

Note 1/29/00: The document below is a technical report of the Computer Science Department of the University of Southern California:

Adleman, L. Towards a Mathematical Theory of Self-Assembly. Technical Report 00-722, Department of Computer Science, University of Southern California, (2000).

Note 1/29/00: The document may be obtained at:

<http://www.usc.edu/dept/cs/tech.html>

Note 1/25/00: Open problem 1. has been settled in the affirmative.

Note 1/25/00: Open problem 6. has been settled in the affirmative from basic results of percolation theory

Toward a Mathematical Theory of Self-Assembly

Extended Abstract

October, 1999

Leonard M. Adleman
University of Southern California

Abstract

Self-assembly is the ubiquitous process by which objects autonomously assemble into complexes. Nature provides many examples: Atoms react to form molecules. Molecules react to form crystals and supramolecules. Cells sometimes coalesce to form organisms. Even heavenly bodies self-assemble into astronomical systems. It has been suggested that self-assembly will ultimately become an important technology, enabling the fabrication of great quantities of small complex objects such as computer circuits. Recent developments in DNA computing have highlighted the intimate connection between self-assembly and computation. Despite its importance, self-assembly is poorly understood. In this paper, an attempt is made to provide a basis for a mathematical theory of self assembly. A simple mathematical model of self-assembly with 'step counting' is presented and used to investigate the time complexity of polymerization. It is hoped that study of this and other models of self-assembly using the tools of computational complexity will shed light on aspects of catalysis, self-replication, thermodynamics and computation.

1. THE IRREVERSIBLE MODEL

The fundamental objects of our study will be square tiles on the infinite two-dimensional plane. Each side of each tile will be covered with a "specific" glue [Winfree98, Rothmund99]. By specific, we mean that each type of glue will have a set of glues to which it can stick and which can stick to it and a disjoint set of glues to which it cannot stick and which cannot stick to it.

Tiles can be thought to have a north side, south side, east side and west side. Formally, a tile is a four-tuple of glues $\langle g_n, g_s, g_e, g_w \rangle$. All tiles are oriented in the same direction - rotation will not be allowed. Two tiles can stick iff they are adjacent and their abutting sides have glues which can stick.

For now we will not allow stuck tiles to become unstuck. Later, we will describe a reversible model of self-assembly which permits unsticking.

A finite set of tiles stuck to one another in a particular way will be called a super-tile or alternatively an s-tile. In an s-tile, all abutting tiles must be stuck together. Further, an s-tile must be connected. That is, for all tiles t and t' in an s-tile, there exists a sequence of tiles $\langle t_1, t_2, \dots, t_z \rangle$ in the s-tile such that $t_1 = t$ and $t_z = t'$ and for $i = 1, 2, \dots, z - 1$: t_i is stuck to t_{i+1} . In particular, a tile is also an s-tile. If s is an s-tile, then $wt(s)$, the weight of s , is the number of tiles it contains. At the chemical level, tiles correspond to atoms and s-tiles to molecules.

Two s-tiles can stick iff they are adjacent and all abutting sides have glues which can stick. We will require that if two s-tiles are adjacent and some of their abutting sides have glues which can stick and some do not, then the s-tiles cannot stick - and neither can any abutting sides whether or not they have glues which can stick. This last requirement is somewhat unnatural; it is adopted here for convenience since it insures that if two s-tiles stick, they form another s-tile rather than a hybrid object of some kind. This requirement will be dropped in the reversible model.

We will assume that the number of glues is countably infinite; hence the number of tiles and s-tiles are also countably infinite. Fix a numbering of the s-tiles s_1, s_2, \dots . We wish to consider coverings the plane with non-overlapping s-tiles. Let W be a countably infinite sequence of reals w_1, w_2, \dots such that:

$$1 = \sum_{i=1}^{\infty} w_i$$

Such a W will be called a signature. Let $C(W)$ be the class of all non-overlapping coverings of the plane such that for all i :

$$w_i = \lim_{a \rightarrow \infty} \# \{s_i \text{ completely within the range } -a < x, y < a\} * wt(s_i) / 4a^2$$

where an s-tile 'contained' within a larger s-tile is not counted. This requirement assures that if a covering is in $C(W)$ for some signature W , then W is unique.

We will be particularly interested in coverings which are in $C(W)$ for some signature W . Such coverings will be called 'constellations' and $C(W)$ will be called a 'constellation class'. Non-constellation coverings include some which have what might be called 'infinite supertiles' - infinite collections of tiles satisfying the connectivity condition. Non-constellation coverings also include those which have a supertile s_i distributed across the plane in such a way that the limit w_i above is not defined.

We will refer to w_i above as the 'volume' of s_i and denote it by $\prec s_i \succ$. $\prec s_i \succ$ is the portion of tiles in the constellation which are in s-tiles of type s_i . We will refer to $\prec s_i \succ / wt(s_i)$ as the 'concentration' of s_i and denote it by $[s_i]$. $[s_i]$ is the notion of concentration commonly used in chemistry. We will define s_i , the 'fraction' of s_i , as follows:

$$s_i = [s_i] \sum_{i=1}^{\infty} [s_i]$$

s_i is the portion of s-tiles which are of type s_i . s_i corresponds to the notion of mole-fraction used in chemistry.

Note that there exist sets of s-tiles which cannot cover the plane at all. For example, a set consisting of copies of a single s-tile in the form of a comb. Such sets of s-tiles can nonetheless be considered in the context of a 'solvent' (or 'vacuum')- a tile which sticks to no other tile nor to itself - in which the s-tiles are 'immersed'. When convenient we will describe constellations in $C(W)$ by listing the non-zero volumes.

Informally, an assembly proceeds by starting with an initial constellation. The s-tiles of the initial constellation are then 'tossed into the air' and allowed to fall randomly onto the plane covering it without overlaps (nor rotations, nor holes). Two s-tiles from the initial constellation which happen to now be adjacent must, if the glues on their sides allow, stick to form a larger s-tile. After all such sticking has occurred, a step has been completed and a new constellation has been created (or if the result is not a constellation, then the step does not count and is repeated)¹. This process is repeated ad infinitum or until a predetermined end point is reached.

2. THE ASSEMBLY OF LINEAR STRUCTURES IN THE IRREVERSIBLE MODEL

As a first problem in self assembly, we will explore the creation of (linear) polymers. Polymers ("many parts") are chains of molecular units. Each molecule of DNA is a polymer made from the bases adenine, guanine, cytosine and thymine. Each protein is a polymer formed from the 20 amino acids. Many plastics such as nylon and rayon are polymers. During the process of creating combinatorial libraries for DNA computing, long pieces of DNA are created by polymerizing shorter pieces of DNA. In the cell, DNA and proteins are created 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 DNA computation the product is created by the recursive concatenation of smaller polymers. Here we consider the latter form of polymerization.

For all n , let T_1, T_2, \dots, T_n be tiles such that for $i = 1, 2, \dots, n - 1$ the east side of T_i can stick to the west side of T_{i+1} and no other sticking is possible.

Create an initial constellation from the class $C(W)$ with $\prec T_i \succ = 1/n$ for $i = 1, 2, \dots, n$.

In the process of taking the first step, a T_i may be juxtaposed with and stuck to a T_{i+1} creating an s-tile that was not present in the initial constellation. In fact, for $1 \leq i \leq j \leq n$ the s-tile $S_{i,j} = T_i T_{i+1} \dots T_j$ (where $S_{i,i}$ denotes T_i) may be created. Clearly as more steps are taken, the volume $\prec s_{1,n} \succ$ of the 'complete' s-tile would be expected to increase and asymptotically converge to 1 - that is, at equilibrium all tiles would become incorporated into complete s-tiles. But what is the number of steps before half of the original tiles are incorporated into complete s-tiles? That is, what is the expected number of steps before a constellation is reached which has

¹For the problem of polymerization considered in this paper the informal notion of a step in an assembly is adequate. A general formal definition is not yet available (see open problems section)

$$\prec S_{1,n} \succ > 1/2$$

For all n , let $f(n)$ be the number of steps for the above condition to be satisfied.

Theorem: $f(n) = O(n \log n)$ and $f(n) = \Omega(n)$

Proof:

In fact, we will show that for all n , $(n - 1) \log_e(2) < f(n) \leq (3n - 1) \log_e(2n)$

We will think of a constellation in terms of the sides contained in each square of the plane rather than the tile contained.

For $i = 1, \dots, n - 1$, if during the assembly the east side of a T_i sticks to the west side of a T_{i+1} , we say that the east side of T_i is 'used' and the west side of T_{i+1} is 'used'. A side which is not used will be said to be 'new'. The east side of T_n and the west side of T_1 will always be new.

Let C be a constellation of the assembly, and let $E_i = E_i(C)$ be the probability among all new east sides of a new T_i east side (whether or not the west side of T_i is new or used - i.e. T_i may be part of a larger s-tile). Define W_i similarly. Let $V_i = V_i(C)$ be the probability among all T_i east sides of a new T_i east side.

V_i and E_i are related as follows:

$$(1) \quad E_i = V_i/Q$$

where $Q = \sum_{i=1}^n V_i$.

Since all east sides of T_n remain new throughout the assembly, $V_n = 1$ and $Q = 1 + \sum_{i=1}^{n-1} V_i$.

Let $C = C_t$ be the constellation arrived at after t steps of the assembly, and let $C' = C_{t+1}$ be the constellation which follows C in the assembly. Let E_i , W_i and V_i refer to the constellation C and let E'_i , W'_i and V'_i refer to the constellation C' .

Then in C' , each new east side from C will abut (and possibly stick to) a new west side from C . For $i = 1, 2, \dots, n - 1$, a new T_i east side from C will stick in C' and hence become used iff it abuts a new T_{i+1} west side from C . For each such new T_i east side this has probability W_{i+1} . Hence:

$$(2) \quad V'_i = V_i - W_{i+1} * V_i.$$

Noticing that for $i = 1, 2, \dots, n - 1$, $E_i = W_{i+1}$ in the initial constellation and that a T_i east side becomes used exactly when a T_{i+1} west side becomes used, it follows that throughout the assembly $E_i = W_{i+1}$. Hence equation 2 can be rewritten as $V'_i = V_i - E_i * V_i$ and using equation 1 as

$$(3) \quad V'_i = V_i - (V_i * V_i/Q) = V_i(1 - V_i/Q)$$

Since for $i, j = 1, \dots, n-1$, $V_i = V_j$ in the initial constellation, it follows from equation 3 that $V_i = V_j$ in all of the constellations of the assembly. We will denote the common value as V and now use V_t to denote its value for the constellation C_t . Similarly Q_t will denote $1 + \sum_{j=1}^{n-1} V_t = 1 + (n-1)V_t$. Equation 3 becomes:

$$(4) \quad V_t = V_{t-1}(1 - V_{t-1}/Q_{t-1})$$

And hence

$$(5) \quad V_t = \prod_{j=0}^{t-1} (1 - V_j/Q_j)$$

Since V is monotonically non-increasing as a function of time, it follows that $1 - V/Q$ is monotonically non-decreasing. Hence:

$$(6) \quad V_t \geq [1 - (V_0/Q_0)]^t = [1 - (1/n)]^t$$

and

$$(7) \quad V_t \leq [1 - V_{t-1}/Q_{t-1}]^t = [1 - V_{t-1}/(1 + (n-1)V_{t-1})]^t$$

Claim 1: For all n , if $t \leq (n-1)\log_e(2)$ then $V_t > 1/2$

Proof: Let $t = (n-1)\log_e(2)$. Then from inequality 6

$$V_t \geq [1 - (1/n)]^{(n-1)\log_e(2)} > 1/e^{\log_e(2)} = 1/2$$

The result holds for all $t \leq n\log_e(2)$ since V is monotonically non-increasing as a function of t .

□

Claim 2: For all n , if $t \geq (3n-1)\log_e(2n)$ then $V_t < 1/(2n)$

Proof: Let $t = (3n-1)\log_e(2n)$ and assume $V_t \not< 1/(2n)$. Then $V_t \geq 1/(2n)$, and also $V_{t-1} \geq 1/(2n)$ (since V is monotonically non-increasing).

From inequality 7:

$$V_t \leq [1 - 1/(n-1 + (1/V_{t-1}))]^{(3n-1)\log_e(2n)}$$

$$\leq [1 - 1/(3n-1)]^{(3n-1)\log_e(2n)}$$

$$< 1/e^{\log_e(2n)} = 1/(2n)$$

→←

For $t > (3n - 1)\log_e(2n)$ the result follows because V is monotonically non-increasing as a function of t .

□

For the lower bound on $f(n)$, let C_t be a constellation of the assembly with $\prec s_{1,n} \succ > 1/2$, then clearly $V_t < 1/2$. Now by Claim 1, $f(n) > (n - 1)\log_e(2)$ as desired.

The upper bound is obtained as follows. Assume that C is a constellation with $\prec S_{1,n} \succ \leq 1/2$ and with the least probability of new east sides among all east sides.

Claim 3: In C , for all i, j with $i > 1$ and $j < n$, $\prec S_{i,j} \succ = 0$.

Proof: Assume that for some such i, j , $\prec S_{i,j} \succ > 0$. Each $S_{i,j}$ has a new T_j east side. Since $j < n$, new T_j east sides occur with the same probability as new T_{j+1} west sides. It follows that for some k , $\prec S_{j+1,k} \succ > 0$. Hence, the probability of new east sides could be reduced by sticking all $S_{j+1,k}$ to $S_{i,j}$. Further, since $i > 1$, the resulting s-tiles cannot be complete so $\prec S_{1,n} \succ$ is still not greater than $1/2$.

□

So all s-tiles $S_{i,j}$ with non-zero volume have $i = 1$ or $j = n$. Given this, it follows that for $j < n$, $\prec S_{1,j} \succ = \prec S_{j+1,n} \succ$. Conceptually, the s-tiles of type $S_{i,j}$ and of type $S_{j+1,n}$ form ‘broken pairs’ resulting from splitting a complete s-tile into two. So a constellation with $\prec S_{1,n} \succ < 1/2$ and with the least probability of new east sides has (except for s-tiles of volume zero) only complete s-tiles and ‘broken pairs’. Each complete s-tile has n east sides of which one (that of the T_n tile) is new. Each ‘broken pair’ has n east sides of which two are new. It follows that in C the probability of a new east side among all east sides is at least $3/(2n)$. So, to keep $\prec S_{1,n} \succ \leq 1/2$ at least 1.5 out of every n east sides must be new.

As derived above, in the constellation C_t that occurs after t steps in the assembly, the probability of a new east side among all east sides is $(1 + (n - 1)V_t)/n$. Hence for C_t to have $\prec S_{1,n} \succ \leq 1/2$, $(1 + (n - 1)V_x)/n$ must be greater than or equal to $1.5/n$. From this it follows that V_t must be greater than or equal to $1/(2n - 2) > 1/(2n)$. But by claim 2, when $t > (3n - 1)\log_e(2n)$ this is not the case. Hence when $t > (3n - 1)\log_e(2n)$, $\prec S_{1,n} \succ > 1/2$. Hence, $f(n) = O(n \log(n))$ as desired.

□

A priori, it is conceivable that complete s-tiles can be assembled in just $O(\log(n))$ steps by a ‘doubling up’ process. This would be achieved if on step one, polymers of size two were created, on step two, polymers of size 4, etc. But the result above shows that this is not what occurs and that in fact the assembly of complete s-tiles requires exponentially more time.

Typically, when a polymerization reaction takes place in the physical world the monomers assemble within a solvent. In our model, a solvent is a tile which sticks neither to itself nor any other tile. Next we will consider the impact of solvents.

Let T_1, T_2, \dots, T_n be as above, let S be a solvent. Create an initial constellation by choosing from the class C with $\prec T_i \succ = \prec T_j \succ$ for $i, j = 1, 2, \dots, n$. Denote this common volume as $\prec T \succ$. Then $\prec S \succ = (1 - n* \prec T \succ)$. Let $g(n)$ be the expected number of steps to reach a constellation with:

$$\prec S_{1,n} \succ > 1/2$$

Then:

Theorem: $g(n) = O((p + n) \log n)$ and $g(n) = \Omega(p + n)$ where $p = \frac{\langle S \rangle}{\langle T \rangle}$.

Proof: omitted in this extended abstract.

In a typical chemical reaction, the number of solvent molecules would be a constant c times the number of monomers. A constant of approximately 10^7 would not be unusual. Nonetheless, asymptotically in this case: $g(n) = O(n \log n)$ and $g(n) = \Omega(n)$.

3. THE REVERSIBLE MODEL

In the section, the reversible model of self-assembly is presented. Here, stuck s-tiles may become unstuck. Adding reversibility will allow the theory to more accurately model real physical systems and will bring thermodynamics to bear on self-assembly problems. It will also allow for the study of important phenomena such as catalysis and self-replication. In this extended abstract, we will be content merely to define the irreversible model.

Given glues g_1 and g_2 , we will associate rational numbers $\sigma = \sigma(g_1, g_2)$ and $v = v(g_1, g_2)$. We require that $0 \leq \sigma < 1$ and $1 \geq v > 0$. During each step of an assembly, two tiles stuck together with glues g_1 and g_2 will have probability v of becoming unstuck. Two adjacent tiles whose abutting sides have glues g_1 and g_2 will have probability σ of sticking.

An s-tile is allowed to have abutting tiles which are not stuck together. However, connectivity is still required. Hence two distinct s-tiles may contain the same tiles in the same relative positions but differ in the way the tiles are stuck together. Two s-tiles stick iff they are adjacent and at least one of their abutting tiles stick.

4. OPEN PROBLEMS AND RESEARCH DIRECTIONS

The models described here are only a first step toward a mathematical theory of self-assembly. These models may need modification before an adequate theory arises. At this stage, there are some questions which can be precisely posed - a few of these are given below. Some of these questions may be easy, some are probably quite hard. Probably the most important questions await a more complete description and understanding of the models. We will be content to indicate research directions.

Open Problems

1. In the irreversible model, determine the precise asymptotic running time of linear polymerization. I suspect that $f(n) = \theta(n \log n)$.
2. In the reversible model, determine the values of on and off rates for which linear polymerization reaches a constellation with $\langle S_{1,n} \rangle \geq 1/2$.

In each case calculate the number of steps required. Express the number of steps as a function of n and the length of the rational numbers representing the on and off rates.

3. In the irreversible model and reversible models, determine the running time for the assembly of squares. I suspect that this is much harder than the case for linear structures.
4. Give a formal definition of a 'step' in an assembly. This should be in the form of a partial mapping from constellation classes to constellation classes. I have confidence that this is possible.

One possible approach, analogous to that used commonly in chemistry, is to consider sets of recurrences which define how the volume (or alternatively the fraction or concentration) of s-tiles in a solution change with time. This is straight forward for simple systems. For example, for the n-linear-polymerization problem the following recurrences apply:

For all n , for all i, j with $1 \leq i \leq j$
 eqn, for all $t \geq 0$:

$$\langle S_{i,j} \rangle_{t+1} = \frac{wt(S_{i,j})\gamma(S_{i,j})_t \Sigma(S_{i,j})_t}{D(t)}$$

where

$$\gamma(S_{i,j})_t = \begin{cases} 1 & \text{if } i = 1, j = n \\ 1 - E_{i-1}(t) & \text{if } i = 1, j \neq n \\ 1 - W_{j+1}(t) & \text{if } i \neq 1, j = n \\ (1 - E_{i-1}(t))(1 - W_{j+1}(t)) & \text{if } i \neq 1, j \neq n \end{cases}$$

$$\Sigma(S_{i,j})_t = \sum \{S_1\}_t \{S_2\}_t \dots \{S_z\}_t$$

where the sum is over all sequences of s-tiles $\langle S_1, S_2 \dots S_z \rangle$ such that $S_1 * S_2 * \dots * S_z = S_{i,j}$ where $*$ denotes concatenation.

Initial condition:

$$\langle S_{i,i} \rangle_0 = \frac{1}{n}$$

Whether such recurrences can be defined in the general case without formal complications is not clear.

Another approach, at least in the irreversible model, is based on the following. Notice that the definition of a constellation allows for constellations to have adjacent s-tiles which are **unstuck** even though all abutting sides have glues which stick to one another. For example, consider a constellation C with tiles R and B laid out in a checkerboard pattern and with signature: $\langle A \rangle = 1/2$ and $\langle A \rangle = 1/2$. Such a constellation exists even if the east side of R and the west side of B have glues that stick. That is, the possibility of sticking does not require sticking. If all possible 'stickings' have been made we call a constellation 'closed', otherwise we call it 'open'. Now assume that R and B are such that other than the east side of R sticking to the west side of B , no other sticking is possible. We could 'close' C by making all possible stickings (without moving tiles). The result would be a new constellation C' , the closure of C , with $\langle R \rangle = 0$, $\langle B \rangle = 0$ and $\langle RB \rangle = 1$ - where RB denotes the result of sticking R to B . On the other hand, if we assume that in C the tiles R and B are covered on all sides with glues which all stick to one another, then the closure of C would be a covering of the plane consisting of a single 'infinite s-tile' - it would not be a constellation at all. For all constellations C , there is a unique closure. This may allow us to define a map on constellation classes such that $C(W) \rightarrow C(W')$ iff with probability 1, the closure of a constellation in $C(W)$ is a constellation in $C(W')$. The problem is, what measure can be used for this?

If several definitions can be given, are they equivalent?

We will assume that an adequate definition of $C(W) \rightarrow C(W')$ can be given. We will write $C(W) \xrightarrow{t} C(W')$ iff there exist signatures W_1, W_2, \dots, W_t such that $W_1 = W$, $W_t = W'$ and for $i = 1, 2, \dots, t-1$, $C(W_i) \rightarrow C(W_{i+1})$. We will write $C(W) \xrightarrow{*} C(W')$ iff there exists a t such that $C(W) \xrightarrow{t} C(W')$. $C(W)$ will be called an equilibrium class iff $C(W) \rightarrow C(W)$. $C(W)$ will be called a primitive class iff there does not exist a signature W' such that $C(W') \rightarrow C(W)$.

Let $W = \langle w_1, w_2, \dots \rangle$ and $W' = \langle w'_1, w'_2, \dots \rangle$ be signatures and let $\varepsilon \in \mathfrak{R}_{>0}$, we write $|W - W'| \leq \varepsilon$ iff $(\forall i)[|w_i - w'_i| \leq \varepsilon]$. We write $C(W) \xrightarrow{\infty} C(W')$ iff for all $\varepsilon \in \mathfrak{R}_{>0}$, there exists a signature V such that $C(W) \xrightarrow{*} C(V)$ and for all signatures V' such that $C(V) \xrightarrow{*} C(V')$, $|V - W'| \leq \varepsilon$. If there exists a W' such that $C(W) \xrightarrow{\infty} C(W')$, then we say that $C(W)$ reaches equilibrium.

5. In the irreversible model, for n and all rational ε , determine the function $f(n, \varepsilon)$ = number of steps required for the initial constellation in linear polymerization on n tiles to come within ε of equilibrium.

Express the number of steps as a function of n and the length of the ε .

6. In the irreversible model there exist W such that there does not exist a W' with $C(W) \rightarrow C(W')$. For example, consider a single tile t which sticks to itself on all sides. Start with a constellation class having $\langle t \rangle = 1$ (i.e all tiles are t but none are stuck). Then after one step, all tiles connect and complete cover the plane, which is not a constellation. Do such W exist in the reversible model? For example as above, but with the tile's glues having on-rate=99/100 and off-rate=1/100? In this case it seems the answer would be yes. Characterize the on-rate, off-rate pairs for which the answer is yes. Notice, that a covering which has an 'infinite s-tile' may still be a constellation, if the sum of the volumes of the s-tiles is 1.
7. If W is a signature with only finitely many non-zero volumes and each of these volumes is rational, then in a natural way, W may be represented by a finite input string R . R will be called a representation of W . We will not describe such a representation in detail in this abstract. Let $|R|$ denote the length of R is binary
 - (a) Let R be a representation of a signature W . Is it decidable whether $C(W)$ reaches equilibrium?
 - (b) Let R and R' be representations of signatures W and W' , is it decidable whether $C(W) \xrightarrow{*} C(W')$?
 - (c) Let R and R' be representations of signatures W and W' , let ε be a positive rational, is it decidable whether there exists a signature W'' such that $C(W) \xrightarrow{*} C(W'')$ and $|W'' - W'| < \varepsilon$?
 - (d) Let R be a representation of a signature W . Assume there exists a W' such that $C(W) \xrightarrow{\infty} C(W')$. Let ε be rational. Let $T(R, \varepsilon) =$ the least time t , such that there exists a W'' with $C(W) \xrightarrow{t} C(W'')$ and $|W' - W''| < \varepsilon$. Is T (linear) polynomial in R and ε ?

Research Directions

It is difficult to give clear directions at this point in the development of a theory of self-assembly. Despite this, and at the risk of making some foolish statements, the author will proceed. He humbly asks for the indulgence of the reader.

Computation

Clearly, a self assembly can be simulated to arbitrary accuracy on a computer. This may suggest that assemblies are not very complex. This is not likely to be the case. In fact Wang [Wang61] showed that Turing machines can be 'simulated' by tiles used to cover the plane. Wang was not concerned with steps, but he did show that there was an effective procedure which given a Turing machine T and an input x , produced a finite set of tiles which would cover the plane without overlaps or holes iff the Turing

machine halted on input x . Whether Turing machines can be simulated within the irreversible model, the reversible model or other models of self-assembly is an important question. Winfree has described a model of self-assembly wherein Turing machines may be simulated to arbitrary accuracy [WinfreePhD]. An affirmative answer brings the entire theory of recursive functions and computational complexity to bear on the assemblies in that model and on the physical phenomena it describes. It also brings undecidability.

Thermodynamics

Presently, the most sophisticated theory of self-assembly is the thermodynamics which describes chemical reactions. From the point of view of thermodynamics, the assemblies above correspond roughly to chemical reactions occurring in a bomb calorimeter - a container which allows no work to be done on the surroundings and allows only heat to be passed to the surroundings.

Though thermodynamics - in the form of statistical mechanics - will provide a very deep theory from which to draw, the theory of self-assembly should not become simply a recapitulation of thermodynamics. Ideally, a theory of self-assembly will create the opportunity for thermodynamics, the theory of recursive functions, computational complexity, information theory and theories of randomness to be combined in a synergistic fashion.

Self-assembly differs from thermodynamics in several ways:

1. The questions asked are not those which would typically be asked by a chemist or physicist. For example, how to form squares. What are the 'assemblable shapes' - (analogous to what are the 'computable functions')? Can quasi-crystals be classified on the basis of their 'assembly complexity' (analogous to 'computational complexity').
2. While thermodynamics is directed at equilibrium states and kinetics is directed at reaction rates, both are intrinsic parts of the basic model of self-assembly presented here. This may permit the power of 'step-counting', which has been developed to great effect in computational complexity over the last few decades, to be applied to problems which are not amenable to the techniques of either thermodynamics or kinetics.
3. The understanding of the kinetics of reactions often relies on experiments and is consequently inductive. The theory of self-assembly presented here is based on proof and is intrinsically deductive.
4. When deductive methods are used in kinetics, they are sometimes based on *ad hoc* assumptions suitable for the questions at hand. Ideally a mathematical theory of self-assembly will aim at exposing general truths applicable to all or at least large classes of situations.

Thermodynamics may offer many opportunities for self-assembly. For example, when considering thermodynamics, one must consider entropy. This is a point of potential importance, in that entropy (in some of its forms) is intimately connected to information theory and Chaitin-Kolmogorov complexity [Bennett82, Oswald91, Zurek99]. The relationship of these computational forms of information and randomness to the thermodynamic concepts of entropy and equilibrium should be explored within the self-assembly framework.

With regard to Chaitin-Kolmogorov Complexity, Rothmund and Winfree have considered and answered the following interesting question in a model of self-assembly related to the ones described here: What is the minimum number of distinct tile types necessary to build an n by n square [Rothmund-Winfree]?

In chemical reactions, sometimes the energy of one reaction is used to drive another reaction. In the assemblies above, the 'energy' which is released when two tiles stick together (via glues with higher on-rates than off-rates) is not being used in this fashion but rather is being directly dissipated as heat. This

may be an important omission. How can one appropriately include such phenomena in the model of self-assembly?

Catalysis

Though many chemistry books contain a definition of 'catalyst', it has been quite difficult (for the author anyway) to find an adequate formal definition within the theory of self-assembly. Such a definition is quite important because without it, a complete formal study of this important phenomena is not possible. Assuming an adequate definition can be found, then providing an example of a catalytic assembly would be quite interesting. What is the 'smallest' assembly in which catalysis occurs? Also, it would be important to characterize, to whatever extent possible, the mechanisms of catalysis. This problem may require modification of the basic models of self-assembly given here. For example, in order to study steric features, it may be worthwhile to introduce 'mega-tiles' - s-tiles with all internal sides stuck with 'irreversible glues' (i.e. off-rate 0) and all external sides coated with 'reversible glues' (i.e. non-zero off-rate).

Replication

It seems almost certain that molecules which reproduced were predecessors to life as it now exists. Replication is closely related to catalysis. In particular, a molecule which catalyzes its own creation is undergoing a form of replication. Give an example of an assembly involving replication (this may be related to the recursion theorem). What is the 'smallest' self-replicating assembly? Characterize, to whatever extent possible, mechanisms of replication. This problem may require modification of the basic models of self-assembly given here.

As a specific problem, consider 'equimolar' (i.e. equal volume) tiles A, T, C and G each of which stick to all others (including itself) on both the west side and the east side. All east-west glues have the same on-rate and all have the same off-rate. Clearly for all lengths, all possible 'DNA sequences' are equally likely to form at equilibrium. What changes in this basic setup might lead to a single, or small number of species predominating (perhaps only transiently)?

Alternative models

There are many alternative and/or augmented models for self assembly. Perhaps some of these are by some standards better than those presented here. However, it is important that the mathematical study of self-assembly not degenerate into the consideration of numerous special *ad hoc* settings. At best, the mathematical theory should expose fundamental aspects of self assembly which apply to large classes (ideally all) models. That said, some possible extensions of the models presented here would be the following:

Generalization to 3-space.

Provision for additional 'operations' such as the merging of self-assemblies (e.g. the addition of new s-tiles to an assembly at a pre-specified time), the separation of self-assemblies (e.g. the specific removal of s-tiles at a pre-specified time). From studies of DNA computation, it is known that such operations provide universal computation under some conditions. There may be constellations which can only be created with, or which can be created more efficiently with such additional operations.

5. ACKNOWLEDGEMENTS

I wish to thank the following researchers for contributing their ideas to this paper: Paul Rothmund, Erik Winfree, Ming-Deh Huang, Ashish Goel, Qi Cheng, Joel Zeitlin, Bob Lucas, John Reif and Dick Lipton.

REFERENCES

- [Bennett82] Charles Bennett. The thermodynamics of computation - a review. In E. Fredkin, R. Landauer and T. Toffoli, editors *Proceedings of the Physics of Computation Conference*, in *Int. J. Theor. Phys.*, issues 21:3/4, 21:6/7, 21:12, pages 905-940. 1982.
- [Oswald91] Jacques Oswald. *Diacritical Analysis of Systems: a treatise on information theory*. Translator: Roy Lisker. Ellis Horwood Limited, 1991.
- [Rothmund99] Paul Rothmund. Programmed self-assembly using lateral capillary forces. Manuscript.
- [Rothmund-Winfree] Paul Rothmund and Erik Winfree. The program size complexity of self-assembles squares. Manuscript, 1999.
- [Wang61] Proving theorems by pattern recognition. II. *Bell Systems Technical Journal*, 40:1-42, 1961.
- [Winfree98] 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.
- [WinfreePhD] Erik Winfree. PhD Thesis *CalTech*, 1998.
- [Zurek99] Wojciech Zurek. Algorithmic randomness, physical entropy measurements and the demon of choice. In Anthony J.G. Hey, editor, *Feynman and Computation*, pages 393-410. Perseus Books, 1999.