

Elongation Control in an Algorithmic Chemistry

Thomas Meyer¹, Lidia Yamamoto¹,
Wolfgang Banzhaf², and Christian Tschudin¹

¹ University of Basel, CH-4056 Basel, Switzerland
{th.meyer,lidia.yamamoto,christian.tschudin}@unibas.ch,
<http://cn.cs.unibas.ch>

² Memorial University of Newfoundland, St. John's, NL, A1B 3X5, Canada
banzhaf@cs.mun.ca, <http://www.cs.mun.ca/~banzhaf/>

Abstract. Algorithmic chemistries intended as computation models seldom model energy. This could partly explain some undesirable phenomena such as unlimited elongation of strings in these chemistries, in contrast to nature where polymerization tends to be unfavored. In this paper, we show that a simple yet sufficiently accurate energy model can efficiently steer resource usage, in particular for the case of elongation control. A string chemistry is constructed on purpose to make strings grow arbitrarily large. Simulation results show that the addition of energy control alone is able to keep the molecules within reasonable length bounds, even without mass conservation, and without explicit length thresholds. A narrow energy range is detected where the system neither stays inert nor grows unbounded. At this operating point, interesting phenomena often emerge, such as clusters of autocatalytic molecules, which seem to cooperate.

1 Introduction

Algorithmic Chemistries [1, 2] are artificial chemistries where algorithms or complex functions emerge as the outcome of stochastic interactions among simple molecules. The role of energy in these chemistries has not been sufficiently emphasized so far. Yet energy plays a crucial role in all chemical processes related to life. Living organisms need a permanent intake of energy in order to drive essential chemical reactions, which otherwise would rarely occur spontaneously. At the same time, energy limitations are important selection factors in evolution.

This paper illustrates one of the potential roles of energy in algorithmic chemistries: that of controlling resource usage in a natural way, in particular, to restrict the lengths of polymers in a string-based artificial chemistry to reasonable bounds, without resorting to arbitrary thresholds.

The paper is structured as follows: Section 2 briefly describes the energy framework used. Section 3 shows elongation control in an evolution scenario with a population of artificial cells based on a constructive chemistry that is particularly aggressive in elongating. The emergence of autocatalytic clusters is observed for a range of energy injection rates. Discussions and conclusions follow in Sects. 4 and 5.

2 Energy Framework

Most algorithmic chemistries operate at the microscopic level of individual molecular collisions, since the outcome of the reaction is computed from the information carried within the molecules [1–3]. Moreover, many algorithmic chemistries are constructive [1, 3, 4], producing new molecular species all the time. This is in contrast to many simulations of real chemistry, where the molecular species and their reactions are known beforehand, and where a deterministic approach based on ODEs (ordinary differential equations) is often sufficient.

We propose an energy framework aimed specifically at algorithmic chemistries, especially constructive ones. It provides a microscopic level stochastic simulation of chemical reactions in a well-stirred vessel under energy constraints, such that the qualitative behavior is similar to real chemistry, yet still sufficiently abstract to incur an acceptable computational cost.

The framework is divided into three steps: collision, reaction, and energy balancing. The first step selects molecules for collision using existing algorithms such as [5, 6].

The second step (reaction) decides whether a reaction will be effective or elastic. This step is based on the fact that the kinetic rate coefficient for a given chemical reaction stems from the activation energy E_a necessary for a reaction to take place, according to the Arrhenius equation: $k = Ae^{-\frac{E_a}{RT}}$ for a reaction such as $X + Y \xrightarrow{k} Z$, where A and R are constants, and T is the absolute temperature. Since the framework operates at the microscopic level, the quantity E_a must be rescaled to a single reaction, obtaining ε_a . Moreover, the kinetic energy ε_k of the colliding reactants must be modeled with low computation penalty. We do this by tracking the overall kinetic energy E_k in the vessel of N molecules rather than the velocity of individual molecules and by drawing ε_k from an exponential distribution, based on results from statistical mechanics [7]: $\varepsilon_k \sim \text{Exp}(E_k/N)$. The reaction is effective if $\varepsilon_k > \varepsilon_a$. It can be easily demonstrated that this procedure leads to Arrhenius behavior at the macroscopic level (mathematical derivation omitted for conciseness).

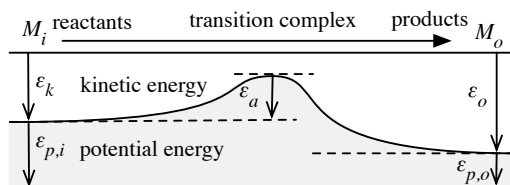


Fig. 1. Energy diagram of an exothermic reaction. The total energy is conserved.

The third step implements energy conservation, i.e. the first law of thermodynamics, by making sure the energy amounts are properly balanced before and after the reaction. Figure 1 shows a typical energy diagram for a reaction.

Energy is conserved as it proceeds from reactants (left side) to products (right side): $\varepsilon_k + \varepsilon_{p,i} = \varepsilon_o + \varepsilon_{p,o}$. Since $\varepsilon_{p,o} < \varepsilon_{p,i}$ in the example, the remaining energy is released, returning to the pool E_k , and the reaction is exothermic. Conversely, the corresponding reverse reaction (from right to left) is endothermic, absorbing energy.

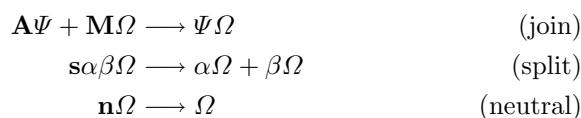
Parameters such as the mapping from molecule to potential energy, as well as the mapping from reaction to activation energy, are left open to be instantiated by the designer to a specific artificial chemistry. Section 3 will show a simple instance of the framework, able to solve the string elongation problem.

2.1 Related Work

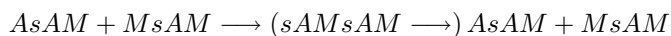
Related energy models aim mostly at the study of real chemistry or biology, especially the origin of life. These existing models do not seem directly appropriate in an algorithmic chemistry context. Some are too complex (down to the quantum level [8]), others too simplified or too specific (e.g. focusing on equilibrium states [9], or on catalytic reactions [10]). Most are unable to handle constructive systems: Many require the designer to specify rate and energy parameters for each species and reaction exhaustively [11]; or assign kinetic coefficients at random [12, 13], without considering the reactants' composition or shape. Some handle shape explicitly [9, 14] but not activation energy.

3 Elongation Control in an Algorithmic Chemistry

A simple string rewriting chemistry is used to illustrate the role of energy in length control. It is a subset of [15] selected on purpose to show a very aggressive elongation behavior. The chemistry consists of polymers strings $s = \Sigma^*$ of arbitrary length over an alphabet of 4 symbols $\Sigma = \{A, M, s, n\}$. The first symbol of a string implicitly defines the string rewriting operation applied to this molecule as follows



where α, β are arbitrary symbols and Ψ, Ω are strings. Strings starting with A or M are in normal form while strings starting with s or n are transient molecules that immediately undergo as many reduction steps as necessary to reach their normal form. For example, the following two rewriting steps are considered one reaction step (the intermediate product $sAMsAM$ is immediately reduced):



Strings in this chemistry tend to increase in length due to the join reaction. Moreover, contrary to nature, mass (in number of symbols) is not conserved, and the split operation nearly doubles the mass of the input.

3.1 Evolution Experiments

The initial population consists of $C = 100$ identical reaction vessels (or “cells”) containing a manually-designed molecular organization that catalyzes its own production: $\{A_{sss}AM, M_{sss}AM\}$. In the absence of resource constraints its concentration rises exponentially.

Using a cell growth-division metaphor, cells divide whenever the number of “membrane” molecules (here: molecules starting with symbol M) reaches a threshold of $N = 1000$, in which case the following operations are carried out: (i) A new offspring cell is generated, (ii) half of the molecules of the dividing cell are randomly selected to move to the offspring, (iii) one of the migrated molecules is randomly mutated, and finally, (iv) the new cell displaces one of the already existing cells, maintaining a constant cell population. Natural selection arises because higher production of membrane molecules leads to earlier division. Thus the most efficient autocatalytic set is expected to survive.

We simulate three scenarios, shown below: one with no restrictions, one with a fixed length threshold, and one with energy control.

Unbounded Growth of Unconstrained System: A significant fraction of all mutations lead to a set of molecules with *infinite closure* [4] that has the potential to reach an infinite sequence space when reacting among each other. Consequently, without any length restriction, such a mutation may easily result in the accumulation of ever-growing strings. The resulting exponentially rising CPU and memory requirements prevent the simulation of such systems to be carried out on today’s computers. We performed 20 simulations, none of which survived 10 generations without exhausting our machine’s resources.

Low Efficiency Under Arbitrary Length Thresholds: We tried two simple methods to prevent strings from growing infinitely: to destroy molecules longer than a certain threshold $l = 10, 20, 30, 100$ (arbitrarily chosen); or to truncate molecules longer than l . Both methods prevent elongation but have an undesired side effect: After several generations, the number of A and M symbols becomes unbalanced due to the stochastic distribution of molecules during cell division. The initial selective advantage of producing more M s than A s goes along with ceasing the production of A s necessary for sustained replication. On the other hand, some cells first create additional A s while still producing M s. Consequently, the rate of reactions among A s and M s rises due to the law of mass action. Since these reactions still generate M s, the membrane production rate increases, too. Such mutants quickly take over the population since their reproductive ratio is about 500 times higher compared to the original program (see Fig. 2(a)). However, this high productivity comes along with a lower *efficiency*, measured as the surface to volume ratio, where the surface is the number of membrane molecules and the volume is defined as the total number of symbols in the cell. At the same time the average molecule length almost reaches threshold l , indicating that the cells fully exploit the length restriction. A typical

representative for $l = 20$ is the following multiset:

$$\{278 MssssssssssssssssAM, 241 MssssssssssssssssAA, 1974 AssssssssssssssssAM, 553 AssssssssssssssssAA\}$$

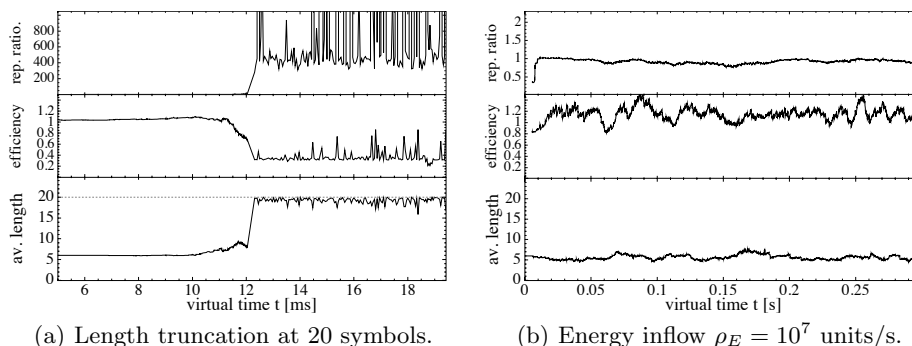


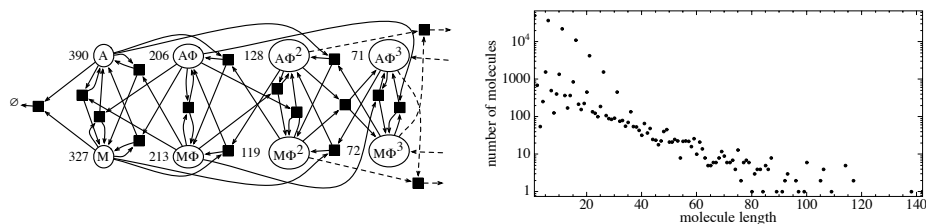
Fig. 2. Simulation results with length restriction (left) and energy control (right). Top to bottom: Average reproductive ratio and average efficiency, both normalized w.r.t. the values of the initial program; and average molecule length.

Emergence of Self-Replicating Clusters Under Energy Control: Instead of externally applying hard length restrictions, the energy framework embeds a notion of resource awareness into the chemistry itself. Less energy slows down reactions and hence the production of new symbols. This method turns out to be very effective in restricting the length of evolved solutions.

We limit the total energy (kinetic plus potential energy) of the system and use a model that, unlike in nature, allows mass to be freely converted to energy and vice-versa: The potential energy of each molecule s is set to its length: $\varepsilon_p(s) = |s|$. The activation energy of each reaction r is set to the difference of potential energy between products and reactants: $\varepsilon_a(r) = \max(0, \varepsilon_p(M_o) - \varepsilon_p(M_i))$, where M_o is the product multiset and M_i is the reactant multiset. Hence, in this chemistry, reactions that build up mass are endothermic, requiring kinetic energy, whereas reactions that destroy symbols are exothermic. Cells receive a constant inflow of energy needed for their growth, whereas the total energy of a displaced cell is lost.

We started with a moderate energy injection rate of $\rho_E = 10^5$ units/s, which results in linear growth. After 100 generations most of the cells still run the initial program; no better solution could be found. Unlike before, now the system cannot increase the reaction rate by excessively producing A molecules. Any production of molecules decreases the temperature and makes endothermic reactions less frequent. Thus a low energy injection rate successfully eliminates mutants with infinite closures.

When increasing energy injection by two orders of magnitude to $\rho_E = 10^7$ units/s, a qualitatively different phenomenon arises: The initial program is able to grow exponentially after each cell division because sufficient energy is available. During cell growth kinetic energy is shared by an exponentially increasing number of molecules. This cools down the cell which gradually returns back to linear growth. Arising mutants that contain syntactically infinite closures start to explore the sequence space by generating longer molecules, but due to the energy restriction the cells are limited in doing so extensively. The existence of longer strings leads to viable mutants that incorporate these new molecules while still being able to survive, i.e. they still have comparable reproduction ratios; the average reproductive ratio remains more or less constant over the whole simulation run (see Fig. 2(b)).



(a) A typical cluster of molecules forming an infinite autocatalytic set. (b) Histogram of molecule lengths after 100 generations.

Fig. 3. Simulation results with energy control for $\rho_E = 10^7$ units/s.

One of the inventions that is often observed in cells after 100 generations is a cluster of molecules of different lengths that all react among each other, as depicted in Fig. 3(a). These molecules are structured as repeated sequences of the duplicate and split pattern $\Phi := sssAM$. In Fig. 3(a), Φ^n represents a string that is the n -fold concatenation of the Φ pattern. The numbers next to the molecules denote their concentration in one of the observed vessels. The cluster contains a lot of very short molecules (A and M), which do not contain the necessary information to replicate themselves. However, they react with a small number of larger molecules that contain multiple copies of the copy and split motif. These longer molecules are maintained at a lower concentration, which exponentially decreases with their length. Even though the cells start to produce larger strings, the average molecule size remains constant and the population maintains its efficiency as shown in Fig. 2(b). Finally, Fig. 3(b) shows the length distribution of a typical simulation run after 100 generations. The cluster consists of molecules of size $5n + 1$: 1, 6, 11, \dots , visible on the upper left part of Fig. 3(b). Cells containing such clusters are prominently present in the population.

When we further increase the inflow of kinetic energy to $\rho_E = 10^8$ units/s, the cells grow exponentially without energy shortage. Consequently, mutants start

to explore the sequence space more aggressively and without hindrance which leads to the same symbolic imbalance as in the fixed length threshold cases.

4 Discussion

Our results indicate that energy control not only helps keeping reasonable resource bounds but also helps to improve the qualitative behavior of the chemistry and steer its evolution. The elongation control experiments show that string length can be kept under control, in spite of an instruction set which is rather aggressive in terms of elongation. The obvious but not optimal solution of a hard length restriction is not needed. Moreover, qualitatively different phenomena are observed according to the amount of energy injected.

A similar elongation phenomenon was described in [3], where a solution based on multi-level selection is also applied. Our method, however, avoids counting on hardware-dependent parameters such as CPU run time.

The experiments reveal a connection between the injected energy and the exploratory capability of a population of cells: If the injected energy flow is too low, the cells grow linearly and the population does not find better solutions. For very high energy injection rates our cells exhibit unbounded growth which leads to symbolic imbalances resulting in a very high reproductive ratio of the affected cells. This is followed by the sudden death of the whole population. Even if the mechanisms behind this behavior are not comparable with those that trigger the elongation catastrophe in [3], interestingly, the resulting effect of the “rise and fall of the fittest” is the same. There exists a range of moderate energy injection rates for which the system is able to survive *and* explore sequence space. “Clusters” of molecules emerge, which altogether form an autocatalytic set. Even though the closure of the set is syntactically infinite, the dynamics of the energy-aware algorithm makes reactions that form long strings more unlikely. This nicely reflects the nature of biochemical reaction system where more and more complex molecules evolve over time in the presence of enough energy.

5 Conclusions

The importance of energy has not been emphasized yet enough, in the context of algorithmic chemistries aimed at performing emergent computation tasks on traditional von Neumann computers. We attempt to fill this gap with an energy framework allowing such algorithmic chemistries to behave thermodynamically and kinetically similar to real chemistry, yet in a simplified form. In spite of “copying” nature’s energy model, the framework still allows for exploration of “life-like computations”, since the degrees of freedom to parametrize the chemistry still remain very large.

We illustrated the framework’s usefulness in a string elongation control task. Even a very aggressive string rewriting chemistry can be restrained by restricting energy alone. Only then and only for a narrow range of energy injection rates the chemistry’s state space exploration can be controlled accurately. An

interesting topic deserving further investigation is how to derive the optimum energy injection rate automatically.

For the future we also aim at evolving programs where the quality of the computed solution is rewarded with energy. Mass and energy conservation could be combined in a more realistic setup. Other exploration venues include computations by artificial metabolisms, as well as the evolution of chemical Carnot cycles [16] with their ability to build up information.

Acknowledgments. This work has been supported by the Swiss National Science Foundation and the European Union, through SNF Project Self-Healing Protocols and FET Project BIONETS, respectively. We would also like to thank Stephan Bürgi for reviewing the chemical physics part of our work.

References

1. Fontana, W., Buss, L.W.: “The arrival of the fittest”: Toward a theory of biological organization. *Bull. Math. Bio.* **56**(1) (1994) 1–64
2. Banzhaf, W., Lasarczyk, C.W.G.: Genetic programming of an algorithmic chemistry. In: *Genetic Programming Theory and Practice II*. Springer (2004) 175–190
3. Decraene, J., Mitchell, G., McMullin, B.: Unexpected evolutionary dynamics in a string based artificial chemistry. In: *Proc. 11th Int. Conf. Sim. Synthesis of Living Systems (ALife XI)*, MIT Press (2008) 158–165
4. Dittrich, P., Speroni di Fenizio, P.: Chemical organization theory. *Bull. Math. Bio.* **69**(4) (2007) 1199–1231
5. Gillespie, D.T.: Exact stochastic simulation of coupled chemical reactions. *J. Phys. Chem.* **81**(25) (1977) 2340–2361
6. Gibson, M.A., Bruck, J.: Efficient exact stochastic simulation of chemical systems with many species and many channels. *J. Phys. Chem. A* **104**(9) (2000) 1876–1889
7. Upadhyay, S.K.: *Chemical Kinetics and Reaction Dynamics*. Springer (2006)
8. Benkő, G., Flamm, C., Stadler, P.F.: Explicit collision simulation of chemical reactions in a graph based artificial chemistry. In: *Proc. 8th Euro. Conf. Adv. ALife (ECAL 2005)*. LNCS, Springer (2005) 725–733
9. Lancet, D., Sadovsky, E., Seidemann, E.: Probability model for molecular recognition in biological receptor repertoires. *PNAS* **90** (1993) 3715–3719
10. Pereira, J.: A biochemistry-inspired artificial chemistry: LAC. In: *Proc. 12th Portuguese Conf. Art. Intel. (EPIA 2005)*. (2005) 79–84
11. Gordon-Smith, C.: Evolution without smart molecules. In: *Extending the Darwinian Framework: New levels of selection and inheritance*” workshop at ECAL 2007, 9th Euro. Conf. ALife. (2007)
12. Bagley, R.J., Farmer, J.D.: Spontaneous emergence of a metabolism. In: *Proc. 2nd Int. Conf. ALife*, Addison-Wesley (1992) 93–140
13. Fernando, C., Rowe, J.: Natural selection in chemical evolution. *J. Theor. Bio.* **247**(1) (2007) 152–167
14. Takeuchi, N., Hogeweg, P.: Evolution of complexity in RNA-like replicator systems. *Biology Direct* **3** (2008)
15. Tschudin, C.: Fraglets - a metabolistic execution model for communication protocols. In: *Proc. 2nd Ann. Symp. Auton. Intel. Net. Sys. (AINS)*. (2003)
16. Smith, E.: Thermodynamics of natural selection II: Chemical carnot cycles. *J. Theor. Bio.* **252**(2) (2008) 198–212