBNs at random. Later investigations [9] use an evolutionary algorithm to search for an ‘interesting’ set of atomic bRBNs; see Sect. 5.5.2.

We use arbitrary alphabetical symbols to name the atoms. We use parenthesised strings of atoms to display the composites. We distinguish the different basin of attraction, where necessary, with a superscript digit. For example (adapted from [9, Fig. 8.5.3]):

$$((((C^1 \text{~} C^1)^2 \text{~} (C^1 \text{~} C^1)^1)^3 \text{~} (C^1 \text{~} C^6)^4)^1 \text{~} B^3)^1 \quad (10)$$



**Fig. 9** The underlying structure of the RBN-world compound  $((A\sim B\sim C)\sim D)$ . The structure  $S$  is given by the  $n$ -ary tree with  $A, B, C, D$  as the atomic leaves. The behavioural model at each subtree is given by the relevant linked RBNs (Sect. 5.2.2 and Fig. 10). The bRBNs at each level of the structure are involved in the linking algorithm. A white circle denotes a linking site not yet linked, with value set to 0; a black circle denotes a linking site that has been linked (in a parent node), with value here set to 1

An RBN-world particle is defined by its  $n$ -ary tree structure, its bRBN, and its current state;  $C = S \times B \times \Sigma$ . At the root of each substructure is a single composite bRBN; the intermediate nodes contains smaller composite bRBNs; the leaves contain the atomic bRBNs (Fig. 9). The properties of the bRBNs at each level of the structure are involved in the linking algorithm. The fact that properties of lower level structures are used in the linking algorithm stops the linking operation from being associative:  $((A\sim B\sim C)\sim D) \neq ((A\sim B)\sim(C\sim D)) \neq (A\sim B\sim C\sim D)$  in general.

## 5.2 bRBN-World Rules

### 5.2.1 bRBN-World Linking Criterion

An RBN is a discrete dynamical system which exhibits a rich possible set of emergent properties that we can use for defining linking criteria (Sect. 5.1.3).

These numerical properties can be compared in a variety of ways. We have investigated [9, 10]:

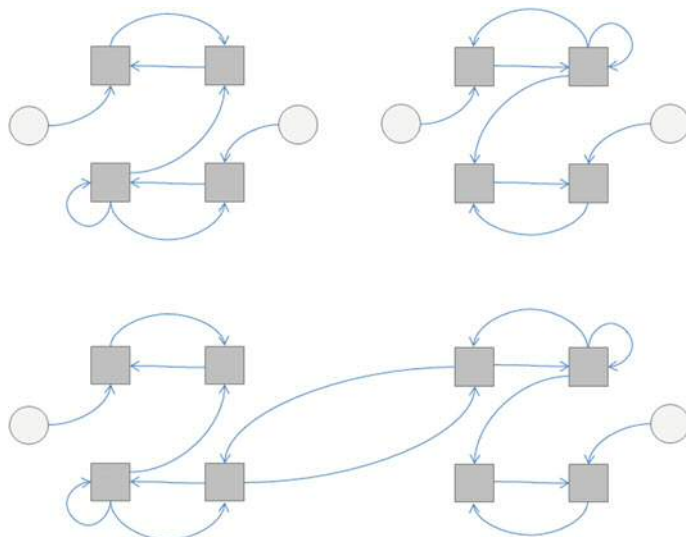
- equal: the same values (to within a small tolerance)
- similar: like equal, but with a larger tolerance bound
- different: not similar
- sum one: the properties sum to one (to within a small tolerance)
- sum zero: the properties sum to zero (to within a small tolerance), hence are opposites

The suitable choices, which also depend on choice of emergent property, are discussed in Sect. 5.5.1.

## 5.2.2 bRBN-World Linking Operation

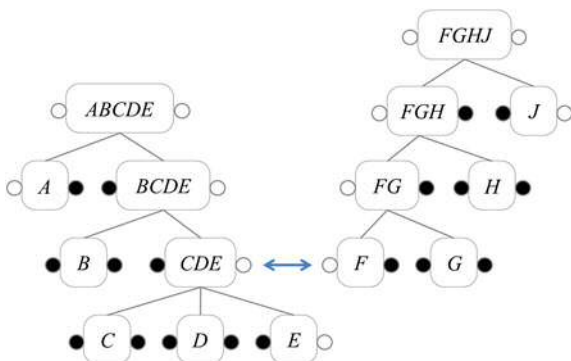
Two individual bRBNs are linked into a larger composite as shown in Fig. 10. Note that although the result is a bRBN, its RBN component does not have the structure of a ‘typical’ RBN: the communication is channeled through specific links, and the result is not a small world network. In particular, large composites comprising long chains of RBNs will have different dynamics from a typical RBN.

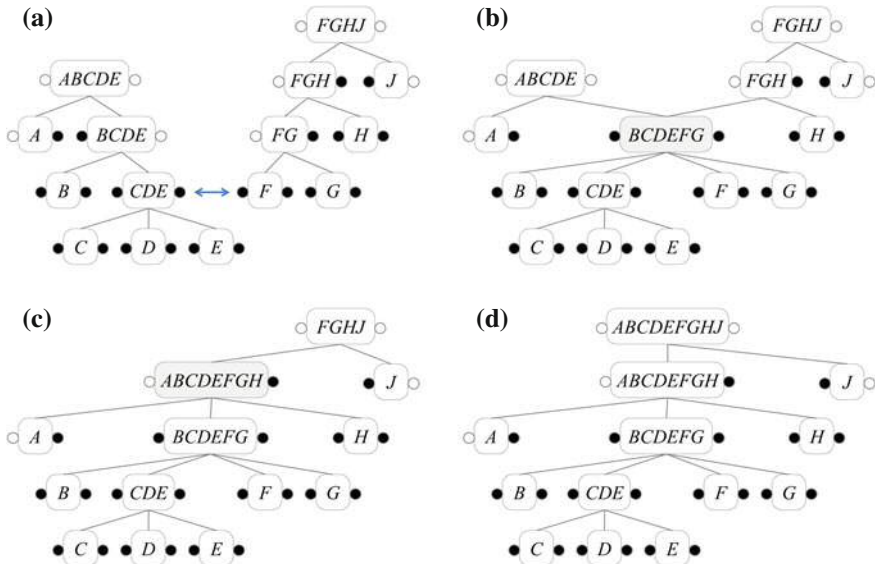
Two structured compounds are linked, at a given linking site in each, as illustrated in Figs. 11 and 12. The link is formed, and the structures are ‘zipped’ together moving up the trees, to form a single tree.



**Fig. 10** Linking two bRBNs. The unlinked bRBNs are stored in the lower nodes of the structure tree; the linked composite bRBN is stored in the higher level of the tree

**Fig. 11** Two composites,  $(A(B(CDE)))$  and  $((FG)H)J$ , to be linked at the linking sites indicated by the arrow





**Fig. 12** The linking operation illustrated: linking  $(A(B(CDE)))$  and  $((((FG)H)J)$  to produce the linked compound  $((A(B(CDE)FG)H)J)$ : **a** set linking sites here and below to ‘linked’ black dots; **b** merge parent nodes; create bRBN that is the join of all the new node’s children (shown in grey); move up to the newly merged parent; **c** while the node has two parents, merge the parent nodes; create the newly merged node’s bRBN from its children; move up to the newly merged parent; **d** while the node has one parent, create the node’s bRBN from its children; move up to the parent

Ancillary information used for linking is:  $S$ , the position in the structure where the link is to be formed;  $b$ , which linking node is to be used;  $\sigma$ , the current state (for the linking property).

### 5.2.3 bRBN-World Decomposition Criterion

The decomposition criterion on a formed link is the same as the linking criterion for forming the link. So, for example in Fig. 10, the criterion for decomposing the top level link uses the two bRBNs  $ABC$  and  $D$ . Although these had to fulfil the linking criterion when the link was formed, they might no longer fulfil it, because now their linked input nodes are set to 1 rather than 0, and so their dynamics have been changed.

### 5.2.4 bRBN-World Decomposition Operation

The decomposition operation is the reverse of the linking operation (Sect. 5.3.3). The designated link is broken, the higher level nodes are ‘unzipped’ to form two separate trees, and the binding site nodes of the lower levels set to 0.

### 5.3 *bRBN-World Algorithm*

#### 5.3.1 Environment

In previously published work on *bRBN-world* [9–11], there are no environmental inputs. In particular, all reactions occur in an aspatial environment.

We report here some preliminary results related to reaction in a constant temperature heat bath.

#### 5.3.2 *bRBN-World Linking Probability*

In the previously published work on *bRBN-world* [9–11],  $Pr = 1$ . That is, if the linking criterion holds, and the link is attempted, then it succeeds with no further criterion to satisfy.

We report here some preliminary results related to reaction probabilities given by a temperature analogue, inspired by reaction kinetics. We use transient lengths as the emergent properties for calculating a temperature-dependent probability. Let  $X_0, X_1$  be the length of the transient from the all-zeros and all-ones state of particle  $X$ . Then consider the linking operation attempting to form composite  $C$  from particles  $A, B$ . We define two energies; the linking energy  $\Delta E_b$  and the decomposition energy  $\Delta E_d$  (Fig. 13):

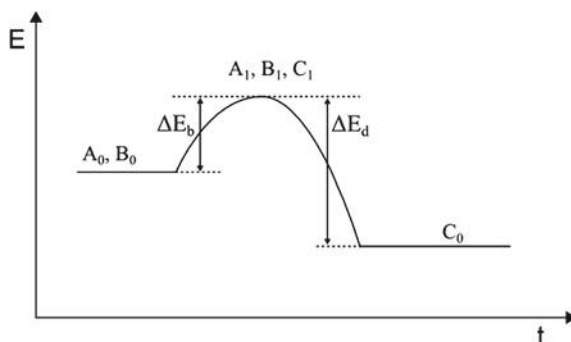
$$\Delta E_b = (A_1 + B_1 + C_1) - (A_0 + B_0) \quad (11)$$

$$\Delta E_d = (A_1 + B_1 + C_1) - C_0 \quad (12)$$

The binding and decomposition probabilities use these energy-analogues in a Boltzmann-like factor, similar to that used in simulated annealing:

$$Pr = \begin{cases} 1 & ; \Delta E < 0 \\ \exp(-\Delta E/T) & ; 0 \leq \Delta E \end{cases} \quad (13)$$

**Fig. 13** ‘Energy’ levels of reactants and linked particles, used to calculate probabilities



The previously published work can be thought of as the infinite temperature limit, with  $Pr = 1$ .

### 5.3.3 bRBN-World Linking Algorithm

Choose one of the linking nodes in each particle, at random.

Get the linking location for these nodes: starting at atomic level, progressively move up the composite tree structure until the linking property holds; use the first linking location found, or FAIL if no linking location is found.

Apply the linking operation (Sect. 5.2.2) at the linking location with probability given by Eq. 13.

### 5.3.4 bRBN-World Decomposition Probability

In the previously published work on bRBN-world [9–11],  $Pr = 1$ . That is, if the linking criterion does not hold, and decomposition is attempted, then it succeeds with no further criterion to satisfy.

We report some preliminary results related to reaction probabilities given by a temperature analogue, inspired by reaction kinetics. This allows otherwise ‘unstable’ links that do not meet the linking criterion to nevertheless persist. See Sect. 5.3.2 for the linking and decomposition probabilities.

### 5.3.5 bRBN-World Decomposition Algorithm

Decomposition proceeds as follows. For a given composite, all the links of its linked component bRBNs are examined. Any link the no longer fulfils the linking criterion (the linking property may have changed on linking) is broken, with probability given by Eq. 13. Any products are similarly decomposed.

### 5.3.6 bRBN-World Reactor Algorithm

The full reactor algorithm is as follows. Two distinct composites are selected; a linking is attempted (Sect. 5.3.3). Decomposition is then attempted (Sect. 5.3.5), either on the initial reactants, or on the successfully linked product. The overall result is the outcome of these two attempts.

A newly-formed composite may decompose because the original linking no longer satisfies the linking criterion (due to changing the state of the relevant linking nodes at the linking location), or because the linking criterion no longer holds elsewhere in the compound. If no link takes place, one or both initial composite particles might themselves decompose, if the linking attempt somehow falsified a linking criterion somewhere in the stateful composite.



## 5.4 Summary of *bRBN-World ssAChem*

- set  $\mathbb{S}$ 
  - structure: n-ary tree:  $S ::= A|S^+, n > 1$
  - model: bRBN with  $b = 2$  linking sites; cycle evolution experiments use  $N = 10$ ,  $K = 2$
  - emergent properties: various transient and attractor cycle properties (Sect. 5.1.3); cycle evolution experiments use ‘proportion’
  - atoms: specific small bRBNs
- rules  $\mathbb{R}$ 
  - linking criterion: matching emergent properties of the bRBNs (Sect. 5.2.1); cycle evolution experiments use ‘sum one’
  - linking operation: form a larger bRBN at the top of the structure tree, zipping together component RBNs lower down the tree
  - decomposition criterion: linking criterion fails to hold for linked structures
  - decomposition operation: break the link, and unzip the trees
- algorithm  $\mathbb{A}$ :
  - environment: aspatial heat bath
  - linking probability: Boltzmann factor from transient emergent properties
  - linking algorithm: link those that pass the linking criterion, with linking probability
  - decomposition probability: Boltzmann factor from transient emergent properties
  - decomposition algorithm: break internal links that fail the linking criterion, with decomposition probability
  - reactor algorithm: link, then decompose.

## 5.5 *bRBN-World Results*

### 5.5.1 Choice of Parameters, Linking Property and Criterion

There are many parameters and properties to choose from: do any result in a good ssAChem?

In our initial work on bRBN-world [11], we arbitrarily chose attractor cycle length as our emergent property, to establish that bRBNs exhibited sufficiently rich behaviours to be the basis for an ssAChem.

In subsequent investigations [9, 10] we determined a better choice, through exploration of the parameter space. That exploration demonstrates that atomic network size  $N$  and connectivity  $K$  has little influence, and that the best choice of emergent

property (Sect. 5.1.3) and linking criterion (Sect. 5.2.1) is either ‘proportion’ as property and ‘sum one’ as criterion, or ‘total’ as property and ‘sum zero’ as criterion. We use ‘proportion/sum one’ for further work on bRBNs.

### 5.5.2 Evolutionary Search for Atomic bRBNs

The space of possible atomic RBNs is vast. The preliminary bRBN-world experiments [10, 11] sample bRBNs at random. To find richer behaviours from small atomic sets, [9] employs a search over the bRBN atomic space, using a genetic algorithm.

The richness of the exhibited behaviour of a candidate atomic set is measure by a fitness function. This function examines the reaction network generated by the atoms for ‘loops’; cyclic reactions such as  $A \rightarrow A(BCD) \rightarrow A(BC) \rightarrow AB \rightarrow A$ . What counts as a ‘loop’ has to be carefully defined to exclude trivial loops such as  $A^1 \rightarrow A^2 \rightarrow A^1$  (where the superscript indicates an atom in a different attractor state), and meaningless loops such as  $A \rightarrow BC \rightarrow A$  (where there is no common element around the loop). See [9, Chap. 8] for details.

A well-mixed reactor vessel is populated with 1000 atoms of each of five types. A reaction network is formed by running the vessel for a sufficient time, using a Gillispie-style algorithm. The fitness function is applied to the resulting network.

The genetic algorithm uses a population size of 100 vessels, and is run for 300 generations. Mutation changes the details of the five types of atoms used. See [9, Chap. 8] for further details.

This evolutionary approach successfully discovered atomic sets capable of producing complex reaction networks. For example, one reaction network analysed in detail exhibited 1,286 reactions and 645 different particles, and a longest reaction loop comprising 8 reactions.

This demonstrates that the bRBN-world ssACem can continue to support more complex behaviours.

### 5.5.3 Addition of a Temperature Analogue

Following on from the work presented in [9–11], we have performed some preliminary experiments with the temperature analogue described in Sect. 5.3.2.

At very low temperatures (‘absolute zero’,  $T = 0$ ) essentially no reactions occur, as the probability is always zero (except when  $\Delta E < 0$ , which happens only rarely).

At very high temperatures ( $T = 15$ ) the system essentially reduces to the previously published bRBN-world behaviours: all reactions that merely satisfy the linking criterion occur, as do all the possible decompositions (no ‘unstable’ links can survive).

At intermediate temperatures ( $2 < T < 10$ ) we have a reduced reaction rate, as expected. We see a mix of behaviours, with different composite species thriving in

different temperature ranges. Additionally, we see new composites that cannot occur at higher temperatures, because they would decompose.

We have also performed preliminary experiments where the temperature of the system changes between three temperatures (low  $T = 2$ , medium  $T = 5$ , high  $T = 10$ ) for five cycles. These systems show a higher degree of reactivity than a similar system held at the medium temperature.

These preliminary experiments demonstrate that a temperature analogue has an interesting effect on the behaviour of the system, by allowing otherwise unstable composites to persist, and engage in further reactions. Having demonstrated the promise of the approach, we need to perform larger scale experiments to quantify the effects.

## 6 ssAChem Design Guidelines

We have provided a framework within which ssAChems can be defined. SMAC and bRBN-world provide two different instantiations of this framework. SMAC focusses on the underlying mathematical structure provided by Jordan algebras and their realisation in the Hermitian matrix model. RBN-world focusses on the underlying dynamical system that provides the atomic properties. These results demonstrate that the ssAChem concept is generically useful. It is not only the original dynamic bRBN-world implementation that demonstrates interesting chemistry: other systems with less dynamics but deeper mathematical structure also demonstrate interesting chemistry.

The process of designing a new ssAChem can proceed as follows:

1. start from the ssAChem framework provided here
2. instantiate it with a particular *model*: an atomic set, the emergent properties, the linking rules
3. develop a computational implementation
4. perform initial experiments, to demonstrate that the chosen model has rich behaviour, and to find good parameter values
5. perform full experiments, to explore full system behaviours

**Acknowledgements** Faulkner is funded by an York Chemistry Department Teaching PhD studentship. Krastev is funded by a York Computer Science Department EPSRC DTA PhD studentship. We thank Leo Caves for some insightful comments on this work. We thank Michael Krotosky, Andrew Balin, and Rudi Mears for their work in exploring some earlier versions of ideas presented here.

## References

1. ALife XV, Cancun, Mexico. MIT Press (2016)
2. Anet, F.A.L.: The place of metabolism in the origin of life. *Curr. Opin. Chem. Biol.* **8**(6), 654–659 (2004)
3. Banzhaf, W.: Self-replicating sequences of binary numbers—Foundations I: General. *Biol. Cybern.* **69**(4), 269–274 (1993)
4. Banzhaf, W.: Self-replicating sequences of binary numbers—Foundations II: Strings of length  $N = 4$ . *Biol. Cybern.* **69**(4), 275–281 (1993)
5. Banzhaf, W.: Self-organization in a system of binary strings. In: *Proceedings of Artificial Life IV*, pp. 109–118 (1994)
6. Banzhaf, W., Baumgaertner, B., Beslon, G., Doursat, R., Foster, J.A., McMullin, B., de Melo, V.V., Miconi, T., Spector, L., Stepney, S., White, R.: Requirements for evolvability in complex systems. *Theory Biosci.* **135**(3), 131–161 (2016)
7. Dittrich, P., Ziegler, J., Banzhaf, W.: Artificial chemistries—a review. *Artif. Life* **7**(3), 225–275 (2001)
8. Eigen, M., Schuster, P.: A principle of natural self-organization. *Naturwissenschaften* **64**(11), 541–565 (1977)
9. Faulconbridge, A.: RBN-world: sub-symbolic artificial chemistry for artificial life. Ph.D. thesis, University of York, UK (2011)
10. Faulconbridge, A., Stepney, S., Miller, J.F., Caves, L.: RBN-world: The hunt for a rich AChem. In: *ALife XII*, Odense, Denmark, pp. 261–268. MIT Press (2010)
11. Faulconbridge, A., Stepney, S., Miller, J.F., Caves, L.S.D.: RBN-World: a sub-symbolic artificial chemistry. In: *ECAL 2009*, Budapest, Hungary. LNCS, vol. 5777, pp. 377–384. Springer (2011)
12. Faulkner, P., Sebal, A., Stepney, S.: Jordan algebra AChems: exploiting mathematical richness for open ended design. In: *ALife XV*, Cancun, Mexico [1], pp. 582–589 (2016)
13. Gillespie, D.T.: Exact stochastic simulation of coupled chemical reactions. *J. Phys. Chem.* **81**(25), 2340–2361 (1977)
14. Hutton, T.J.: Evolvable self-replicating molecules in an artificial chemistry. *Artif. Life* **8**(4), 341–356 (2002)
15. Hutton, T.J.: Evolvable self-reproducing cells in a two-dimensional artificial chemistry. *Artif. Life* **13**(1), 11–30 (2007)
16. Kauffman, S.A.: Metabolic stability and epigenesis in randomly constructed genetic nets. *J. Theoret. Biol.* **22**(3), 437–467 (1969)
17. Kauffman, S.A.: Autocatalytic sets of proteins. *J. Theoret. Biol.* **119**(1), 1–24 (1986)
18. Kauffman, S.A.: Requirements for evolvability in complex systems. *Physica D* **42**, 135–152 (1990)
19. Kauffman, S.A.: *The Origins of Order*. Oxford University Press (1993)
20. Krastev, M., Sebal, A., Stepney, S.: Emergent bonding properties in the Spiky RBN AChem. In: *ALife XV*, Cancun, Mexico [1], pp. 600–607 (2016)
21. McCrimmon, K.: Jordan algebras and their applications. *Bull. Am. Math. Soc.* **84**(4), 612–627 (1978)
22. Ogawa, A.K., Yiqin, W., McMinn, D.L., Liu, J., Schultz, P.G., Romesberg, F.E.: Efforts toward the expansion of the genetic alphabet: information storage and replication with unnatural hydrophobic base pairs. *J. Am. Chem. Soc.* **122**(14), 3274–3287 (2000)
23. Ono, N., Ikegami, T.: Model of self-replicating cell capable of self-maintenance. In: *Advances in artificial life*, pp. 399–406. Springer (1999)
24. Pross, A.: Causation and the origin of life: metabolism or replication first? *Orig. Life Evol. Biosph.* **34**(3), 307–321 (2004)
25. Suzuki, H., Ono, N., Yuta, K.: Several necessary conditions for the evolution of complex forms of life in an artificial environment. *Artif. Life* **9**(2), 153–174 (2003)