

Linear Self-Assemblies: Equilibria, Entropy, and Convergence Rates

Leonard Adleman* Qi Cheng† Ashish Goel‡ Ming-Deh Huang§
Hal Wasserman¶

Draft: do not distribute
April 28, 2000

Abstract

Over the last few decades, much of the emphasis in computer science has been on building computers that are more complex and correspondingly more powerful. Recently, however, there has been an increasing awareness of the need for alternative models of computation, quantum computation and DNA computing being two notable examples. The latter necessitates the development of a theory of self-assembly. Winfree proved that self-assembling tile systems in a plane are capable of doing universal computation, and when restricted to a line are exactly as powerful as discrete finite automata. Rothmund and Winfree studied the program size complexity of tile systems needed to produce $n \times n$ squares. Both of the above works considered irreversible self-assemblies. In this paper we study reversible linear tile systems.

Motivated by physical chemistry, we define the “profit” of a linear assembly to be the sum of its entropy density and its “bond-stability density.” We prove that reversible linear tile systems must possess an equilibrium. We also show that any linear assembly must be at equilibrium when profit is maximized, and prove several interesting properties of the equilibria of linear assemblies. We then study the rate of convergence to equilibrium for two important cases of linear tile systems: n -linear polymerization and single tile-type in solvent.

*Professor of Computer Science and Molecular Biology, University of Southern California. Research supported by grants from NASA/JPL, NSF, ONR, and DARPA. Email: adleman@usc.edu

†Department of Computer Science, University of Southern California. Research supported by NSF Grant CCR-9820778. Email: qcheng@cs.usc.edu

‡Assistant Professor of Computer Science, University of Southern California. Email: agoel@cs.usc.edu

§Professor of Computer Science, University of Southern California. Research supported by NSF Grant CCR-9820778 Email: huang@usc.edu

¶Email: Halwass@aol.com

1 Introduction

Over the last few decades, much of the emphasis in computer science has been on building computers that are more complex and correspondingly more powerful. Recently, however, there has been an increasing awareness of the need for alternative models of computation, quantum computation [6] and DNA computing [1] being two notable examples. The former has been extensively studied from a theoretical perspective. The latter has led to an interest in the study of self-assembling systems, necessitating the development of a theory of self-assembly. Winfree [10, 11] proved that self-assembling tile systems in a plane are capable of doing universal computation, and when restricted to a line are exactly as powerful as discrete finite automata. Rothmund and Winfree [8] studied the program size complexity of tile systems needed to produce $n \times n$ squares. In both of the above works, a computation is said to be complete when a desired structure has been produced. These papers study irreversible self-assemblies: tiles are allowed to stick together (assemble) but not break apart (disassemble). In this paper we study reversible tile systems and ask the following basic questions:

1. Can the evolution of these tile systems be seen as proceeding in discrete steps?
2. Do reversible self-assembling systems achieve equilibria?
3. Do there exist good characterizations of these equilibria?
4. What is the number of steps required for these self-assemblies to converge to equilibria?

These questions were first asked by Adleman [2] and partially answered in the context of what he called the irreversible n -linear polymerization problem.

Our motivations for studying the equilibria and convergence rates of reversible systems are manifold. Since self-assembly is universal, a good characterization of the equilibria of these systems will allow us to predict which computations can be carried out with adequate yields and what quantities of initial substrates these computations require. Convergence rates are an obvious measure of the time complexity of computation through self-assembly.

To answer these questions in the context of general tile systems is unexpectedly difficult (in two dimensions just the question of whether infinite tiles are formed is a *percolation* problem for which no exact answers are known in the general case [9]). In this paper, we restrict ourselves to linear tile systems, in which the tiles are assumed to lie on a line. Motivated by thermodynamics and physical chemistry [5], we define a “profit” function which is similar to free energy in a chemical reaction. We show that profit is maximized at equilibrium, and prove a weak equivalent of the second law of thermodynamics.

While our results have been motivated by computational questions, self-assembly is a ubiquitous process by which objects autonomously assemble into complexes. Atoms react to form molecules; molecules react to form crystals; cells coalesce to form organisms. In the context of computation, self-assembly is important to both DNA computing [1, 7] and amorphous computing [3]. Thus we believe that a well developed mathematical theory of self-assembly will be useful in a variety of domains.

Our model

We assume that there are n types of tiles arranged on the doubly infinite line, with C_i being the relative fraction of tiles of type i . Some of these tiles may be bonded to their neighbors. We “toss” all the tiles so that they fall at random positions on the line without overlaps or holes and without breaking any existing bonds. Then an existing bond between tiles of types i and j is broken with probability τ_{ij} (where we require τ_{ij} to be larger than 0 for reversible systems). Similarly, if an unbonded right surface of type i comes in contact with an unbonded left surface of type j , a bond is formed with probability σ_{ij} . A supertile is a maximal contiguous set of tiles that are bonded together. A detailed description of our model is presented in Section 2. We will also observe that the supertiles formable by a tile system may be thought of as a language of strings, and that the class of all such languages is closely related to the class of regular languages closed under the substring operation.

Results

We prove that any linear tile system must possess an equilibrium. We define the *profit* of the system as the sum of a term related to information-theoretic entropy and a term related to bond stability. The latter term is motivated by the definition of free energy in chemical reactions. Our main result is that when profit is maximized the system must be at equilibrium, and when the system is at equilibrium profit must be at a critical point (all derivatives zero). We also show that any equilibrium of a tile system must demonstrate *independence*: i.e., whether a tile of type i is bonded on its right with a tile of type j is independent of which tiles (if any) are bonded on its left. Furthermore, any equilibrium must be *strong* in a sense defined below.

For standard thermodynamic systems, it is always true that the system goes to a unique equilibrium. However, we have not been able to prove that the same holds for our tile systems.

We conclude with a study of several important examples of linear self-assembling systems: the reversible and irreversible cases of n -linear polymerization as treated by Adleman, and the case of a single tile-type in solvent. For these systems, we give tight bounds on the number of steps required to reach distance ε from equilibrium, showing in most cases that the distance from equilibrium decays exponentially over time.

Organization

Section 2 presents formal definitions of our model. Section 3 presents some interesting facts about the equilibria of linear self-assemblies. Section 4 presents our main result: that maximizing profit implies equilibrium, and equilibrium implies a critical point for profit. Sections 5 and 6 study the n -linear polymerization problem and the single tile-type in solvent problem, respectively. Section 7 lists open problems.

2 Linear self-assemblies: definitions

Definition 2.1 *An n -tile system consists of:*

1. *A set of tile types $\{T_1, T_2, \dots, T_n\}$*

2. Fractions $\langle C_1, C_2, \dots, C_n \mid C_i > 0, \sum_i C_i = 1 \rangle$
3. A set of (sticking, breaking) probabilities $\langle (\sigma_{ij}, \tau_{ij}) \mid 1 \leq i \leq n, 1 \leq j \leq n, 0 \leq \sigma_{ij} < 1, 0 < \tau_{ij} \leq 1 \rangle$.

Each C_i represents the fraction of tiles that are of type T_i . Also note that we in general require $\tau_{ij} \neq 0$, thereby ensuring reversibility. By a slight abuse of notation, we will use T_i as well as i to refer to the type of a tile. Informally, a tile of type i bonds on its right to a tile of type j with probability σ_{ij} , and such a bond breaks with probability τ_{ij} . Tiles are arranged on a doubly infinite line, one tile at each integer position. Let $X(k)$ represent the tile at position k and $I(k)$ the type of $X(k)$.

Definition 2.2 A linear tiling is a pair $\langle I, B \rangle$, where $I : \mathcal{Z} \rightarrow \{1, \dots, n\}$ and $B : \mathcal{Z} \rightarrow \{0, 1\}$.

As explained above, $I(k)$ denotes the type of the tile at position k . $B(k)$ is 1 if $X(k)$ is bonded to $X(k+1)$ and 0 otherwise.

Definition 2.3 A supertile of a linear tiling is a pair of integers (k, w) , $w > 0$, such that $B(k-1) = B(k+w) = 0$ and $B(k+i) = 1$ for all $0 \leq i < w$. The sequence $S = \langle I(k), \dots, I(k+w-1) \rangle$ is called the type of the supertile (k, w) . The quantity $w(S) = w$ is called the weight of the supertile-type S .

Intuitively, a supertile in a linear tiling corresponds to a maximal contiguous set of tiles bonded to each other.

Definition 2.4 Given a supertile-type S , let $w_i(S)$ denote the number of tiles of type i in S , and let $b_{ij}(S)$ denote the number of times a tile of type i lies immediately to the left of a tile of type j in S .

Definition 2.5 A supertile-type S is said to be permissible if $\sigma_{ij} = 0 \Rightarrow b_{ij}(S) = 0$

Let L be any regular language which is closed under the substring operation. Then we may construct an associated tile system whose permissible supertiles in some sense correspond to the strings of L . (There will be one tile-type for each pair $\langle \alpha, s \rangle$, where α ranges over the letters of the alphabet and s ranges over the states of the associated finite automaton; we then associate supertile $\langle \alpha_1, s_1 \rangle \dots \langle \alpha_k, s_k \rangle$ with string $\alpha_1 \dots \alpha_k$, ignoring s_1, \dots, s_k .) This relation between tile systems and finite automata further motivates our study of self-assembly.

For a given linear tiling, let $N_k(S)$ denote the number of supertiles of type S among the $2k+1$ supertiles at positions $-k$ through k . Furthermore, let $S_1 \cdot S_2$ denote the supertile-type which is obtained by concatenating S_1 with S_2 .

Definition 2.6 A linear assembly is a linear tiling such that:

1. For all k , if $\sigma_{I(k)I(k+1)} = 0$ then $B(k) = 0$. In other words, supertile-types which are not permissible cannot be present.
2. The linear tiling does not contain any infinite supertiles.

3. The limit $f_S = \lim_{k \rightarrow \infty} N_k(S)/(2k+1)$ exists for all supertile-types S . The quantity f_S is called the fraction of tiles of type S .
4. The limit $\bar{w} = \lim_{k \rightarrow \infty} \sum_S N_k(S)w(S)/(2k+1)$ (representing average supertile weight) exists and is finite.
5. $C_i = \sum_S f_S w_i(S)/\bar{w}$ for all i . This ensures that the fraction of tiles of type i is exactly C_i .

Henceforth, we will restrict our attention to linear assemblies.

Definition 2.7 The signature \mathcal{S} of a linear assembly is the mapping f from supertile-types to their fractions.

A signature may also be thought of as a probability distribution defined over supertile-types. We sacrifice formalism for understandability in the following definition.

Definition 2.8 A step (or toss) is the transformation of a linear assembly \mathcal{L} to a linear assembly \mathcal{L}_2 by the following probabilistic process (where all random choices are independent):

1. Obtain a “random” linear assembly \mathcal{L}_1 from the class of all linear assemblies which have the same signature as \mathcal{L} .
2. Obtain a linear assembly \mathcal{L}_2 from \mathcal{L}_1 by allowing bonds of type i, j to break with probability τ_{ij} and to form with probability σ_{ij} .

Informally, at each step, the current supertiles are put in “random” order, and then adjacent tiles of types i, j bond with probability σ_{ij} if they were free, or break apart with probability τ_{ij} if they were bonded.

A toss can be reinterpreted as a probabilistic process defined on the signature of a linear assembly. Under an appropriately defined measure space (we omit the details from this version) it can be shown that the “toss” operation defined on linear assemblies is an ergodic transformation.

Definition 2.9 A linear assembly is said to be in:

1. Species equilibrium if a step preserves the assembly’s signature with probability 1 (in which case such a signature is called an equilibrium signature);
2. Bond equilibrium if a step preserves the quantity $R_{ij} = \sum_S f_S b_{ij}(S)/\bar{w}$ for all i, j with probability 1;
3. Strong equilibrium if the assembly is in species equilibrium and for all supertile-types S_1, S_2, S such that $S = S_1 \cdot S_2$, $f_S \tau_{ij} = f_{S_1} f_{S_2} \sigma_{ij}$ (where the rightmost tile of S_1 is of type i and the leftmost tile of S_2 is of type j).

Intuitively, strong equilibrium occurs when the $S_1 S_2 \Rightarrow S$ reaction and the inverse $S \Rightarrow S_1 S_2$ reaction are equally frequent. The motivation for this definition of strong equilibrium will later be made clear.

Definition 2.10 A linear assembly is independent if $f_{S_1.T.S}/f_{S_1.T} = f_{S_2.T.S}/f_{S_2.T}$ for all supertile-types S_1, S_2, S and all tile-types T such that $f_{S_1.T}$ and $f_{S_2.T}$ are non-zero.

Intuitively, independence implies that what happens to the right of a tile is independent of what happens to the left.

Definition 2.11 We define the entropy density of a linear assembly to be

$$S_I = \frac{-\sum_S f_S \ln f_S}{\bar{w}}$$

and the bond-stability density of a linear assembly to be

$$S_X = \frac{\sum_S b(S) f_S}{\bar{w}},$$

where $b(S) = \sum_{\beta} \ln(\sigma_{\beta}/\tau_{\beta})$, the summation being over all bonds β in S .

The profit Π of a linear assembly is then defined to be the sum $S_I + S_X$.

Intuitively, the entropy density expresses the amount of information-per-tile-position stored in the linear assembly. The definition of bond-stability density is motivated by a corresponding definition in physical chemistry. If r_f and r_r are the equilibrium forward and reverse reaction rates of a chemical process, then the process releases energy proportional to $\ln r_f/r_r$, and this energy is then available to increase the entropy of the external environment. As a consequence, the total entropy of the universe is maximized when the sum of the entropy of a system and the energy released by the system (appropriately normalized) is maximized. Thus a plausible conjecture would be that a linear assembly as defined in this paper is at equilibrium when its profit – the sum of its entropy density and its bond-stability density – is maximized.

However, we note that, while the resemblance between our systems and chemical reactions is striking, we are not aware of any work in chemistry or thermodynamics that directly implies our results.

3 Linear self-assemblies – some basic properties of equilibria and independence

We now present some facts about linear assemblies that are interesting in their own right and will also help us prove our main result. The proof of theorem 3.1 is in appendix A and the proofs of theorems 3.2 and 4.1 are omitted.

Theorem 3.1 Any equilibrium signature must be independent (per Definition 2.10). An equilibrium signature which is independent must also be at strong equilibrium. Thus the notions of strong equilibrium and species equilibrium are identical. Furthermore, bond equilibrium and independence together imply species equilibrium, and vice versa.

Theorem 3.2 The signature that achieves maximum profit (refer to Theorem 4.1) must satisfy independence. Furthermore, $f_S > 0$ for all permissible supertile-types S in this signature.

Theorem 3.3 At equilibrium, profit equals $-\sum_{i=1}^n C_i \ln f_{\mathcal{T}_i}$, where \mathcal{T}_i denotes the singleton supertile-type $\langle T_i \rangle$.

4 A thermodynamic characterization of the equilibria of linear assemblies

In what follows, f_S denotes the fraction of a supertile-type S , and unless otherwise specified summations are over all supertile-types S such that $f_S \neq 0$.

For each supertile-type S , recall that $b(S) = \sum_{\beta} \ln(\sigma_{\beta}/\tau_{\beta})$, where the summation is over all bonds β in S . Then the profit function is

$$\Pi = \left(- \sum_S f_S \ln f_S + \sum_S b(S) f_S \right) / \bar{w}$$

By the definition of a linear assembly, for all i , $\sum_S w_i(S) f_S = C_i \bar{w}$. But $\bar{w} = \sum_S w(S) f_S$. Thus $\sum_S a_{S,i} f_S = 0$, where $a_{S,i} = w_i(S) - C_i w(S)$. Hence we have the following side conditions:

$$\begin{aligned} \sum_S f_S &= 1 \\ \sum_S a_{S,i} f_S &= 0, \quad i = 1, \dots, n \\ f_S &\geq 0, \quad \forall S \end{aligned}$$

Theorem 4.1 *There exists a signature where the profit function is maximized, and this signature is in strong equilibrium.*

Proof: The side conditions define a compact convex polytope which we denote by F . (That F is bounded is easily seen from the constraints $\sum_S f_S = 1$ and $f_S \geq 0, \forall S$.) It follows that the profit function assumes a maximum value in F [4, p. 20].

Recall (from Theorem 3.2) that the profit function can only be maximized at a point where $f_S > 0$ for all permissible supertile-types S . The following general lemma, which we state without proof, implies that such a point must therefore lie in the interior of F .

Lemma 4.2 *Let P be the subset of affine n -space defined by: $Ax = b, x \geq 0$, where A is an $m \times n$ real matrix with $m < n$ and b is an m -vector. Then every point x which satisfies $Ax = b$ and $x > 0$ lies in the interior of P .*

Now that we have shown that the profit function has a maximum which lies in the interior of F , to finish the argument it suffices to show that

Claim 4.3 *At a local maximum of the profit function in the interior of F , the strong equilibrium condition is satisfied.*

To prove the claim we observe that by the Lagrange multiplier method, it follows that at a local extremum in the interior of F , there exist $\lambda, \lambda_i, i = 1, \dots, n$, such that

$$\frac{\partial}{\partial f_S} \left(\Pi - \lambda \left(\sum_S f_S - 1 \right) - \sum_i \lambda_i g_i \right) = 0, \quad \forall S, \quad (1)$$

where $g_i = \sum_S a_{S,i} f_S$. Recalling that $\bar{w} = \sum_S w(S) f_S$, we may calculate

$$\frac{\partial \Pi}{\partial f_S} = \frac{-\ln f_S - 1 + b(S) - w(S)\Pi}{\bar{w}}.$$

It then follows from Eq. (1) that

$$\lambda + \delta(S) = \frac{-\ln f_S - 1 + b(S) - w(S)\Pi}{\bar{w}}, \quad \forall S,$$

where $\delta(S) = \sum_i \lambda_i a_{S,i}$. Multiplying both sides of the above equation by f_S and then summing over all S yields, after further simplification using the definition of Π and the identities $\sum_S f_S = 1$ and $\sum_S a_{S,i} f_S = 0$, that

$$\lambda = -\frac{1}{\bar{w}}. \quad (2)$$

Eqs. (1) and (2) imply

$$-\ln f_S + b(S) - w(S)\Pi = \delta(S)\bar{w}, \quad \forall S.$$

Thus

$$f_S = \frac{e^{b(S)}}{e^{w(S)\Pi + \delta(S)\bar{w}}}.$$

Now we are ready to verify the strong equilibrium condition, that for every pair of supertile-types S_1, S_2 which can concatenate via a bond of type β to form $S = S_1 \cdot S_2$,

$$f_{S_1} f_{S_2} \sigma_\beta = f_S \tau_\beta.$$

Since $\sigma_\beta/\tau_\beta = e^{b(S)-b(S_1)-b(S_2)}$ and $w(S_1) + w(S_2) = w(S)$ and $\delta(S_1) + \delta(S_2) = \delta(S)$, it follows that

$$\frac{e^{b(S_1)}}{e^{w(S_1)\Pi + \delta(S_1)\bar{w}}} \frac{e^{b(S_2)}}{e^{w(S_2)\Pi + \delta(S_2)\bar{w}}} \sigma_\beta = \frac{e^{b(S)}}{e^{w(S)\Pi + \delta(S)\bar{w}}} \tau_\beta.$$

Hence the strong equilibrium condition is met. ■

The following theorem is implicitly proved in the above argument.

Theorem 4.4 *Equilibrium implies that the linear assembly is at a critical point of profit.*

We leave open the question of whether equilibrium is unique, and whether any critical point of profit must be a maximum.

Theorem 4.1 is similar in spirit to the second law of thermodynamics. Ideally we would like to prove that profit is non-decreasing at each step. We have not been able to prove this and instead present a weaker result.

Assume that we are given a linear assembly which is independent per Definition 2.10. Then observe that, with probability 1, the signature of the assembly is entirely determined if we know the current portions of i, j bonds for all i, j .

Let r_{ij} be the fraction of T_i tiles that are bonded to a T_j tile on the right. Let l_{ij} be the fraction of T_j tiles that are bonded to a T_i tile on the left. Let r_i be the fraction of T_i tiles that are unbonded on the right, and similarly let l_j be the fraction of T_j tiles that are unbonded on the left. Clearly $r_i + \sum_j r_{ij} = 1$ and $l_j + \sum_i l_{ij} = 1$. Now define $R_{ij} = C_i r_{ij}$, $R_i = C_i r_i$, $L_{ij} = C_j l_{ij}$, and $L_j = C_j l_j$. Clearly $\sum_j R_{ij} + R_i = C_i$ and $\sum_i L_{ij} + L_j = C_j$. R_{ij} can be interpreted as the

fraction of neighboring pairs of tiles such that the left tile is of type T_i , the right tile is of type T_j , and there is a bond between the two.

It is possible to express the profit of the linear assembly as a function of the variables R_{ij} . Let $\Delta(R_{ij})$ denote the change in R_{ij} as the result of a toss. Let $\partial\Pi/\partial R_{xy}$ denote the derivative of Π with respect to R_{xy} as all other R_{ij} 's are held fixed.

Theorem 4.5 *For any independent linear assembly, $\Delta(R_{ij}) \cdot (\partial\Pi/\partial R_{ij}) \geq 0$ for all i, j .*

Proof: Omitted. ■

This implies that, at every step, each R_{ij} moves in the direction which increases profit. The fact that our systems evolve in discrete time-steps seemingly makes it difficult to prove the stronger claim that profit is non-decreasing at each step, and may even make this claim untrue.

5 n-linear polymerization

As a first problem in self-assembly, we will explore the creation of linear polymers. Polymers are chains of molecular units. A molecule of DNA is a polymer made from the bases adenine, guanine, cytosine and thymine. Proteins are polymers formed from the twenty amino acids. Many plastics such as nylon and rayon are also polymers. In the cell, DNA and proteins are created by using highly specialized enzymes (polymerases) or enzyme complexes (ribosomes) to catalyze the addition of successive units to one end of a growing chain. In the case of some synthetic polymers and in DNA computation, the product is created by the recursive concatenation of smaller polymers. Here we wish to model the latter form of polymerization.

This is our motivation for *n-linear polymerization*, which we define as follows:

Start with equal quantities of tiles T_1, \dots, T_n , where $n \geq 2$. The gluing rule is that the right side of each T_i , for $1 \leq i \leq n - 1$, can glue to T_{i+1} only. Hence the allowable combinations are all subsequences of $T_1 T_2 \dots T_n$. Here we consider the *uniform* case, in which all allowable bonds have the same on-rate $\sigma \in [0, 1]$ and off-rate $\tau \in [0, 1]$.

If $\tau = 0$ (the *irreversible* case), then the equilibrium behavior is evidently for all tiles to form complete $T_1 T_2 \dots T_n$ strands. (We exclude the trivial case $\sigma = \tau = 0$.) Here we will be principally concerned with the equilibrium behavior for $\tau \neq 0$ (the *reversible* case).

Let time-dependent variables v_1, \dots, v_{n-1} be defined as follows: v_i is the portion of the T_i tiles which are currently free to the right. We assume below that $v_1 = v_2 = \dots = v_{n-1}$. This assumption is justified for several reasons. First, if it is ever true, then it is preserved over time. Hence it is justified if the initial state has this symmetry (for example, if we employ the natural initial state $v_1 = \dots = v_{n-1} = 1$). Second, we shall prove in Section 5.3 that, even if this symmetry does not initially hold, the system rapidly evolves toward it.

Hence let v denote the common value of v_1, \dots, v_{n-1} . How does v evolve over one time-step? Of the $1 - v$ portion of T_i tiles currently bonded to the right, each will become free with probability τ . Meanwhile, of the v portion of T_i tiles currently free to the right, each will become bonded with probability σ if the tile to its right is a T_{i+1} . What then is the probability that the tile to the right of a free T_i is a T_{i+1} ? Since all T_1 tiles are free to the left, while a v portion of T_2, \dots, T_n tiles are free to the left, this probability is $v/[v(n - 1) + 1]$. Hence the update rule for v (where v' denotes

the new value) is given by:

$$v' - v = \frac{-v^2\sigma}{v(n-1) + 1} + (1-v)\tau. \quad (3)$$

Now assume $\tau \neq 0$ and let $\rho := \sigma/\tau$. Then it is easily seen that, for v on legal domain $[0, 1]$, this transformation has a unique fixed point $v = \beta$, where β is defined as follows:

$$\beta = \frac{(n-2) + \sqrt{n^2 + 4\rho}}{2(\rho + n - 1)}. \quad (4)$$

Hence, when $v = \beta$, the system is at bond equilibrium. But then we can specify a strong equilibrium for the system, namely when each T_i (for $1 \leq i \leq n-1$) is free to the right with *independent* probability β . Indeed, since this state has bond equilibrium and independence, it follows from Theorem 3.1 that it is a strong equilibrium.

In the following sections, we further investigate the dynamics and equilibrium behavior of n -linear polymerization. In Section 5.1, we determine the rate at which v goes to equilibrium in the irreversible case. In Section 5.2, we prove that, in the reversible case, v goes to equilibrium value β , and we study its rate of convergence. In Section 5.3, we moreover prove that any reversible polymerization system must go to the unique strong equilibrium specified above. Due to space constraints, we defer all proofs to appendix B.

5.1 Dynamics of irreversible polymerization

Here we consider n -linear polymerization in the irreversible case ($\tau = 0, \sigma \neq 0$). For this case, the equilibrium is evidently for all tiles to form complete $T_1T_2 \dots T_n$ strands. Equivalently, the v variable, representing the fraction of gluable surfaces free, goes to zero. We will next determine the rate at which v goes to this equilibrium.

Theorem 5.1 *For irreversible n -linear polymerization with on-rate σ , and for $\varepsilon > 0$, let t be the smallest time such that the v variable, starting from any initial value, is guaranteed to be at most ε if time elapsed is at least t . Then t is at least*

$$\frac{1}{2\sigma(e-1)}[1/(e\varepsilon) - 1] + \frac{n-1}{2\sigma}[\ln(1/\varepsilon)]$$

and at most

$$\frac{e}{\sigma(e-1)}[e/\varepsilon - 1] + \frac{n-1}{\sigma}[\ln(1/\varepsilon)].$$

Another useful measure of system dynamics is the time required for half of the tiles to form complete $T_1T_2 \dots T_n$ strands. We will now determine upper and lower bounds for this time.

Theorem 5.2 *For an irreversible n -linear polymerization system with on-rate σ , start the system with no bonds yet formed. Then the time required for the volume of complete $T_1T_2 \dots T_n$ strands to reach $1/2$ is at least*

$$\frac{1}{2\sigma(e-1)}[(n-1)/e - 1] + \frac{n-1}{2\sigma}[\ln(n-1)]$$

and at most

$$\frac{e}{\sigma(e-1)}[2e(n-1) - 1] + \frac{n-1}{\sigma}[\ln(2(n-1))].$$

We may summarize as follows the magnitude of the bounds we have discovered. The time to reduce v to ε is $\Theta((1/\varepsilon + n \log(1/\varepsilon))/\sigma)$. The time for half of the tiles to form complete $T_1 T_2 \dots T_n$ strands is $\Theta((n \log n)/\sigma)$.

5.2 Dynamics of reversible polymerization

We now turn to reversible n -linear polymerization. System parameters are $n \geq 2$, on-rate $\sigma \in [0, 1]$, off-rate $\tau \in (0, 1]$, and $\rho := \sigma/\tau$. Recall that $v \in [0, 1]$ represents the fraction of gluable surfaces currently free, and that v is fixed over time iff its current value is β as given in (4). Here we will prove that v in fact converges exponentially fast to equilibrium value β . In the final version of this paper, we intend to also provide a corresponding lower-bound for the reversible case.

Theorem 5.3 *For reversible n -linear polymerization, the v variable, starting from any initial value, is guaranteed to be within distance ε from equilibrium value β if time elapsed is at least $\max\{3, 1/\tau\} \cdot \ln(1/\varepsilon)$.*

Observe that, for the irreversible case, the time for v to reach distance ε from equilibrium, expressed as a function of ε , is $\Theta(1/\varepsilon)$. For the reversible case, on the other hand, this time is $O(\log(1/\varepsilon))$. This may seem paradoxical but is in fact easily explained. At irreversible equilibrium, the quantities of tile strands which need to combine are brought down to zero, so that it becomes increasingly difficult for the remaining fragments to find each other.

5.3 Uniqueness of the n -linear equilibrium

Irreversible n -linear polymerization evidently goes to a unique equilibrium—i.e., a strong equilibrium corresponding to all tiles forming complete $T_1 T_2 \dots T_n$ strands. We now turn to the reversible case, for which, in Section 5, we specified a particular equilibrium—a strong equilibrium in which each gluable surface is free with independent probability β . We will now prove that any system must in fact go to this unique equilibrium.

Theorem 5.4 *Starting from any initial state, a reversible n -linear polymerization system will go to a unique equilibrium, namely the state in which each gluable surface is free with independent probability β , where β is as given in (4).*

6 Single tile-type in solvent

Here the system has only two types of tiles: substance tiles (type 1) and solvent tiles (type 2), with $1 > \sigma_{11} > 0$, $1 > \tau_{11} > 0$, and all other σ_{ij} being zero. We will assume independence; thus we need just one time-dependent variable to determine the signature at time t , namely $\gamma(t)$, the fraction of bonded right sides of substance tiles. We omit details of the proof as well as of γ^* and the exact time-evolution.

Theorem 6.1 *This tile system has a unique equilibrium γ^* . Also, if $\Delta(t) = |\gamma(t) - \gamma^*|$, then $\Delta(t)$ decays exponentially in t .*

7 Open problems

1. Is profit a non-decreasing function of time?
2. Do all linear assemblies ultimately reach equilibrium as a result of repeated tossing?
3. Does every linear tile system have a unique equilibrium signature?
4. Analyze n -linear polymerization with respect to arbitrary (non-uniform) on and off rates.
5. Generalize to higher dimensions:
 - (a) Does an analogue of the profit function exist?
 - (b) What is the proper definition of a step?
 - (c) Does every system have a unique equilibrium signature? (For systems of dimension 2 or higher, the answer seems to be no.)
 - (d) Analyze the assembly of squares. (We believe that this problem is difficult even for the case of 2×2 squares.)

References

- [1] Leonard M. Adleman. Molecular computation of solutions to combinatorial problems. *Science*, 266:1021–1024, November 11, 1994.
- [2] Leonard M. Adleman. Towards a mathematical theory of self-assembly. Technical Report 00-722, Department of Computer Science, University of Southern California, 2000.
- [3] H. Abelson, D. Allen, D. Coore, C. Hanson, G. Homsy, T. Knight, R. Nagpal, E. Rauch, G. Sussman and R. Weiss. Amorphous computing. *AI Memo 1665*, August 1999.
- [4] G.E. Bredon. *Topology and Geometry*, Springer-Verlag, 1993.
- [5] D. Eisenberg and D. Crothers. *Physical Chemistry with Applications to the Life Sciences*, Benjamin Cummings, 1979.
- [6] Centre for Quantum Computation web page, <http://www.qubit.org>.
- [7] Paul Rothmund. Programmed self-assembly using lateral capillary forces. *Manuscript*.
- [8] Paul Rothmund and Erik Winfree. The program-size complexity of self-assembled squares. *STOC 2000*.
- [9] D. Welsh. Percolation and the random cluster model: Combinatorial and algorithmic approaches, in *Probabilistic Methods for Algorithmic Discrete Mathematics*, M. Habib, C. McDiarmid, J. Ramirez-Alfonsin, B. Reed (eds.), Springer, 1998.
- [10] Erik Winfree, Furong Liu, Lisa A. Wenzler, and Nadrian C. Seeman. Design and self-assembly of two-dimensional DNA crystals. *Nature*, 394:539-544, 1998.
- [11] Erik Winfree. PhD thesis. Cal Tech, 1998.

A Proof of theorem 3.1

Proof: We will first prove that any species equilibrium must be independent. Suppose there are supertile-types S_1, S_2 , and S (where one or more of S_1, S_2 , and S may be the empty supertile-type) and a tile-type T such that at a given species equilibrium, $f_{S_1 \cdot T \cdot S} / f_{S_1 \cdot T} \neq f_{S_2 \cdot T \cdot S} / f_{S_2 \cdot T}$. The quantity $f_{S_1 \cdot T \cdot S} / f_{S_1 \cdot T}$ can be interpreted as the probability of a tile of type T having S as the maximal contiguous set of tiles bonded to its right, given that S_1 is the maximal contiguous set of tiles bonded to its left; we will denote this probability by $\Pr [T \cdot S | S_1 \cdot T]$. $\Pr [T \cdot S | S_2 \cdot T]$ is defined similarly. Let $\varepsilon > 0$ be an arbitrarily small constant. Let $\Gamma = \min_{ij} \min\{1 - \sigma_{ij}, \tau_{ij}\}$. Notice that the definition of a tile system guarantees that $\Gamma > 0$.

Then take a linear assembly with the given equilibrium signature and perform $1 + (1/\Gamma^2) \ln(1/\varepsilon)$ tosses. After each toss, the probability that any two given tiles are both unbonded on the right is at least Γ^2 . Therefore, given any two tiles of type T , the probability that there was no toss which resulted in both tiles being simultaneously unbonded on the right is less than ε . Let us now look at two tiles X_1 and X_2 of type T which, after the last toss, are bonded on the left to S_1 and S_2 , respectively. If the two tiles were simultaneously unbonded on the right after any toss, then the probabilities of their now being bonded on the right to S must be equal. Therefore

$|\Pr[T \cdot S|S_1 \cdot T] - \Pr[T \cdot S|S_2 \cdot T]| < \varepsilon$. Since ε can be made arbitrarily small, this implies that the two conditional probabilities must in fact be identical.

This concludes our proof that any species equilibrium must be independent. Any species equilibrium must also correspond to bond equilibrium (by definition) and given both bond equilibrium and independence, we can explicitly compute the signature. An easy calculation shows that this signature demonstrates strong equilibrium. ■

B Proofs from section 5

Proof of theorem 5.1: From (3), the update rule for v is

$$v' = v - \frac{v^2 \sigma}{v(n-1) + 1}. \quad (5)$$

First note the following claim:

Claim B.1 *Update rule (5) implies that if $v_0 \geq v_1$ then $v'_0 \geq v'_1$.*

Proof: (5) may be rewritten as

$$v' = v \left[1 - \frac{\sigma}{n-1} \right] + \frac{\sigma}{(n-1)^2} \left[1 - \frac{1}{v(n-1) + 1} \right].$$

Hence the derivative of v' with respect to v is

$$1 + \frac{\sigma}{n-1} \left[\frac{1}{[v(n-1) + 1]^2} - 1 \right].$$

This is non-negative. ■

Next observe, again from (5), that v is multiplied at each step by a factor of

$$1 - \frac{v\sigma}{v(n-1) + 1} = 1 - \frac{\sigma}{n-1} \left[1 - \frac{1}{v(n-1) + 1} \right]. \quad (6)$$

Now let us ask: if the current value of v is v_0 , how long will it take till $v \leq v_0/e$? Since $v \leq v_0$, v is multiplied at each step by a factor which is at least

$$1 - \frac{\sigma}{n-1} \left[1 - \frac{1}{v_0(n-1) + 1} \right].$$

Define Q as follows:

$$Q := \left[\frac{\sigma}{n-1} \left[1 - \frac{1}{v_0(n-1) + 1} \right] \right]^{-1} = \frac{n-1}{\sigma} + \frac{1}{\sigma v_0}.$$

Then v is multiplied at each step by a factor which is at least $1 - 1/Q$. But $Q \geq 2$, so $(1 - 1/Q)^Q \geq 1/4$. Thus, after $Q/2$ steps, v will have been multiplied by a factor which is at least $(1 - 1/Q)^{Q/2} \geq (1/4)^{1/2} > 1/e$.

Hence, after these

$$Q/2 = \frac{1}{2} \left[\frac{n-1}{\sigma} + \frac{1}{\sigma v_0} \right]$$

steps, v will still be larger than v_0/e . But then how many additional steps will it take till $v \leq v_0/e^2$? By Claim B.1, it will take at least the number of steps to decrease v from v_0/e to v_0/e^2 ; and, repeating the above argument, this must be at least

$$\frac{1}{2} \left[\frac{n-1}{\sigma} + \frac{1}{\sigma v_0/e} \right].$$

But then, for $\varepsilon > 0$, how much time is necessary to reduce v from any initial value to at most ε ? By Claim B.1, the greatest time is for initial value $v = 1$; and, repeating the above argument as needed, a lower-bound on this time is

$$\begin{aligned} & \frac{1}{2} \left[\frac{n-1}{\sigma} + \frac{1}{\sigma} + \frac{n-1}{\sigma} + \frac{1}{\sigma/e} + \cdots + \frac{n-1}{\sigma} + \frac{1}{\sigma/e^{\lfloor \ln(1/\varepsilon) \rfloor - 1}} \right] \\ &= \frac{1}{2\sigma} \left[\frac{e^{\lfloor \ln(1/\varepsilon) \rfloor} - 1}{e-1} \right] + \frac{n-1}{2\sigma} \lfloor \ln(1/\varepsilon) \rfloor \\ &\geq \frac{1}{2\sigma(e-1)} [1/(e\varepsilon) - 1] + \frac{n-1}{2\sigma} \lfloor \ln(1/\varepsilon) \rfloor. \end{aligned}$$

This completes our proof of the lower-bound as stated in the theorem. For the matching upper-bound, we return to (6) and ask again: how long will it take for v to decrease from v_0 to v_0/e ?

As long as $v \geq v_0/e$, v is multiplied at each step by a factor which is at most

$$1 - \frac{\sigma}{n-1} \left[1 - \frac{1}{v_0(n-1)/e+1} \right].$$

Define \tilde{Q} as follows:

$$\tilde{Q} := \left[\frac{\sigma}{n-1} \left[1 - \frac{1}{v_0(n-1)/e+1} \right] \right]^{-1} = \frac{n-1}{\sigma} + \frac{1}{\sigma v_0/e}.$$

Then v is multiplied at each step by a factor which is at most $1 - 1/\tilde{Q}$. Thus, after \tilde{Q} steps, v will have been multiplied by a factor which is at most $(1 - 1/\tilde{Q})^{\tilde{Q}} < 1/e$. Hence, after these

$$\tilde{Q} = \frac{n-1}{\sigma} + \frac{1}{\sigma v_0/e}$$

steps, v will be at most v_0/e .

But then, for $\varepsilon > 0$, how much time is sufficient to reduce v from any initial value to at most ε ? By Claim B.1, the greatest time is for initial value $v = 1$; and, repeating the above argument as needed, an upper-bound on this time is

$$\begin{aligned} & \frac{n-1}{\sigma} + \frac{1}{\sigma/e} + \frac{n-1}{\sigma} + \frac{1}{\sigma/e^2} + \cdots + \frac{n-1}{\sigma} + \frac{1}{\sigma/e^{\lfloor \ln(1/\varepsilon) \rfloor}} \\ &= \frac{e}{\sigma} \left[\frac{e^{\lfloor \ln(1/\varepsilon) \rfloor} - 1}{e-1} \right] + \frac{n-1}{\sigma} \lfloor \ln(1/\varepsilon) \rfloor \end{aligned}$$

$$\leq \frac{e}{\sigma(e-1)}[e/\varepsilon - 1] + \frac{n-1}{\sigma}[\ln(1/\varepsilon)].$$

■

Proof of theorem 5.2: First note that we start the system with a (trivial) statistical independence between bond probabilities. This independence will then be continued over time. Thus the state of the system at any given time is completely expressed by the variable v , which represents the independent probability that each gluable surface is free.

The volume of complete $T_1 T_2 \dots T_n$ strands is equivalently the probability that a tile chosen at random is part of a complete $T_1 T_2 \dots T_n$ strand. But this probability is $(1-v)^{n-1}$. Thus we must bound the time required to reduce v till $(1-v)^{n-1} \geq 1/2$.

If $v = 1/(n-1)$, then $(1-v)^{n-1} < 1/e < 1/2$. Hence, in order to satisfy the desired condition, we must drive v lower than $1/(n-1)$. Plugging $\varepsilon = 1/(n-1)$ into the lower-bound of Theorem 5.1 then yields the lower-bound of Theorem 5.2.

Similarly, let $v = 1/(2(n-1))$. Then, since $2(n-1) \geq 2$, we have

$$(1-v)^{n-1} = \left[\left(1 - \frac{1}{2(n-1)}\right)^{2(n-1)} \right]^{1/2} \geq (1/4)^{1/2} = 1/2.$$

Plugging $\varepsilon = 1/(2(n-1))$ into the upper-bound of Theorem 5.1 then yields the upper-bound of Theorem 5.2. ■

Proof of theorem 5.3: Write v as the sum $\beta + \Delta$; thus Δ represents the current distance from equilibrium. Then (3) transforms to the following update rule for Δ :

$$\begin{aligned} \Delta' - \Delta &= \frac{-\Delta^2[\sigma + \tau(n-1)] - \Delta[2\beta(\sigma + \tau(n-1)) - \tau(n-2)]}{\Delta(n-1) + \beta(n-1) + 1} \\ &= -\Delta\tau M, \text{ where} \\ M &= \frac{\Delta(\rho + n - 1) + 2\beta(\rho + n - 1) - (n - 2)}{\Delta(n - 1) + \beta(n - 1) + 1} \\ &= 1 + \frac{\rho}{n - 1} + \frac{\beta(\rho + n - 1) - (n - 1) - \rho/(n - 1)}{\Delta(n - 1) + \beta(n - 1) + 1} \\ &= 1 + \frac{\rho}{n - 1} + \frac{\frac{1}{2}\sqrt{n^2 + 4\rho} - n/2 - \rho/(n - 1)}{v(n - 1) + 1}. \end{aligned}$$

We will now bound the value of the multiplier M . It is easily proved that $\frac{1}{2}\sqrt{n^2 + 4\rho} - n/2 - \rho/(n-1) \leq 0$; hence M is an increasing function of v . Thus M is minimized when $v = 0$, in which case

$$M = 1 + \frac{1}{2}\sqrt{n^2 + 4\rho} - n/2.$$

This is minimized when $\rho = 0$, in which case $M = 1$. Hence we have proved $M \geq 1$, which implies that $-(\Delta' - \Delta) = \Delta\tau M \geq \Delta\tau$.

Similarly, M is maximized when $v = 1$, in which case

$$\begin{aligned} M &= 1 + \frac{\rho}{n-1} + \frac{\frac{1}{2}\sqrt{n^2 + 4\rho} - n/2 - \rho/(n-1)}{n} \\ &= \frac{1}{2} + \frac{\rho}{n} + \frac{1}{2}\sqrt{1 + 4\rho/n^2}. \end{aligned}$$

This is a decreasing function of n , so it is maximized when $n = 2$, in which case

$$M = \frac{1}{2}(1 + \rho + \sqrt{1 + \rho}).$$

But this upper-bound implies that

$$-(\Delta' - \Delta) = \Delta\tau M \leq \frac{1}{2}\Delta(\tau + \sigma + \sqrt{\tau^2 + \sigma\tau}).$$

This is an increasing function of σ and τ , so it is maximized when $\sigma = \tau = 1$, in which case it equals $\Delta(1 + \sqrt{2}/2) < 1.71\Delta$.

In summary, we have shown that $-(\Delta' - \Delta)$, the amount subtracted from Δ in the next time-step, is between $\tau\Delta$ and 1.71Δ . This implies that $|\Delta|$ is multiplied at each step by a factor which is at most $\max\{1 - \tau, 0.71\}$. The theorem as stated then readily follows. ■

Proof of theorem 5.4: Consider the evolution of the system from any initial state. For $1 \leq i \leq n - 1$, let v_i denote the portion of the T_i tiles which are currently free to the right. Then v —the overall portion of glueable surfaces free—equals $(v_1 + \dots + v_{n-1})/(n - 1)$. It is easily seen that the update rule for v_i may then be written as follows in terms of the current values of v_i and v :

$$v'_i - v_i = \frac{-v_i^2\sigma}{v(n-1) + 1} + (1 - v_i)\tau.$$

But then, for any i and j ,

$$\begin{aligned} v'_i - v'_j &= (v_i - v_j) - \frac{(v_i^2 - v_j^2)\sigma}{v(n-1) + 1} - (v_i - v_j)\tau \\ &= (v_i - v_j) \left[1 - \frac{(v_i + v_j)\sigma}{v(n-1) + 1} - \tau \right]. \end{aligned}$$

The quantity in brackets is easily seen to be at most $1 - \tau$ and at least $-2/3$. But this implies that $|v_i - v_j|$ is multiplied at each time-step by a factor which is at most $\max\{1 - \tau, 2/3\}$.

Hence all of v_1, \dots, v_{n-1} are rapidly squeezed together toward a common value. Thus, after sufficient time has passed, they will not differ measurably from each other. Let us assume then (proceeding somewhat informally) that sufficient time has passed that we may take v_1, \dots, v_{n-1} to have common value v . But this is the assumption necessary to apply update rule (3). And, as we saw in Section 5.2, this rule implies that v will then evolve rapidly to unique equilibrium value β .

It remains only to prove that the system goes to independence. Let $\Gamma = \min\{\tau, 1/2\}$. Then, after each step, any given tile is free to the right with probability at least Γ —independent of its tile-type or how it is bonded to the left. (Indeed, if the tile was bonded to the right before the step, then it becomes free with probability τ . If it was free to the right before the step, then with probability at least $1/2$ the tile to its right is not the correct type for it to bond with, in which case it certainly remains free.) But then the proof method of Theorem 3.1 may be used to show that the system rapidly goes to independence. ■