

# Emergent Bonding Properties in the Spiky RBN AChem

Mihail Krastev<sup>1,3</sup>, Angelika Sebald<sup>2,3</sup>, Susan Stepney<sup>1,3</sup>

<sup>1</sup>Department of Computer Science, University of York, UK

<sup>2</sup>Department of Chemistry, University of York, UK

<sup>3</sup>York Centre for Complex Systems Analysis  
mk599@york.ac.uk

## Abstract

We present a subsymbolic Artificial Chemistry (ssAChem) in which all properties relevant to bonding are emergent from the underlying dynamical system (an RBN). We explore this ssAChem by evolving a seed set of atomic particles and showing the type of composite particles the system can produce.

## INTRODUCTION

The field of Artificial Chemistry (AChem) has produced numerous models of chemical systems over the years. These models often have a symbolic representation of atoms or molecules, a defined set of possible reactions, and rely on some form of environment such as a 2D lattice or a well mixed reactor (Dittrich et al., 2001).

In these approaches bonding occurs via explicitly defined reaction rules through a grammar-like notation. A reaction results in a new symbol replacing one in the reactor (McMullin, 1997; Varela et al., 1974; Ono and Ikegami, 2001; Madina et al., 2003). These systems are not always mass conserving. In mass-conserving systems bonds form links between two particles (Hutton, 2002, 2005, 2007). In both cases reaction rules are explicitly defined aiming to explore specific behaviours. Another approach (Fontana and Buss, 1994; Dittrich and Banzhaf, 1998; Banzhaf et al., 1999; Hickinbotham et al., 2010) uses one reactant as an operator and another one as an operand with the result being the reaction product. In such systems reactions are emergent from the properties of the particles themselves.

In subsymbolic Artificial Chemistries (ssAChems) (Faulconbridge et al., 2010, 2011; Faulconbridge, 2011) particles are defined as systems with internal properties. Bond formation is a result of these interacting in a predefined way. Reactions are an emergent property of interaction. Overall the subsymbolic system can be seen as being less complicated, since it does not define individual behaviours like most symbolic approaches do. However, because bonding properties are now emergent and the reaction algorithm universal for all possible particles, the system is significantly more expressive with a huge combinatorial search space of possible reactions.

Random Boolean Networks (RBNs) (Kauffman, 1969; Kauffman et al., 2003) are an attractive dynamic system on which to base ssAChems. They provide a large number of exploitable properties, are computationally inexpensive and can easily be combined to produce analogues to molecular structures. RBN-World (Faulconbridge et al., 2011) itself, while having emergent bonding properties, still has many of its properties externally defined. Here we introduce the *Spiky RBN* model, and a simple reaction mechanism which allows for bond formation and decomposition. Like in RBN-World we use an RBN as the subsymbolic system. However, our mapping of subsymbolic to atomic particle properties is significantly different. Our aim is to build an ssAChem where all properties relevant to bonding are fully emergent from the underlying structure and dynamics.

## The Spiky RBN Model

Here we first introduce the Spiky RBN, which defines our atomic particles and their properties. Next we give the collision and stability criteria, which determine if a link can be formed and if a link is stable. Then we explain the linking mechanism that describes what a link consists of and how to form one. To emphasise that we are not simulating real atoms or chemicals, we use the following terminology. Particles are connected by *links*. An *atomic* particle is a particle with no links (in our model it is a single Spiky RBN); no operations within our system can break down an atomic particle. A *composite* particle consists of two or more atomic particles connected via link. References to a *particle* mean either an atomic or a composite particle.

The boolean networks are random since our interest is in emergence of properties and dynamics as oppose to engineering of properties towards specific dynamic behaviour. Engineering specific boolean networks would limit the systems dynamics to only those that we have encoded.

## Atomic Particle

An atomic particle is a small RBN, and its properties emerge from the dynamics. Particles are linked to form larger RBNs, with their own emergent dynamics. In the original RBN-

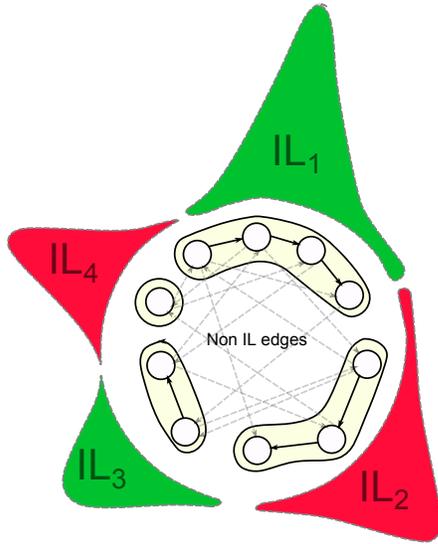


Figure 1: Diagram of the Spiky RBN. Nodes are separated into Interaction Lists (IL) of varying size based on Algorithm 1. Edges in solid lines are those which are part of the IL and can be used in linking. Dashed line edges are part of the RBN topology that cannot be changed due to linking. Note that there is an IL with only one node and no edges; it cannot form a link. Each IL spike can be seen, with colour denoting sign and size denoting magnitude.

world, particles had two special ‘bonding nodes’ added arbitrarily; in our Spiky RBN model, the number, location, and properties of these nodes are emergent.

The core of the model is the subsymbolic representation of the atomic particle. The RBN is split into Interaction Lists (IL) made up of RBN nodes as shown in Fig. 1. Each IL is a list of nodes where each subsequent node takes direct input from the previous node in the list. The first node is the only one with unspecified inputs. ILs are constructed by following connected nodes where the next node is chosen based on the number of its outgoing edges, as defined in Algorithm 1. We attempted a few methods of picking first and subsequent nodes least influential gave a good balance between number of nodes per IL and number of ILs.

The ILs partition the nodes: every node belongs to exactly one IL, and every IL has at least one node in it. ILs have no effect on the topology of the RBN; they are a logical grouping.

ILs form the basis for interaction between particles in the Spiky RBN model, replacing the arbitrary binding site in RBN-World (Faulconbridge et al., 2011). The size of the IL (the number of nodes it contains) and the number of ILs in a particle are all derived purely from the topology of the RBN. Each IL has a numerical property referred to as the *spike* (Fig. 1), which determines if a link will form and remain stable. The spike value is calculated over the attractor cycle

```

Data:  $N$ : list of all nodes in the RBN ordered by least
infectious first
while  $N$  is not empty do
  Remove first node  $n$  from  $N$ ;
  Create new Interaction List  $IL_i$ ;
  Add  $n$  to  $IL_i$ ;
  while  $\exists n' \in N$  where  $n$  is an input to  $n'$  do
    Remove  $n'$  from  $N$ ;
    Add  $n'$  to  $IL_i$ ;
     $n \leftarrow n'$ ;
  end
   $i++$ ;
end

```

Algorithm 1: Building Interaction Lists

as follows.

The value of node  $x$  at RBN state  $s$ ,  $x_{s\text{value}}$  is 1 if it is in a ‘true’ state and  $-1$  if it is in a ‘false’ state.

$$x_{s\text{value}} = \begin{cases} 1, & \text{if } x_{s\text{state}} = T \\ -1, & \text{if } x_{s\text{state}} = F \end{cases} \quad (1)$$

The value of node  $x$  over one cycle of the attractor  $x_{\text{value}}$  is

$$x_{\text{value}} = \sum_{s=t}^{s=t+c} x_{s\text{value}} \quad (2)$$

where  $t$  is the first state of the attractor and  $c$  is the attractor length.

The spike for  $IL_1$  of particle  $A$ ,  $S_{A1}$ , is the sum of all node values for nodes in that IL.

$$S_{A1} = \sum_{x \in IL_1} x_{\text{value}} \quad (3)$$

This gives us a spike with both a magnitude and a sign. It is constrained by the attractor length  $c$  and the number of nodes in  $IL_{A1\text{size}}$ :

$$-IL_{A1\text{size}}c \leq S_{A1} \leq IL_{A1\text{size}}c \quad (4)$$

We calculate the attractor of a RBN from an initial state of all ‘false’.

An atomic particle has three properties: the number of ILs, the size of each IL, and the spike of each IL. The first two are a function of the RBN topology and the third is a function of the RBN dynamics. Because all are deterministically calculated, two identical RBNs will produce two identical atomic particles with identical behaviours. The number of ILs gives the maximum number of links that a particle can form. The size of each IL determines if it can be part of a link and how severely a link will change the topology of the particle. The spike dictates which specific set of bonds is possible.

## Collision and Stability Criteria

The second component of the model are the collision and stability criteria. The collision criterion dictates what must be true in order for a link to form. The stability criterion dictates what must be true in order for a link to continue to exist.

A link can form between ILs chosen from two different particles. The ILs must have size  $>1$  in order for a link to be possible. How particles and ILs are chosen depends on the reactor type. In an aspatial well mixed reactor two random particles can be chosen and a random IL on each. In a 2D lattice reactor particles in adjacent sites and the nearest ILs can be chosen.

The collision criterion states that:

$$S_{iA} + S_{jB} = 0 \quad (5)$$

where  $S_{iA}$  is the spike of the  $i^{th}$  IL of particle  $A$ . If the collision criterion is not met, the collision is considered elastic and the two particles do not form a link. If the collision criterion is met then a link forms as described below. Link formation results in a change in RBN topology and consequently in a possible change to partial linking properties. After the bond construction all partial linking properties are recalculated and used to check against the stability criterion.

Like collision criterion, the stability criterion states that:

$$S'_{iA} + S'_{jB} = 0 \quad (6)$$

where  $S'_{iA}$  is the spike of the  $i^{th}$  IL of particle  $A$  after the bond has been formed. The stability criterion is checked not only for the newly formed link, but for every pair of ILs that are part of a link in the new composite particle. Decomposition results in a particle splitting into two or more fragments. Each IL that was part of a now broken link is free again. This means the topology has changed such that stability criterion is checked recursively until all links in all fragments meet the criterion. Because the stability criterion is checked for all links it holds true for every composite particle that is not currently attempting to link.

## Link Structure and Formation

If the collision criterion holds then a link is formed. This is done by swapping pairs of nodes inputs between the two ILs as shown in Fig. 2. Edges which are part of the IL are swapped, starting with the edge outputting from the first node in the IL.

The maximum number of swaps possible is  $n - 1$ , where  $n$  is the size of the smaller of the two ILs.

A link can be constructed between any two IL of size  $>1$ . ILs of size 1 cannot link because they have no edges to swap. (It is possible to have a node that takes input from itself. We do not consider these.)

After link formation the spikes of all ILs in the new composite particle are recalculated, since link formation results

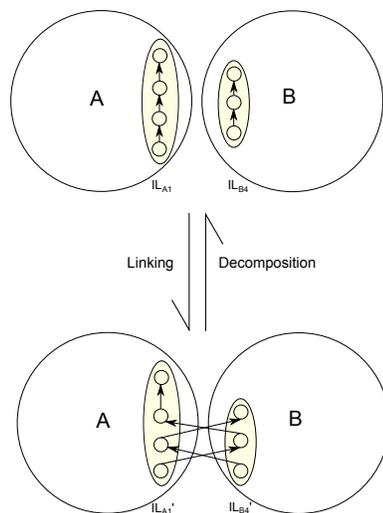


Figure 2: An example of the structure of a link between particles  $A$  and  $B$ . Note that two pairs of edges are swapped since  $IL_{B4}$  has a size of 3.

in a change in the underlying RBN topology. For any links that do not meet the stability criterion the link is decomposed by reversing the input swaps. When a link decomposes the break results in two new particles; again the spikes of the ILs are recalculated, and any further decomposition needed is performed. This process continues until the products are stable (meet the stability criterion), hence a single interaction between two reactants can result in multiple product particles. The algorithm for two atomic particles is shown in fig.3.

Our links have a richer structure than those in the original RBN-World (Faulconbridge et al., 2011). Small ILs mean fewer swaps to form a link, resulting in less perturbation to the linking particles. This implies a higher chance that the spikes do not change and the link is stable. Larger ILs produce a larger change in topology and are therefore more likely to result in different spikes and so bond instability.

## Experiment: Growing a Seed Set

The aim of the experiment is to use an evolutionary approach to generate interesting seed sets of atomic particles. One of the long term goals of the project is to add further parameters such as kinetics, variable link strengths, and geometry to the model, in order to see the effects these properties have on the dynamics of the system. In order to understand and compare the effects of these parameters we need an exemplar dynamic system: a set of atomic particles with nontrivial dynamic properties. Finding such a set can be difficult since the search space is vast and there is no way to predict dynamic behaviour without simulation. We can state undesirable characteristics:

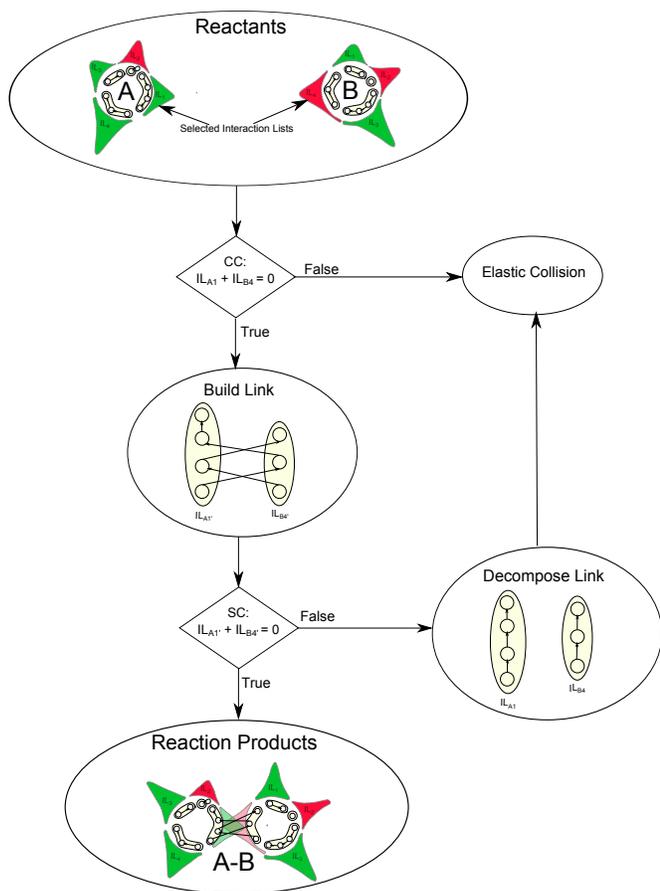


Figure 3: Reaction algorithm between two atomic particles. First an IL is selected on each. The collision criterion is checked. If it passes then a link is formed (product particle  $AB$ ). The stability criterion is checked for the ILs that are part of the link. If it fails then the link decomposes. Note that ILs not involved in links may have had their spikes changed due the effect of the link on the new composite RBN.

- Overly restrictive : most reactions are elastic and do not result in larger composite particles. The system quickly reaches a stable state with no reactions occurring.
- Overly permissive : almost all reactions result in stable links. The system quickly congeals into a single large composite particle.
- Chaotic : almost all reactions between composite particles result in decomposition of links. The system is reactive but larger composites are quickly destroyed.

## Reactor

For this experiment we use an aspatial reactor initialised with 20 unique atomic particles. The reactor attempts 1,000 links by picking two particles at random, picking an IL on each at random, and attempting a link. If the reaction is successful all reactants *and* products are added to the reactor.

At any time the reactor contains one copy of each composite particle that has been generated so far, plus the initial 20 atomic particles. In effect our system is a well stirred reactor with equal concentrations of all particles. This is a rough exploration of the possible behaviour the seed set can produce.

## Fitness Metrics

We use a fitness function to describe the type of system we are looking for. We calculate a fitness per seed set in the population, as well as a fitness per atomic particle. There are five measures on which we base our fitness function:

- $C$  The number of unique composite particles that the system creates after a set number of reaction attempts.
- $V$  Variance in observed composite particle size.
- $L$  Variance in number of links per atomic particle in a composite particle.
- $R$  Percentage of attempted reactions for which the new bond is stable.
- $P$  Number of unique links that an atomic particle has been observed as forming. For example if atomic particle  $A$  has only ever formed bonds with atomic particle  $B$  and itself, then  $P = 2$ .

These characteristics form the basis of our fitness function.  $C$ ,  $V$ ,  $L$  and  $R$  provide an overall fitness for our seed set.  $P$  provides an individual fitness for each atomic particle within the seed set. We use a rank based approach, which removes the need to provide weights for the components of the fitness function.

## Fitness Functions

The fitness of reactor  $i$  is:

$$fr_i = Rank(C_i) + Rank(V_i) + Rank(L_i) + Rank(R_i) \quad (7)$$

where  $Rank(C_i)$  is the rank of reactor  $i$  when the reactors are ordered by lowest to highest. Since our population is made of 20 individuals and there are four ranks,  $fr_i$  is constrained to

$$4 \leq fr_i \leq 80 \quad (8)$$

The fitness of atomic particle  $j$  in reactor  $i$  is the number of unique bonds it can form,  $P_{ij}$

$$f_{ij} = P_{ij} \quad (9)$$

With a seed set of 20 atomic particles  $f_{ij}$  is constrained to

$$0 \leq f_{ij} \leq 20 \quad (10)$$

The mutation function replaces the atomic particle  $j$  in reactor  $i$  with a new random one with a probability proportional to  $M_{ij}$ , where

$$M_{ij} = \frac{84 - fr_i}{1 + f_{ij}} \quad (11)$$

That is, fitter reactors and fitter particles are mutated less.

$$0.19 < M_{ij} \leq 80 \quad (12)$$

## Exploratory Algorithm

To generate atomic particle sets we use an algorithm similar to clonal selection (De Castro and Von Zuben, 2000, 2002) (see Fig. 4). Our population is made of 20 reactors. Unlike normal clonal selection, each population member produces exactly one clone by mutating atomic particles based on  $M_{ij}$ . This is because our aim is not specific optimisation but rather exploration for possible seed sets with favourable behaviours. In order to ensure that good seeds propagate through the generations we include a low 5% crossover chance. The crossover function replaces the three lowest participation particles in a set with the three highest participation particles from another set. When crossover occurs we ensure that the resultant seed set has 20 unique atomic particles by making sure the incoming particles are not already in the seed set. The crossover probability is kept low because again we are interested in diversity. Also, due to the nature of the system, high fitness of a particle in one reactor does not necessarily imply high fitness in another reactor. This is a desirable trait since high fitness in all reactors would suggest the particle is overly permissive and can bond with almost everything.

## Results

The experiment was run with RBNs of  $K = 2, N = 12$  forming atomic particles. We first look at the behaviour of the reactors over the generations and then give an example particle from the best reactor at the end of the run.

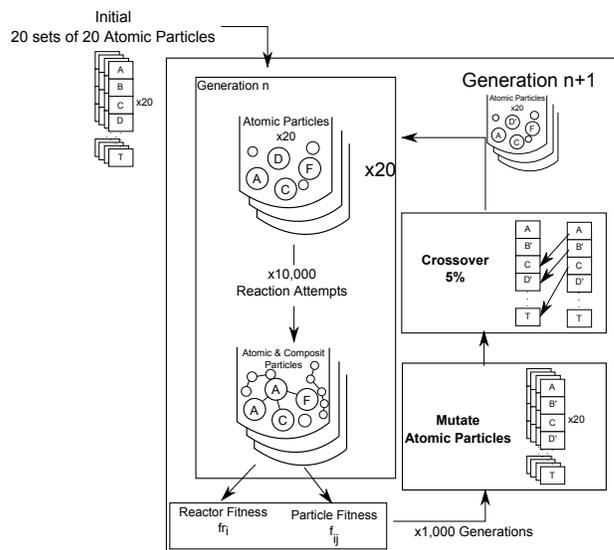


Figure 4: Exploration algorithm. The algorithm is initialized with 20 sets of 20 atomic particles each. Each reactor then attempts 1,000 reactions. We then calculate reactor fitness and particle fitness, mutate and perform crossover to get the new population for gen  $n + 1$ . This repeats for 100 generations.

## Exploratory Algorithm Performance

Fig. 5 shows how the values of  $C$ ,  $V$ ,  $L$  and  $R$  change over the generations.

Over the generations there is an increase in the distribution's upper quartiles suggesting some reactors are improving and finding better seed sets.

The median values in each graph fluctuate, which is to be expected: even if no mutation or crossover is experienced there is no guarantee that a successful reactor will reproduce its behaviour in the following generation. Because reactions are randomly chosen it is possible that a very reactive composite particle is not generated even if one is possible.

The median variance in size (Fig. 5b) stays very low for most of the experiment. This is due to reactors producing only composite particles of one size (most commonly size 2) giving a variance of 0.

Generation 68 shows a large increase in variance in particle size (Fig. 5b) compared to the previous generation. For that generation there is also an increase in median number of unique particles (Fig. 5a), and number of bonds formed (Fig. 5d) compared to the previous generation. The median variance in number of bonds per particle (5c) is lower than the previous generation however. This suggests that gen 68 produced many large composites which were mostly straight ribbons of particles with low branching.

The large number of outliers shows that while these low-reactivity systems are common we can also find more interesting examples. The reduction in outlier numbers towards

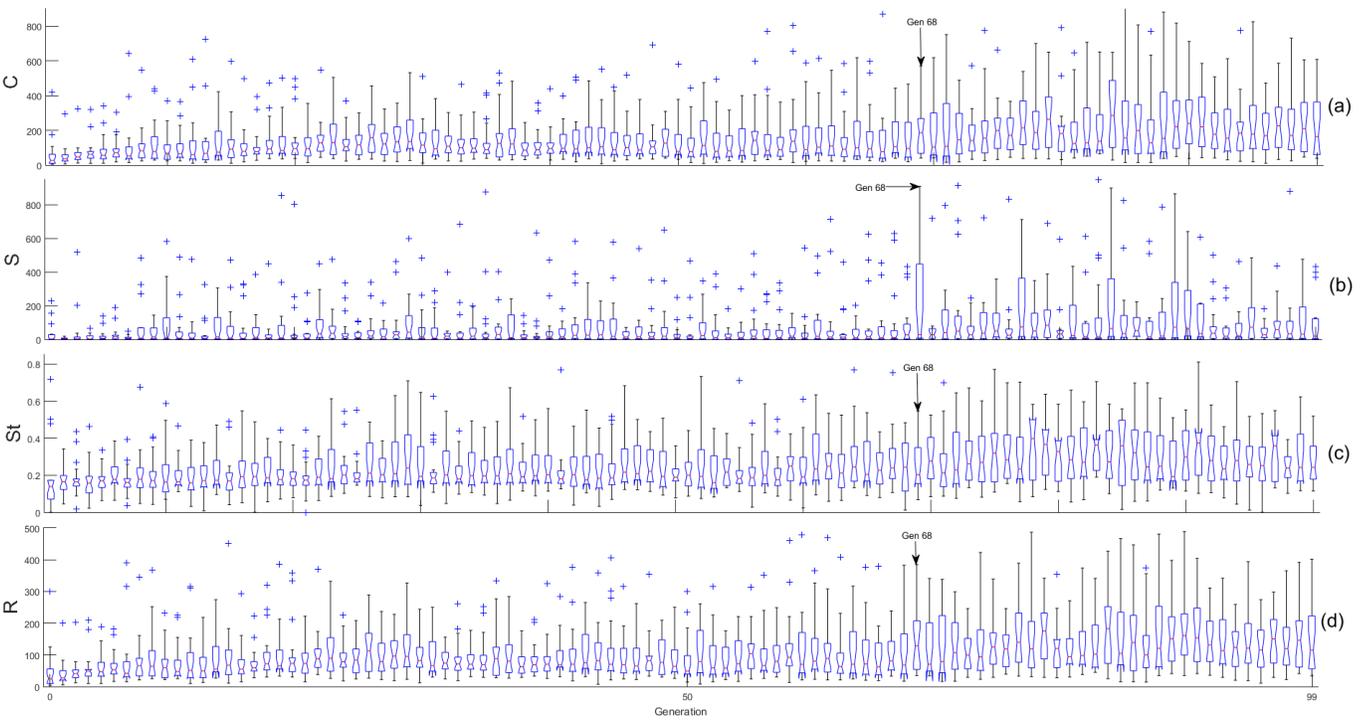


Figure 5: Distribution of the reactor measures per generation; some outliers are omitted. (a)  $C$ , number of unique particles; (b)  $V$ , variance of particle size; (c)  $L$ , variance of number of links; (d)  $R$ , stability. Generation 68 is highlighted for reference.

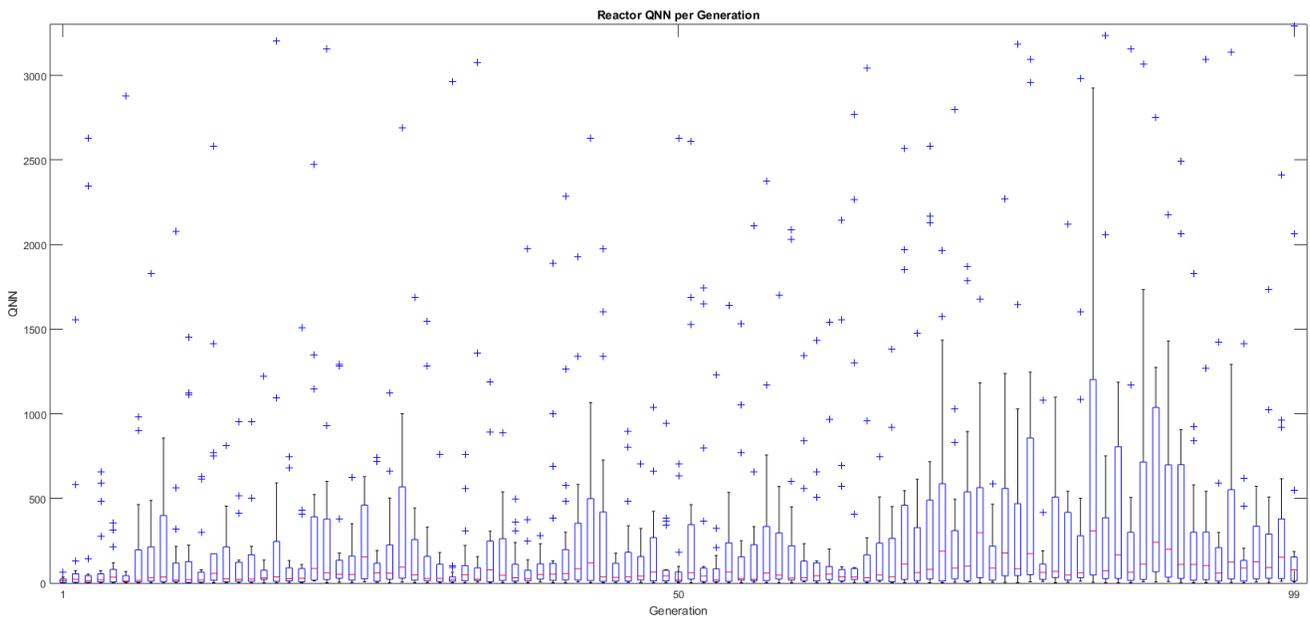


Figure 6: Distribution of the evolutionary activity measure QNN per generation for each reactor.

Reactor	$C$	$V$	$L$	$R$
0	46	2.3667	0.18255	40
1	63	4.3401	0.19970	59
2	129	8.3091	0.20915	83
3	586	430.3	0.51866	307
4	76	7.6058	0.17437	58
5	229	21.694	0.38975	137
6	59	1.8248	0.17602	54
7	139	6.8823	0.27666	116
8	501	1607.6	0.36968	402
9	453	128.54	0.27755	298
10	149	16.221	0.20390	93
11	39	2.5063	0.12235	34
12	566	116.41	0.36923	280
13	274	398.26	0.11454	166
14	608	368.44	0.29997	338
15	266	31.143	0.28607	158
16	183	25.329	0.34497	140
17	58	4.797	0.18159	45
18	179	61.056	0.48978	114
19	50	2.2724	0.12439	38

Figure 7: Gen 99 Reactors

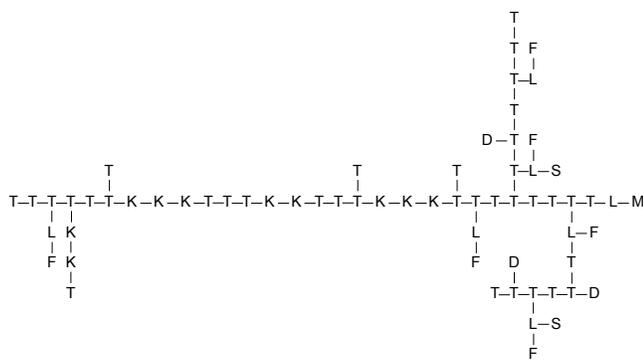


Figure 8: A composite particle produced by reactor 3, made of 65 atomic particles from 7 different species (T,K,L,S,F,M,D).

the end of the run together with an increase in distribution variance suggests that fit individuals have a positive influence on the population.

In order to check if the exploratory algorithm is producing positive evolutionary activity we use the QNN measure (Droop and Hickinbotham, 2012). Fig. 6 shows the QNN distribution per reactor over the generations. Large increases in the QNN distribution, especially towards the latter third of the experiment, suggest a period of strong evolutionary activity for most reactors. Large positive outliers are single reactors which are showing high activity, likely due to mutation. Again due to the random nature of the reactor we see fluctuations between generations.

### Example product

Fig. 7 shows the reactor metrics for each reactor at the end of the last generation. In the final run four of the 20 reactors produced over 500 unique composite particles. Of these the best (based on  $f_r$ ) was reactor 3 producing 586 particles of which 307 had newly created stable bonds.

Fig. 8 shows one of the largest generated composite particles. The main chain consists of T and K atomic particles. We see branching along the chain showing that the T atomic particle is capable of three links. Interestingly we also see that particle L allows other, non-K or -T particles to join the chain (specifically F, M and S). This gives the product compositional diversity. Like the T particle, L is capable of forming up to three links. However the reactor does not contain any long chain L composites, suggesting that L-L links are unstable. While the particle in Fig. 8 is stable there are still T atoms with only one link, suggesting that it could grow even further.

This particle is a product of 47 unique reactions. Most produce exactly one new unique composite particle. However three of the reactions produce two unique particles. This suggests that most reactions are combinatorial in nature. While reactions can produce multiple products, only previously unobserved products are recorded and added back to the reactor.

Branching is common in the final reactors; 15 out of 20 reactors have at least one composite where an atom has three links. However we have not observed a particle that can form four or more links. It is possible that changing the way ILs are constructed to ensure that each particle has at least 4 linking sites of size  $>2$  would give more branching. However engineering particles in such a way is contrary to the core principle of having emergent properties. A more consistent approach would be to find naturally occurring particles of that nature and introduce them into the seed set.

### Conclusions and Outlook

We have presented a new ssACChem based on the Spiky RBN model of an atomic particle. The core design principle of the model is to derive all properties relevant to linking from

the emergent properties of the RBN. We have shown that the method is capable of producing large composite particles with varying structures. Seed sets of atomic particles have been found which are reactive and produce a variety of possible composite particles, as well as a range of reaction paths which can be further explored.

Overall the Spiky RBN model seems to be a viable option for a fully emergent ssAChem. Future work will focus on expanding the available mechanisms beyond simple bonding and stability.

Firstly energetics will be introduced. Bonding and decomposition will depend on meeting collision and stability criteria as well as a probability proportional to reactor temperature. This could result in a relaxation of the stability criteria allowing for more composite species to exist for a short time.

Secondly a spatial 2D reactor will be introduced. Geometry of composites will be determined by number of bonds per particle as well as angles between bonds. The values will again be emergent from the sRBN properties.

As well as bonding, weaker inter-particle interactions could be considered. This could allow the emergence of organisation within a spatial reactor.

The sRBN model provides us with the flexibility to introduce the above mechanisms in a way that is fully emergent from the underlying organisation.

### Acknowledgment

Krastev is funded by an EPSRC DTA PhD studentship.

### References

- Banzhaf, W., Dittrich, P., and Eller, B. (1999). Self-organization in a system of binary strings with spatial interactions. *Physica D: Nonlinear Phenomena*, 125(1):85–104.
- De Castro, L. N. and Von Zuben, F. J. (2000). The clonal selection algorithm with engineering applications. In *Workshop on Artificial Immune Systems and their Applications, Workshop Proceedings of GECCO*, pages 36–39.
- De Castro, L. N. and Von Zuben, F. J. (2002). Learning and optimization using the clonal selection principle. *IEEE Transactions on Evolutionary Computation*, 6(3):239–251.
- Dittrich, P. and Banzhaf, W. (1998). Self-evolution in a constructive binary string system. *Artificial Life*, 4(2):203–220.
- Dittrich, P., Ziegler, J. C., and Banzhaf, W. (2001). Artificial chemistries—a review. *Artificial Life*, 7(3):225–275.
- Droop, A. and Hickinbotham, S. (2012). A quantitative measure of non-neutral evolutionary activity for systems that exhibit intrinsic fitness. *Artificial Life*, 13:45–52.
- Faulconbridge, A. (2011). *RBN-World: sub-symbolic artificial chemistry for artificial life*. PhD thesis, University of York.
- Faulconbridge, A., Stepney, S., Miller, J. F., and Caves, L. S. (2010). RBN-World: The hunt for a rich AChem. In *ALife XII*, pages 261–268. MIT Press.
- Faulconbridge, A., Stepney, S., Miller, J. F., and Caves, L. S. (2011). RBNWorld: Sub-symbolic artificial chemistry. In *ECAL 2009*, pages 377–384. Springer.
- Fontana, W. and Buss, L. W. (1994). What would be conserved if “the tape were played twice”? *Proceedings of the National Academy of Sciences*, 91(2):757–761.
- Hickinbotham, S. J., Clark, E., Stepney, S., Clarke, T., Nellis, A., Pay, M., and Young, P. (2010). Diversity from a monoculture: Effects of mutation-on-copy in a string-based artificial chemistry. In *ALife XII*, pages 24–31. MIT Press.
- Hutton, T. J. (2002). Evolvable self-replicating molecules in an artificial chemistry. *Artificial Life*, 8(4):341–356.
- Hutton, T. J. (2005). Replicators that make all their own rules. In *ECAL Workshop on Artificial Chemistry*.
- Hutton, T. J. (2007). Evolvable self-reproducing cells in a two-dimensional artificial chemistry. *Artificial Life*, 13(1):11–30.
- Kauffman, S., Peterson, C., Samuelsson, B., and Troein, C. (2003). Random boolean network models and the yeast transcriptional network. *Proceedings of the National Academy of Sciences*, 100(25):14796–14799.
- Kauffman, S. A. (1969). Metabolic stability and epigenesis in randomly constructed genetic nets. *Journal of Theoretical Biology*, 22(3):437–467.
- Madina, D., Ono, N., and Ikegami, T. (2003). Cellular evolution in a 3d lattice artificial chemistry. In *ECAL 2003*, pages 59–68. Springer.
- McMullin, B. (1997). SCL: An artificial chemistry in Swarm. Working paper 97-01-002, Santa Fe Institute.
- Ono, N. and Ikegami, T. (2001). Artificial chemistry: Computational studies on the emergence of self-reproducing units. In *ECAL 2001*, pages 186–195. Springer.
- Varela, F. G., Maturana, H. R., and Uribe, R. (1974). Autopoiesis: the organization of living systems, its characterization and a model. *Biosystems*, 5(4):187–196.