SANDIA REPORT
SAND2000-0875
Unlimited Release
Printed April 2000

Prediction of Self-Assembly of Energetic Tiles and Dominos: Experiments, Mathematics and Software

Sorin Istrail, Alan Hurd, Ross Lippert, Brian Walenz, Serafim Batzoglou, John H. Conway and Freddie W. Peyerl

Prepared by
Sandia National Laboratories
Albuquerque, New Mexico 87185 and Livermore, California 94550

Sandia is a multiprogram laboratory operated by Sandia Corporation, a Lockheed Martin Company, for the United States Department of Energy under Contract DE-AC04-94AL85000.

Approved for public release; further dissemination unlimited.

Sandia National Laboratories

Issued by Sandia National Laboratories, operated for the United States Department of Energy by Sandia Corporation.

NOTICE: This report was prepared as an account of work sponsored by an agency of the United States Government. Neither the United States Government, nor any agency thereof, nor any of their employees, nor any of their contractors, subcontractors, or their employees, make any warranty, express or implied, or assume any legal liability or responsibility for the accuracy, completeness, or usefulness of any information, apparatus, product, or process disclosed, or represent that its use would not infringe privately owned rights. Reference herein to any specific commercial product, process, or service by trade name, trademark, manufacturer, or otherwise, does not necessarily constitute or imply its endorsement, recommendation, or favoring by the United States Government, any agency thereof, or any of their contractors or subcontractors. The views and opinions expressed herein do not necessarily state or reflect those of the United States Government, any agency thereof, or any of their contractors.

Printed in the United States of America. This report has been reproduced directly from the best available copy.

Available to DOE and DOE contractors from
Office of Scientific and Technical Information
P.O. Box 62
Oak Ridge, TN 37831

Prices available from (703) 605-6000
Web site: http://www.ntis.gov/ordering.htm

Available to the public from
National Technical Information Service
U.S. Department of Commerce
5285 Port Royal Rd
Springfield, VA 22161

SAND2000-0875
Unlimited Release
Printed April 2000

Prediction of Self-Assembly of Energetic Tiles and Dominos: Experiments, Mathematics and Software [1]

Sorin Istrail, Alan Hurd, Ross Lippert, Brian Walenz
Applied Mathematics Department
Sandia National Laboratories
P.O. Box 5800
Albuquerque, New Mexico, 87185-0819

Serafim Batzoglou
Massachusetts Institute of Technology
Cambridge, MA

John H. Conway
Princeton University
Princeton, NJ

Freddie W. Peyerl
Harvard University
Cambridge, MA

Abstract

Self-assembly is, essentially, the study of the optimal conformations of a set of discrete units (amino acids, nano-bricks, etc), where optimality is determined by internal nearest neighbor interactions, and homogeneous external interactions. This report summarizes the recent advances made by self-assembly researchers at Sandia.

Work on general self-assembly theory is motivated by our ability to realize physical processes in which the theory has relevance. The fabrication and self-assembly of Janus bricks is explored

[1]Research supported in part by the Sandia National Labs, Laboratory Directed Research and Development program and in part by the DOE, Math Information and Computional Sciences Program.

in a laboratory setting, and the results of our early experiments with the Janus bricks are presented.

We present a framework for general self-assembly problems, in which the optimal conformations are characterized by the extreme points of a set of (possibly unknown) constraints. The origin of these constraints is in the combinatorial geometry of the discrete units, and an algorithm for determining these constraints via numerical experiments is presented.

The study of constraints can be specialized to an important problem arising in protein folding. We prove approximate constraints on the conformations of certain popular protein folding models which characterize the properties of the energetically optimal conformation to within a small (possibly zero) error.

The analysis of protein conformations requires very flexible tools for visualization. This leads us to the development of TORTILLA a general protein modeler. TORTILLA is our answer to the classical "ball and stick" models of yesterday, providing us the ability to rapidly visualize conformations from an ever-expanding set of protein models.

Acknowledgements

This work was started in 1998 with support from Fred Howes as part of the MICS program. His enthusiasm was an inspiration for all of us.

We would like to thank Bill Camp for initiating research in statistical mechanics and computer science which was the starting point for this project. Discussions and support from Ken Dill, Grant Heffelfinger, Jonathan King and Russell Schwartz are acknowledged with pleasure.

This project was supported by: the Sandia Laboratories' Laboratory Directed Research and Development office, the Sandia Laboratories' Computer Science Research Foundations, and the DOE Math Information Computer Science Program,

Part I

Experiments on the Development and Self-Assembly of Janus Nanobricks

Self-assembly may someday allow one to create a desired engineering part using only preprogrammed small pieces that will interact in a very specific fashion, much the way a virus self assembles. For this process to become practical, or even possible, simple interactions must be controlled through fundamental understanding. Hydrophobic and hydrophilic interactions have proven to be of great interest in self-assembly process because they are very strong and can be controlled. Ceramic and glass Janus nanobricks were made and several self assembly experiments were carried out.

Introduction

Self-assembly is a process in which several small units interact with one another to form one larger product. This process is observed in many different biological systems, but it is most striking and well studied in viruses. Viruses produce all the components required to form a mature virus within a host cell. After all the components are produced, they come together to form a complete mature virion.

The forces that drive the assembly of virus particles include hydrophobic and electrostatic interactions – only rarely are covalent bonds involved in holding together the multiple subunits. In biological terms, this means that protein-protein, protein-nucleic acid, and protein-lipid interactions are present[1].

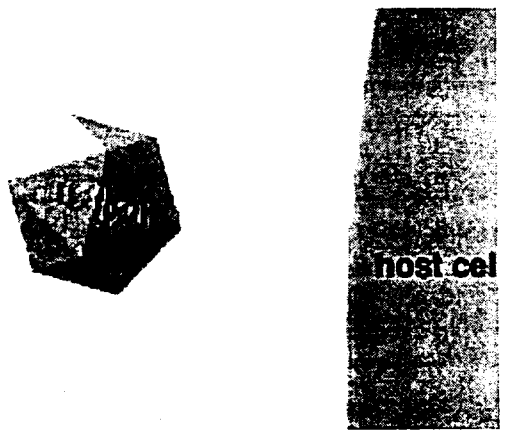

Figure 1 A virus capsid self-assembled from only 20 equilateral triangular tiles. (Picture from Scientific American)

One interesting part of viruses is the capsid, which encapsulates the genetic material of the virion. The capsid is made up of proteins in the shape of thin equilateral triangles. Only 20 of these triangles are required to form the icosahedral capsid (Figure 1). The simplest regular polyhedron is the tetrahedron with only four triangular tiles.

This research project involves both investigation and understanding of some of these simple interactions, specifically hydrophobic and hydrophilic. In addition to understanding the interactions, the pieces, Janus nanobricks, need to be created. We use the term Janus, the two-faced God of Gateways, to signify bodies that have two or more types of surface affinities[1].

Programmed with an unlimited range of surface affinities, any shape could be self-assembled if the bricks had sufficient mobility and time to find their proper neighbors. If this process could be controlled, current fabrication process would be revolutionized. A recent paper by Rothemund shows how self-assembly in 2d can be used to compute, in a way similar to DNA computing.[2]

Experiments

I. Hydrophobic and Hydrophilic Interactions

One of the simplest interactions to investigate and manipulate is the interaction mediated water. In 3d systems, hydrophobic (water repelling) or hydrophilic (water attracting) are important; in 2d systems, such as particles trapped at an air-water interface, capillary forces are dominant.

To study hydrophobic/hydrophilic interactions, a two-sided primitive Janus plate is the simplest configuration. The plate must be hydrophobic on one side but hydrophilic on the other. In 3d suspension, these plates should associate in pairs, hydrophobic-to-hydrophobic surfaces touching.

Scotchlite™ Glass Bubbles, or "microballoons," were obtained ranging in size from 35 to 135 □m in diameter. Glass Bubbles are thin-walled (approximately 2 □m) borosilicate, hollow glass spheres. The advantage of using microballoons is that outside surface can be chemically treated separately from the inside.

The microballoons were sieved to narrow the size distribution to 90 to 100 □m in diameter. Approximately 10 ml were sieved and then coated with a solution of 5% Me_3SiCl silane in hexane. The microballoons were placed in 10 ml of the silane solution and soaked for 30 minutes. They were then soaked in reagent-grade ethanol (EtOH) for one hour and washed with EtOH.

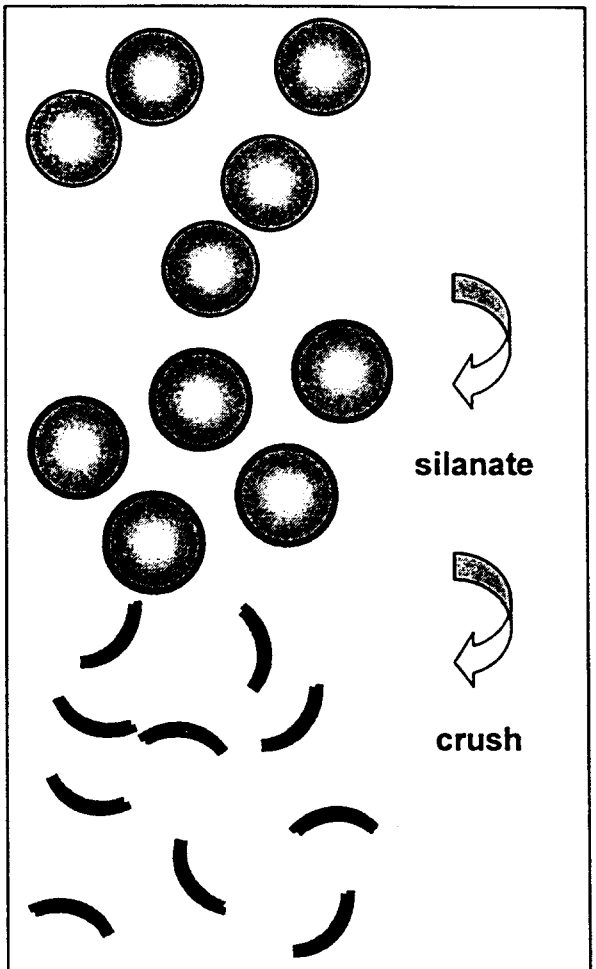

Figure 2 .A graphic representation of the primitive Janus bricks from crushed microballoons.

The dried microballoons were placed between two aluminum blocks (ca 10 cm^2) and subjected to 1000 psi of pressure in a press. The pressure crushed the microballoons into small shards of glass, approximately 10 □m in extent. Each shard has two different surface affinities, with the outside hydrophobic and the inside hydrophilic.

These primitive Janus bricks were then spread across the surface of water in one of two ways. For side-view microscopy, the water substrate was contained in a quartz cuvette was 1 mm by 1 cm and overfilled to be slightly convex; for top-views, a one-centimeter-diameter circular well in a glass microscope slide was used. In some experiments the shards were applied with EtOH as a spreading agent and in others by sprinkling. Particles tended to center themselves near the convex apex of the water substrate. Studies were also done using a two-phase liquid system (oil and water). In that case the crushed microballoons were stirred into the solvent mixture and allowed to separate.

II. Janus Nanobrick Formation using LIGA molds

The Janus bricks well suited for microcomponents would be regular shapes that tile space; ceramic materials offer the most advantage but the highest challenge. In two dimensions, our target

was a perfect square or an equilateral triangle made of silica or alumina, as these shapes are needed for simple aggregate structures while the materials are good model ceramics.

LIGA technology was used to fabricate the molds. LIGA is an acronym from German words for lithography and galvanic electroplating. LIGA molds consist of a silicon wafer substrate covered by a patterned polymer (PMMA) sheet of uniform thickness. The pattern of holes is made by x-ray exposure from a highly collimated synchrotron source through a mask. Thus the sidewalls of the PMMA holes are precisely vertical. The PMMA is chemically developed to create a high aspect ratio, parallel-wall mold[2].

We filled the mold with a desired material then dissolved the PMMA or decomposed it with heat leaving a three-dimensional micropart. This micropart could be separated from the substrate after further processing to set the affinities of the surfaces.

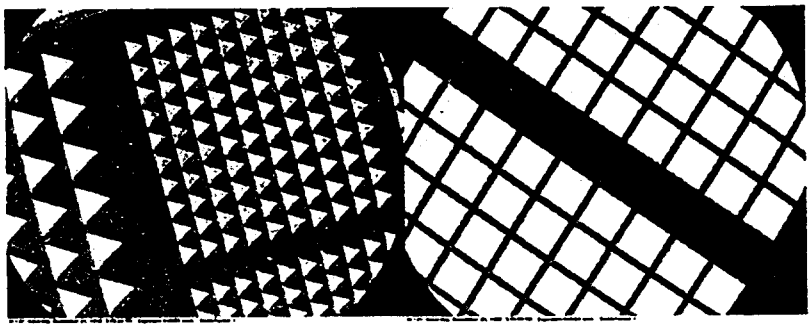

Figure 3 Alumina plates made by dry pressing and bisque firing. The larger triangles and squares are 1 mm on each side, thickness 0.2 mm.

Sol-gel processing in LIGA molds. Initial attempts were made to form bricks from sol-gel materials. While these attempts failed, the procedure is noted here. Two different types of sol-gel were used. The first was a 'B2', a base-catalyzed gel which becomes a gel in about 15 minutes.

The 'B2' sol-gel was made from a TEOS (tetraethylorthosiloxane) solution and mixed with a catalyst (0.7N NH_4OH) in a 10:1 ratio. The mixture was then applied to the surface of the LIGA mold and placed in a desicator for one hour. The desiccator was filled with EtOH (ethanol) to help reduce the rate of evaporation.

To ensure complete filling of the mold, additional volumes of the sol were added every half-hour for a total of three applications. The sample was then either placed in a 50°C oven overnight or left at room temperature overnight.

The same procedure was performed with the 'A2', an acid-catalyzed sol with 2.0N HCl in the same ratio of 10:1 ('A2' stock : acid) as the base-catalyzed B2 samples.

To consolidate the samples, the A2 and B2 samples going into the oven were placed in one covered petri dish. The two samples, which sat at room temperature, were also placed in one covered petri dish. Unfortunately the large amount of shrinkage and concomitant cracking of these silica materials during drying made them unusable for forming molded bricks.

Bisque-fired ceramic powder processing. A successful process was developed for making Janus plates of precise shape and thickness using dry ceramic powders. A fine powder was pressed into the PMMA molds using ca 1000 psi press between flat anvils. Excess powder was scraped off with a sharp blade, then the wafers were bisque-fired at 1000 C for 1 hr (ramping up and down over a period of 1 hr each). In bisque firing, the particles become lightly but firmly sintered together with minimal part shrinkage (and not full density). After firing, the PMMA had been cleanly burned away, leaving the Janus plates firmly adhered to the wafer where they could be surface treated before removal using a sharp blade.

Results and Discussion

I. Hydrophobic and Hydrophilic Interactions between JanusMicroballoon Shards

In order to study the interactions between the shards, we first had to be sure that the surfaces were successfully coated with the silane. The simplest check was simply comparing the normal and the silanated whole microballoons. By comparing the differences in contact angles of the water with the

microballoons, it became very obvious that the silanation procedure made the surface hydrophobic (Figure 4).

The next step was to observe the interactions between the crushed microballoons. The most successful procedure for spreading the glass shards was simply sprinkling them onto the surface of the water. When the normal untreated shards were spread on the surface, they floated around and didn't appear to interact with one another.

The silanated shards were spread in the same manner and immediately packed together. The sprinkling also caused some shards to land on top of others. Those shards on top of others almost immediately pulled spontaneously to the surface of the water.

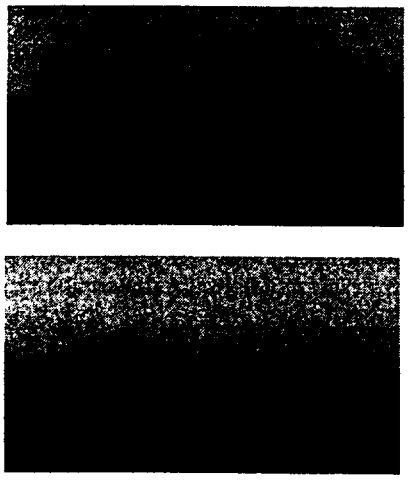

Figure 5 Side view of hydrophobic, silanated microballoons (top) and normal untreated microballoons (bottom) at the surface of water. The hydrophobic balloons exhibit contact angles of 90° while the untreated balloons do not.

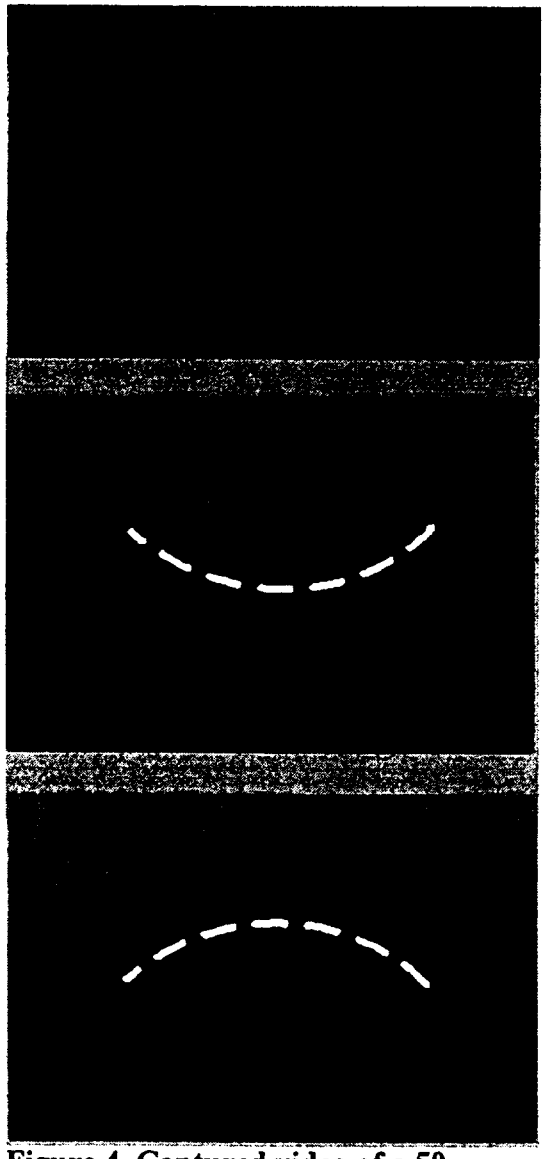

Figure 4 Captured video of a 50-micron Janus shard flipping over to achieve lowest free energy (side view) at a water-oil interface, with the oil on top. The flipping sequence between frames 2 and 3 occurred in less than 2 video frames (1/30 sec).

The silanated shards did interact with one another. In fact, the interactions cause the shards to pull together. These interactions pulled the shards to one another even when they are separated by 0.5 mm to 1.0 mm.

It was also observed that the shards surprisingly oriented themselves with the hydrophobic side up. This was confirmed by sprinkling the shards on a droplet of water and then evaporating the water. The sample was then examined under SEM.

Another unique feature of the shards is that they act as small convex lenses. With the shards on the surface of the water, the microscope can be focused on the shards and then focused on the light being transmitted through the shards, which formed focal spots above the liquid surface. The difference in

focus positions was approximately 100 μm. This focal length approximates the diameter of the microballoons proving that the shards were oriented with the curved side facing up (none were seen to focus on the opposite side of the water surface).

We tried to observe how the shards flip to be concave down on the water surface but the process was much too fast to see or record with our equipment. However, we were able to see the process by slowing down the flipping using a viscous overlying phase, as follows.

The two-phase mixing experiment allowed for the shards to be mixed with oil and water and then to separate based on density and hydrophobic-hydrophilic interactions (Figure 5). It was observed that no matter how the shards settle on the oil-water interface they quickly assume the lowest free energy orientation with hydrophilic side facing the water. The strength of this reorienting force is strong enough to rotate a 50-micron shard 180° within 0.05 sec even though the initial "grip" on the shard must be only at the edges.

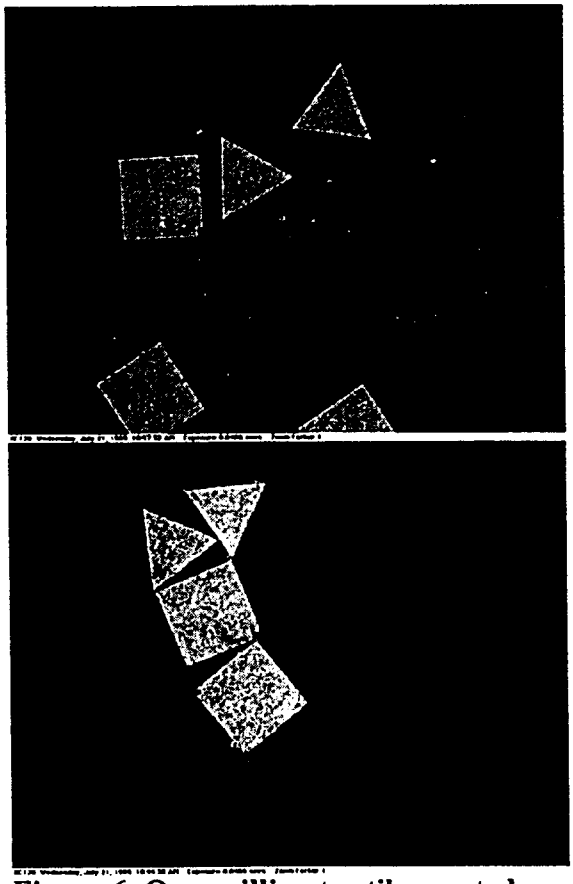

Figure 6 One-millimeter tiles coated with gold and a hydrophobic silane exhibit no lateral capillary interactions on an ethanol subphase (top) but strong corner-to-corner interactions on water (bottom).

II. Capillary Interactions between Janus Tiles

The alumina tiles made by LIGA molds were treated in various ways to tailor their aggregation characteristics and to effect specific binding forces between edges. While this study showed promising results, it is unfinished in that a general way to impart different affinities and binding forces from one surface to another was not found.

We aimed to have highly specific pairs bound to the tile edges, so that only edges that were programmed to stick together would do so. The surface available for this joining was (for the large tiles in Fig 3) 0.2 mm by 1.0 mm. For the necessary specificity we looked to biological adhesion molecules; we considered using virus head-tail pairs, biotin-avidin, and antigen-antibodies. Although the virus head-tail pairs are the most remarkable option (these are viral pieces called "packaging material" separately grown by recombinant techniques) and possibly the most specific, we chose to pursue biotin-avidin, which is among the strongest bonds in biology.

Our general strategy to program the tiles with two surface affinities was to treat all the surfaces first, then use line-of-sight processing to change selected surfaces. Thus, it is a decisive advantage to perform the line-of-sight treatments while the tiles are aligned on the substrate, as in Figure 3. Available line-of-sight techniques were lasers, x rays, physical vapor depositions such as evaporation or sputtering, and plasma cleanings.

We rejected x rays as being too penetrating, but found that none of the other techniques were completely effective. For example, we used an RF discharge in air to attempt to clean silane treatments from tile edges but the discharge was too isotropic even with baffles in the chamber, resulting in all the surfaces being fully cleaned (except for the base against the substrate). Although the mean-free path of reactive species in the plasma should have been several centimeters, it is likely that their excursions were much

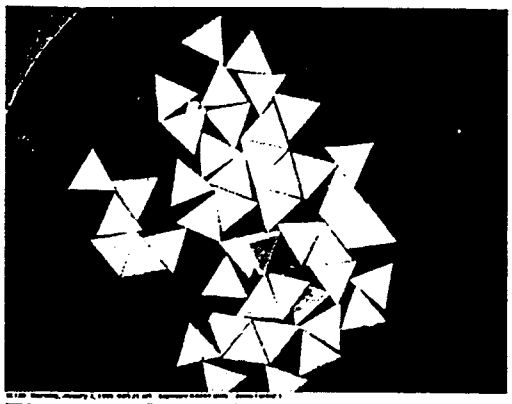

Figure 7 One-millimeter alumina triangles interacting on a water subphase. With no agitation, the aligning effects of corner-to-corner interactions can be seen.

shorter given the frequency of the driving field. A DC discharge may be more effective since it would be more line-of-sight.

Similarly, our effort to sputter-coat only certain surfaces with gold in order to bind receptor molecules using thiol chemistry resulted in trace gold deposits on the shadowed sides—enough to eliminate the affinity contrast we sought. Here evaporation over a long path (as opposed to sputtering) might be more effective.

While we obtained a photosensitive avidin for attempting optical programming methods, this technique and the refinement of the other techniques were not accomplished in the project period.

Nevertheless, using hydrophobically treated tiles, we performed simple experiments on the surface of a liquid subphase to check surfaces energies and tile-tile interactions. In Figure 6, hydrophobic tiles floated on water exhibit long-range corner-to-corner interactions while those on EtOH do not. These corner forces result from capillary hydrostatics, which in the limit of small deviations h(x,y) and slopes h_x for the liquid surface from its unperturbed height has the form[5]

$$\nabla^2 h(x,y) = -\tfrac{\rho g}{2\gamma} h(x,y)$$

which has strong cusps at corners of the form

$$h_r \approx r^{-1/3}$$

for squares and

$$h_r \approx r^{-2/5}$$

for triangles, where hr is the slope and r is the distance from the corner.[6] These "singularities" account for strong corner-to-corner interactions, which can be seen in Figure 7, showing an ensemble of hydrophobic triangular tiles on water. With no agitation, the triangles are seen to tile the plane densely in local regions because the corner-to-corner interactions tend to bring the tiles into registry. This happy circumstance can be exploited in future self-assembly experiments.

Conclusion

We have shown that simple interactions such as those with water can be applied and manipulated for the use of a Janus self-assembly process. A successful protocol was developed to make millimeter-size Janus tiles from microballoons and from LIGA molds. While the hydrophobic and hydrophilic interactions have proven successful for orienting these tiles, biological adhesion molecules offer the key to specific binding. Capillary hydrostatic interactions on a water subphase offer strong, long-range, self-orienting interactions for future studies.

Acknowledgment. This work was funded by Sandia National Labs (DOE contract No. DE-AC04-94AL85000).

References

(1) Cann, A. J. Principles of Molecular Virology second edition 1997 p. 23.
(2) Rothemund, P.W.K., Using lateral capillary forces to compute by self-assembly, Proc Nat Acad Sci 2000, 97, pp 984-989.
(3) LIGA Technology, http://daytona.ca. sandia.gov/LIGA/index1.html.
(4) Chopard, B.; Luthi, P.; Droz, M. Microscopic Approach to the Formation of Liesegang Patterns, J. Stat. Phys. 1994, 76, pp. 661-677.
(5) Adamson, H. T., *Statistical Mechanics of Phases, Interfaces, and Thin Films* (Wiley-VCH, New York, 1996), chapter 7, section 7.3, page 345.
(6) Jackson, J.D., *Classical Electrodynamics* (John Wiley & Sons, Inc, New York, Second Edition, 1975), chapter 2, section 11, page 77.

Part II
Energetic Tiles

1 Energetic tilings

In this report, we introduce an energetic tiling scheme. In this scheme, we will assume that the edges of our tiles sit on some lattice and interact by some local physical process for which there is an energy cost (positive or negative), dependent upon the number and types of contacts which are made.

1.1 Proposed definition of an energetic tiling

Let A be a set of polygons called *tiles*. Let $\Xi_{\mathcal{A}}$ be the set of configurations, which tilings of the plane with polygons from the set A.

An energy function can be defined for $x \in \Xi_{\mathcal{A}}$ by the following

$$E(x) = \sum_{i,j} n_{ij}\epsilon_{ij},$$

where $n_{ij}(x)$ is the number of tiles, per unit volume, having face i which makes contact with a face j (note: $n_{ij} = n_{ji}$). Since n_{ij} is symmetric, we may take $\epsilon_{ij} = \epsilon_{ji}$. We may call ϵ the *interaction energy*.

If $\mathcal{A}$ contains a tile h such that for any $t \in \mathcal{A}$, $\epsilon_{ij} = \epsilon_{ti'j}$ where i, ti are faces of h and j, tj are faces of t, then we call h a *hole*. It is not difficult to show that (up to a constant) we may rewrite the energy as

$$E(x) = \sum_{i,j} n_{ij}\epsilon_{ij} + \mu_t\rho_t,$$

where $\epsilon_{ij} = 0$ when i (or j) is a face of h, and ρ_t is the number, per unit volume, of tile $t \in \mathcal{A} - h$. This is simply a formal way to capture the notion of a gap in the tiling by distinguishing one tile as a hole and assigning to the non-hole tiles a *chemical potential*, μ_t.

There is some question whether the tiling statistics, ρ_t and n_{ij} are well-defined for every configuration x. We would like to ignore that question for now.

1.2 Tiling statistics

Let $\mathcal{S}_{\mathcal{A}}$ be the set of valid (n_{ij}), that is, those (n_{ij}) such that for any δ there exists a configuration $x \in \Xi_{\mathcal{A}}$ with tiling statistics within $O(\delta)$ of (n_{ij}). This set is bounded and closed, and therefore compact. Additionally, it can be proven that $\mathcal{S}_{\mathcal{A}}$ is convex.

Thus, the energetic tiling problem is one of maximizing a cost function on the convex body $\mathcal{S}_{\mathcal{A}}$. A study of this set is therefore essential. As of this writing we have no general theory which further describes $\mathcal{S}_{\mathcal{A}}$ all but the most trivial $\mathcal{A}$.

If $\mathcal{S}_{\mathcal{A}}$ is a polytope, then it has only a finite number of vertices, at least one of which will maximize any given energy. Additionally, all the extreme points of $\mathcal{S}_{\mathcal{A}}$ are isolated, i.e. the various optimizing configurations are stable to perturbations of the energy function.

If $\mathcal{S}_{\mathcal{A}}$ is not a polytope, at least part of the boundary of $\mathcal{S}_{\mathcal{A}}$ must be a non-linear constraint. This divides the extreme points of $\mathcal{S}_{\mathcal{A}}$ into two classes: the isolated extreme points, and the continuous extreme points, or "sensitive" and "insensitive" extreme points (a finite number of the former and an infinite number of the latter).

One can compute an approximate $\mathcal{S}_{\mathcal{A}}$, assuming one has a tool to produce approximately optimal tilings for any given energy function. A heuristic procedure to "flesh out" $\mathcal{S}_{\mathcal{A}}$ would be
• generate a set of statistics from finite (periodic) tilings which optimize randomly generated energy functions.
• Take the convex hull of the statistics. The vertices of this hull are approximations to the extreme points of $\mathcal{S}_{\mathcal{A}}$.
• Use the inequalities of the convex hull as energy functions and attempt to generate configurations with statistics outside the the current convex hull.
• Add these new points to the current set of vertices and take the convex hull again.

Repeating the last two steps should result in one of two outcomes. If $\mathcal{S}_{\mathcal{A}}$ is a polytope, then eventually all the extreme points of $\mathcal{S}_{\mathcal{A}}$ will lie in the current convex hull, and that current convex hull will be $\mathcal{S}_{\mathcal{A}}$. If $\mathcal{S}_{\mathcal{A}}$ is not a polytope, then the nonlinear part of the boundary will always contain extreme points not in the present convex hull, and the process will not terminate. The unfortunate part of this heuristic is that it does not provide a means of deciding is $\mathcal{S}_{\mathcal{A}}$ is a polytope or not, nor when one has found it, if it is.

1.3 Triangular tiling

We look at some simple energetic tiling schemes with triangles. Let $\mathcal{A}_0$ be the set consisting of a single equilateral triangle with distinct edges. Let $\mathcal{A}_1$ be a set consisting of two triangles, one of which is considered a hole.

1.3.1 Energetic tiling with $\mathcal{A}_0$

The statistics set, $\mathcal{S}_{\mathcal{A}_0} = \mathcal{S}_0$, is the polytope ($n_{ij} \geq 0$, and $\sum_j n_{ij} = 1, \forall i$) with vertices realized by the three tilings, (up to rotation) shown (as unit cells) in figures 1.3.1—1.3.1

1.3.2 Energetic tiling with $\mathcal{A}_1$

As of this writing, it has not been determined whether the statistics set, $\mathcal{S}_{\mathcal{A}_1} = \mathcal{S}_1$ is a polytope or not. Thus we cannot present a complete set of vertices in this section. Two easily proven isolated vertices are shown in figures 1.3.2 and 1.3.2 (using gray triangles to represent holes). Note: the all-hole tiling ($\rho = 0$) and all the vertices of $\mathcal{S}_0$ are also isolated vertices of $\mathcal{S}_1$.

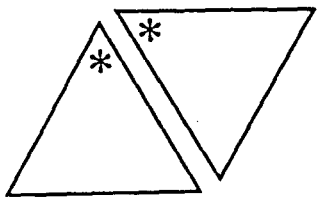

Figure 1: Vertex of $\mathcal{S}_0$, with $n_{ij} = \begin{pmatrix} 0 & \frac{1}{2} & \frac{1}{2} \\ \frac{1}{2} & 0 & \frac{1}{2} \\ \frac{1}{2} & \frac{1}{2} & 0 \end{pmatrix}$

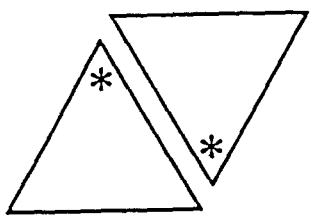

Figure 2: Vertex of $\mathcal{S}_0$, with $n_{ij} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}$

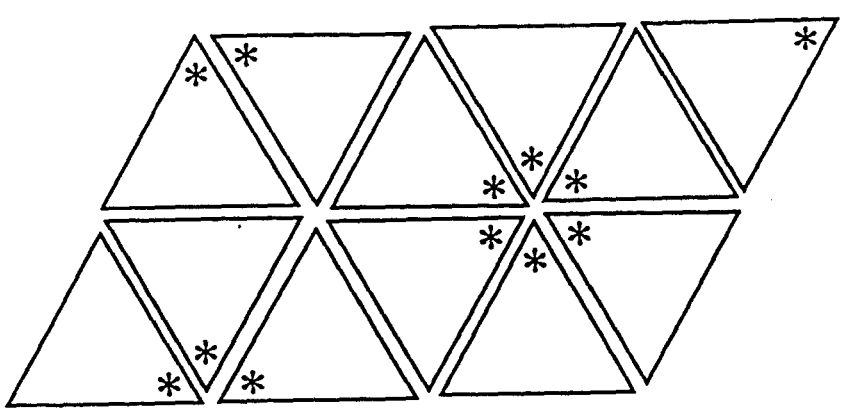

Figure 3: Vertex of $\mathcal{S}_0$, with $n_{ij} = \begin{pmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \end{pmatrix}$

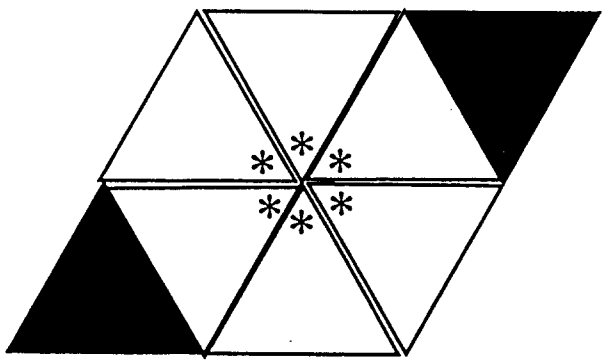

Figure 4: Known vertex of $\mathcal{S}_1$, with $\rho = \frac{3}{4}, n_{ij} = \begin{pmatrix} 0 & \frac{3}{4} & 0 \\ \frac{3}{4} & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}$

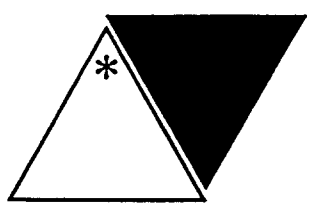

Figure 5: Known vertex of $\mathcal{S}_1$, with $\rho = \frac{1}{2}, n_{ij} = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}$

We have performed a series of numerical experiments to try and isolate the rest of the vertices. These experiments strongly suggest that three additional constraints of the form

$$n_{11} + n_{22} - n_{33} + n_{12} - n_{23} - n_{13} \leq 1 \tag{1}$$
$$n_{11} - n_{22} + n_{33} - n_{12} - n_{23} + n_{13} \leq 1 \tag{2}$$
$$-n_{11} + n_{22} + n_{33} - n_{12} + n_{23} - n_{13} \leq 1 \tag{3}$$
$$\tag{4}$$

are obeyed by the elements of $\mathcal{S}_1$. If we assume these constraints (in addition to $\rho, n_{ij} \geq 0$, and $\sum_j n_{ij} = \rho, \forall i$), we find several new isolated vertices (figures 1.3.2—1.3.2).

It is also true that, if we assume that Eq's 1—3 are true, we require additional constraints, because taking 1—3 and the trivial constraints alone yields a polytope with the vertex

$$\rho = \frac{3}{5}, n_{ij} = \begin{pmatrix} \frac{3}{5} & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix},$$

which is a set of statistics not realizable by any tiling that we have been able to create, thus we know there are additional constraints. There also may be non-linear constraints, but, as of this writing, we have no conjectures about them.

Part III

Energetic Dominos

2 Energetic dominos

In this chapter we derive constraints on the interactions of hydrophobic contact between bipoles in the cubic and face-centered cubic lattices. Although we do not prove the tightness of the inequalities we are able to derive lower bounds that guarantee a predictable degree of approximation. This allows us to prove that the biplane self-assembly is within a small percent of optimal.

3 Definitions and Notation

We denote by $\mathcal{F}$ a figure composed of hydrophobic residues, on a lattice. We denote by $\mathcal{G}$ a figure composed of bipoles (a hydrophobic/hydrophilic pair) on a lattice. Similarly for $\mathcal{F}_i, \mathcal{F}', \mathcal{G}_i, \mathcal{G}'$, etc.

We denote by $N(\mathcal{F})$ or $N(\mathcal{G})$ the number of hydrophobic residues in $\mathcal{F}$ or $\mathcal{G}$, respectively. When there is no ambiguity, we will denote this quantity with N. Similarly we denote with

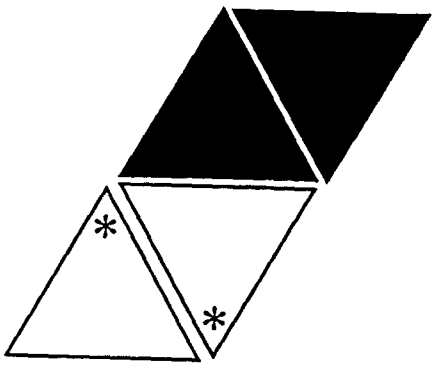

Figure 6: Probable vertex of $\mathcal{S}_1$, with $\rho = \frac{1}{2}, n_{ij} = \begin{pmatrix} \frac{1}{2} & 0 & 0 \\ 0 & \frac{1}{2} & 0 \\ 0 & 0 & 0 \end{pmatrix}$

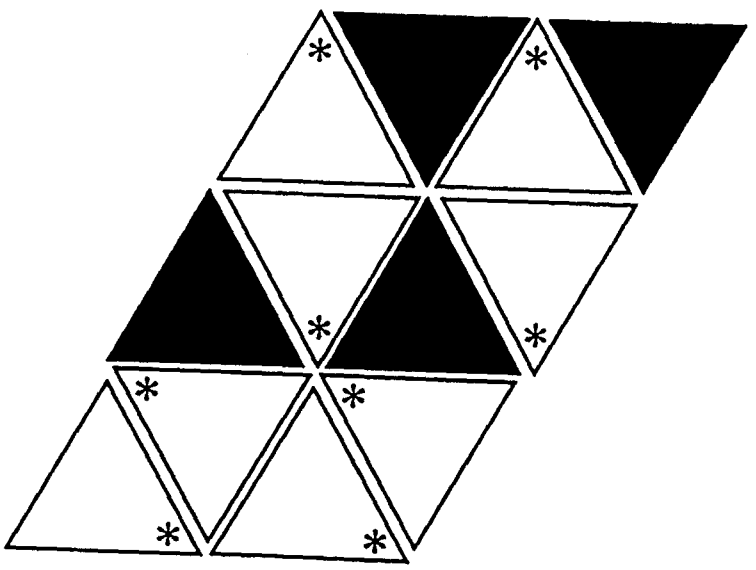

Figure 7: Probable vertex of $\mathcal{S}_1$, with $\rho = \frac{2}{3}, n_{ij} = \begin{pmatrix} \frac{1}{3} & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & \frac{2}{3} \end{pmatrix}$

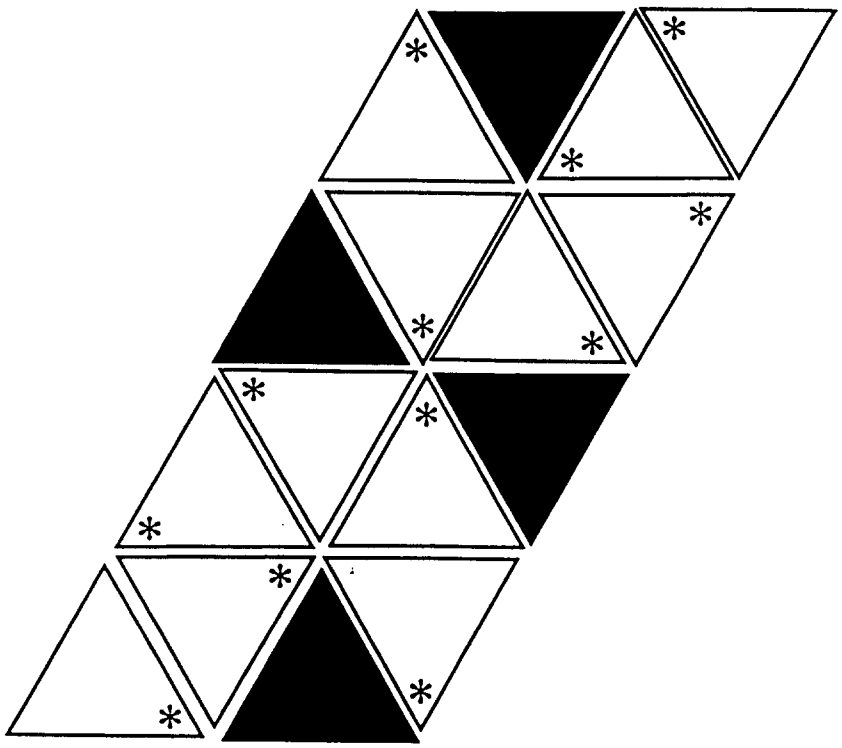

Figure 8: Probable vertex of $\mathcal{S}_1$, with $\rho = \frac{3}{4}, n_{ij} = \begin{pmatrix} \frac{1}{4} & 0 & \frac{1}{2} \\ 0 & 0 & 0 \\ \frac{1}{2} & 0 & \frac{1}{4} \end{pmatrix}$

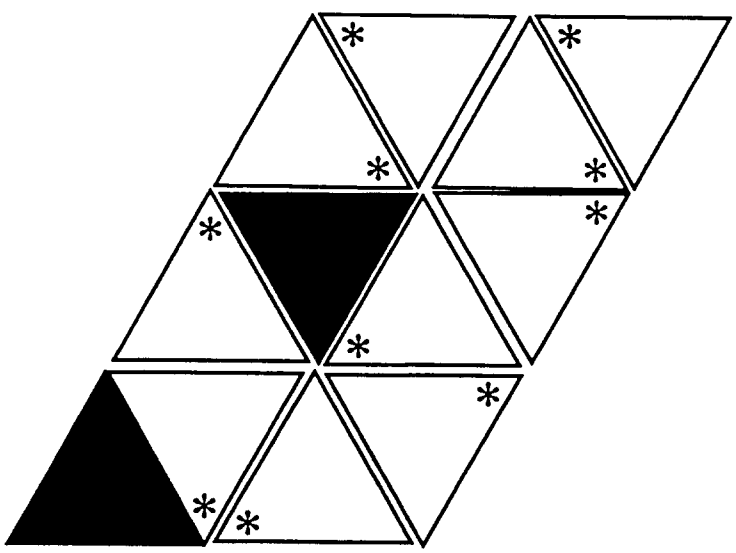

Figure 9: Probable vertex of $\mathcal{S}_1$, with $\rho = \frac{5}{6}, n_{ij} = \begin{pmatrix} \frac{1}{2} & \frac{1}{3} & 0 \\ \frac{1}{3} & 0 & 0 \\ 0 & 0 & \frac{5}{6} \end{pmatrix}$

H, L, W (or $H(\mathcal{F}), L(\mathcal{F}), \ldots, W(\mathcal{G})$ when needed) the dimensions Height, Length, and Width of $\mathcal{F}$ or $\mathcal{G}$on the cubic lattice. Without loss of generality we let $H \geq L \geq W$. We give directions to the cubic lattice, namely $+z, -z, +y, -y, +x, -x$, corresponding to H, L, W in the obvious way. A figure on the FCC lattice can be partitioned in four different ways so that the slices are 2D triangular lattices. We denote with H, L, W_1, W_2 the four dimensions of $\mathcal{F}$ or $\mathcal{G}$, and let wlog $H \geq L \geq W_1 \geq W_2$. We use the corresponding direction names z, z', y, y', x, x' for the directions $+z, -z, +y, -y, +x, -x$ respectively.

A figure $\mathcal{F}$ can be decomposed into H *levels* of sizes $h_1, \ldots, h_H$ so that $N = h_1 + \cdots + h_H$. We denote by L_i, W_i the dimensions of the ith level of $\mathcal{F}$ in the cubic lattice.

We showed how to decompose a figure $\mathcal{F}$ on the square lattice into levels, such that $N(\mathcal{F}) = h_1 + \cdots + h_H$. Alternatively we can decompose $\mathcal{F}$ with respect to the other two dimensions, namely with respect to the x- or y- axes. We define thus the H-, L-, and W-partitions of a figure $\mathcal{F}$. The H-partition partitions $\mathcal{F}$ into *H-slices*, or levels, as we showed above. We denote the ith H-slice by $\mathcal{H}_i$, and its size by $|\mathcal{H}_i| = l_i$. The H-slice $\mathcal{H}_i$ has dimensions length and width, that we denote by $L(\mathcal{H}_i)$ and $W(\mathcal{H}_i)$. The L-partition similarly partitions $\mathcal{F}$ into *L-slices*. We denote the ith L-slice by $\mathcal{L}_i$, and its size by $|\mathcal{L}_i| = l_i$. We define its dimensions $H(\mathcal{L}_i)$ and $W(\mathcal{L}_i)$. Similarly we define the W-partition, the W-slices $\mathcal{W}_i$, their sizes $|\mathcal{W}_i| = w_i$, and their dimensions $H(\mathcal{W}_i), L(\mathcal{W}_i)$. In the FCC lattice we make corresponding definitions. There, we denote the W_j-slices ($j = 1, 2$) by $\mathcal{W}_{j,i}$ and their sizes by $|\mathcal{W}_{j,i}| = w_{j,i}$.

We say that a perimeter residue of level i, in the cubic lattice is a *corner residue* if it has exposed faces on both the x- and y-directions. Notice that in a 3D figure on the cubic lattice there are 12 different types of corners. We denote a corner using the two defining faces of it. Thus we define $(x'y)$-, $(x'y')$-, (xz)-, ..., $(y'z')$-corners. Then, a *corner residue* of type (uv) is a residue that has exposed faces on directions u and v where $u \neq -v$.

4 Bipole Contact Upper Bounds on the Cubic and FCC lattices

We would like to obtain a good upper bound on the number of H-H contacts possible when N bipoles self-assemble on the cubic, or FCC lattice. Denote the number of contacts by $[HH](N)$, or simply $[HH]$. A way to obtain an upper bound on the number of contacts possible is to observe that each residue can only make as many contacts as the number of its neighbors, minus one for its polar part. That gives 5 contacts on the cubic lattice, and 11 contacts on the FCC lattice. Therefore we obtain the following inequalities, for a configuration of size N:

$$[HH] \leq 5N \tag{1.CUBIC}$$
$$[HH] \leq 11N \tag{1.FCC}$$

In order to give tighter upper bounds, we introduce some more notation. Let $[HH]$ be the number of H-H contacts, $[HP]$ be the number of H-P contacts excluding the bipole edge, and

$[HW]$ be the number of H-Water contacts, i.e. contacts between hydrophobic residues and empty points in the lattice. Then clearly the following equations hold:

$$[HH] + [HP] \qquad +[HW] = 5N \qquad \text{(2.CUBIC)}$$
$$[HH] + [HP] \qquad +[HW] = 11N \qquad \text{(2.FCC)}$$

We now introduce the concept of a "conflict" contact [REF Agarwala]. Let (H_1, P_1) and (H_2, P_2) be two bipoles. Let there be a contact (H_1, H_2). This is a *conflict* H-H contact for H_1 if and only if there is a contact (H_2, P_1). Similarly it is a conflict contact for H_2 if and only if there is a contact (H_1, P_2). Call (H_1, H_2) a conflict contact if it is a conflict contact for at least one of H_1, H_2. In [REF Agarwala] it is shown that on the FCC lattice every H-residue that has a P-residue as a neighbor in an $\{H, P\}$-sequence, can at most 7 non-conflict H-H contacts and 4 conflict H-H contacts.

Denote by $[HHc]$ the conflict H-H contacts, and by $[HHe]$ the non-conflict H-H contacts. Then $[HH] = [HHe] + [HHc]$. In the cubic lattice $[HHc]$ always equals 0, but in the FCC lattice we have $[HHe] \leq 7N$ and therefore

$$[HHc] \geq [HH] - 7N \qquad \text{(3.FCC)}$$

Finally define similarly a "conflict" H-P contact. Let (H_1, P_1) and (H_2, P_2) be two bipoles, and (H_1, P_2) be a H-P contact. This is a *conflict* H-P contact if and only if (H_1, H_2) is also a contact. Clearly every H-H conflict contact corresponds 1-1 to a H-P conflict contact (we are implicitly assuming that the contact (H, H') is different from the contact (H', H)). Therefore,

$$[HP] = [HPc] + [HPe] = [HHc] + [HPe] \geq [HH] - 7N + [HPe] \qquad \text{(4.FCC)}$$

Now we can prove a tighter upper bound on the number of H-H contacts for the FCC lattice, and also demonstrate an idea that we will use to prove a tighter bound for the cubic lattice. For the FCC lattice, expand the equation (2.FCC) to get:

$$\begin{aligned}
& [HHe] + [HHc] + [HPe] + [HPc] + [HW] = 11N && \text{(from (2.FCC))} \\
\Longrightarrow\ & [HHe] + [HHc] \leq 11N - [HPe] - [HPc] - [HW] \\
\Longrightarrow\ & 2[HHc] \leq 4N - ([HPe] + [HW]) && \text{(from (3.FCC), (4.FCC))} \\
\Longrightarrow\ & [HHc] \leq 2N - \frac{1}{2}([HPe] + [HW]) \\
\Longrightarrow\ & [HH] \leq 9N - \frac{1}{2}([HPg] + [Hw])
\end{aligned}$$

In the cubic lattice the following corresponding inequality holds:

$$[HP] \leq 5N - ([HP] + [HW])$$

Therefore we have proven the following result:

Theorem 1 *The following upper bounds on the number of H-H contact edges for a conformation of N bipoles on the Cubic, and FCC lattices, hold:*

$$[HH] \leq 2.5 \times N - 0.5 \times ([HP] + [HW]) \quad \text{(5.CUBIC)}$$
$$[HH] \leq 4.5 \times N - 0.25 \times ([HPg] + [HW]) \quad \text{(5.FCC)}$$

5 Stronger Bounds for the Cubic Lattice

Using (5.CUBIC) we can derive a stronger upper bound on the number of H-H contacts of N bipoles in the Cubic lattice, by deriving a lower bound on the number of H-P or H-Water contacts. Intuitively, a place where a figure $\mathcal{G}$ should miss a certain number of contacts, is the edges and corners in the exterior of the embedded figure $\mathcal{F}$ consisting of the hydrophobic residues of $\mathcal{G}$. For example in the cubic lattice it is easy to see that any figure $\mathcal{F}$ has at least 8 corner residues and each one of them makes at most 3 H-H contacts, at least 2 H-Water contacts, and exactly one contact to its polar part. Moreover, every "level" or H-slice of the figure has at least 4 "corners", namely at least one H-residue that has no H-residue neighbors on the $(+x, +y)$ directions, one that has no H-neighbors on the $(+x, -y)$ directions, one that has no H-neighbors on the $(-x, +y)$ directions, and one that has no H-neighbors on the $(-x, -y)$-directions.

For every bipole conformation $\mathcal{G}$ there is a corresponding H-residue conformation $\mathcal{F}$ consisting of the hydrophobic residues of $\mathcal{G}$. Every H-residue in $\mathcal{G}$ has 5 free neighboring positions for either H-H, H-P, or H-W contacts. Every H-residue in $\mathcal{F}$ that contacts at most 4 other H-residues, corresponds to a residue in $\mathcal{G}$ that has at least one H-W or H-P contact. Therefore every residue in $\mathcal{F}$ that has more exposed faces than 1, contributes at least the number of exposed faces minus 1 to the sum $[HW] + [HP]$. That is, the exposed "area" of a figure $\mathcal{F}$, $A(\mathcal{F})$, minus the number of exposed faces of $\mathcal{F}$, $E(\mathcal{F})$ is clearly a lower bound on $[HP]+[HW]$. Of course if $\mathcal{F}$ corresponds to a bipole conformation $\mathcal{G}$ then $E(\mathcal{F}) = N(\mathcal{F})$. On the other hand, if we obtain a lower bound on $A(\mathcal{F}) - E(\mathcal{F})$ for any conformation $\mathcal{F}$ of N H-residues, that would clearly be a lower bound for $[HP] + [HW]$ for any bipole conformation $\mathcal{G}$ of size N.

Let $\mathcal{F}$ be a figure of H-residues. As we defined above, H, L, W are the dimensions height, length, and width of $\mathcal{F}$ respectively. Assume wlog that $H \geq W \geq L$. Define H_a to be the number of H-slices that have L-dimension ≥ 2. Similarly define L_a to be the number of L-slices that have W-dimension ≥ 2 and W_a the number of W-slices that have L-dimension ≥ 2. Then denote by H_b, L_b, W_b to be the number of H-, L-, and W-slices respectively that do not satisfy this condition. In order to prove a lower bound on $A(\mathcal{F}) - E(\mathcal{F})$, we will identify residues that have at least 2 exposed faces. For each such residue, we will call one of its exposed faces the

bipole face. Intuitively, such a face wlog would connect to the polar part of the corresponding bipole. Mathematically, assigning polar faces will help us give a lower bound on $\mathsf{A} - \mathsf{E}$: we can count as contributions to this quantity all the faces of a particular residue, except its polar face. Those additional faces we will also call *water/polar faces* or w/p faces. Then $\mathsf{A} - \mathsf{E}$ is at least as large as the number of w/p faces we can count.

First consider the H-partition of $\mathcal{F}$. Take an arbitrary H-slice $\mathcal{H}_i$.

- If $W(\mathcal{H}_i) = 1$ then $\mathcal{H}_i$ is either a single cube, or a column of cubes with possible gaps. In both cases we can demonstrate 2 w/p-faces on the x, x' directions by placing all the polar faces of residues in $\mathcal{H}_i$ in the y-direction as shown in figure [REF FIGURE].

- if $W(\mathcal{H}_i) = 2$ then we can similarly count 4 w/p-faces on the x, x' directions: 2 of them in the "row" of $\mathcal{H}_i$ with the highest y-coordinate (placing polar faces in the y direction), and two of them in the "row" with the lowest y-dimension (placing polar faces in the y' direction. Figure [REF FIG] demonstrates that.

Therefore we can count a total of $4H_a + 2H_b$ w/p-faces, all of them assigned in the directions x, x'.

Next we consider the W-partition of $\mathcal{F}$. We will be careful to count contacts that are independent of the ones we counted in the H-partition so as to avoid double counting.

Let $\mathcal{W}_i$ be a W-slice in this partition. Similarly as before we can count w/p-faces in the z, z' directions. Figure [REF FIG] demonstrates that. The polar faces we are assigning for the W-partition are in the y, y' directions, as for the H-partition. The w/p-faces we are counting for the W-partition are in the z, z' directions, while they are in the x, x' for the H-partition. Therefore the numbers of w/p-faces counted for the H- and W-partitions add up. We count a total of $4W_a + 2W_b$ w/p-faces for the W-partition.

¿From the above we obtain lower bound on $[HP] + [HW]$ for the Cubic lattice:

$$[HP] + [HW] \geq 4(H_a + W_a) + 2(H_b + W_b)$$

Because the N residues of $\mathcal{F}$ fit in the above dimensions, we immediately obtain the following *volume constraints*:

$$H_a \times W_a \times L + H_b \times W + H \times W_b \geq N$$

where $H \geq W \geq L$. The above formula can be used for any specific N, in order to give tighter lower bounds on the number of contacts achieved by N bipoles on the Cubic lattice.

Theorem 2 *Let $\mathcal{F}$ be a figure of N bipoles on the Cubic lattice, with dimensions $H \geq W \geq L$ and such that $H = H_a + H_b$ and $W = W_a + W_b$, where H_a, H_b, W_a, W_b all non-negative. Then*

$$[HH] \leq 5N - 4(H_a + W_a) + 2(H_b + W_b) \qquad \text{(H-H.UB)}$$

where

$$H_a \times W_a \times L + H_b \times W + H \times W_b \geq N \qquad \text{(VOL)}$$

The above costraints can readily be used to calculate lower bounds on the number of contacts missed for a configuration of N bipoles. For N between 50 and 100, we performed a simple calculation using *Matlab* and obtained the following results:

For $N = 50$, $[HH] \leq 5 \times 50 - 28$. That is, number of contacts missed is ≥ 28.
For $51 \leq N \leq 55$, number of contacts missed ≥ 30.
For $56 \leq N \leq 75$, number of contacts missed ≥ 32.
For $76 \leq N \leq 80$, number of contacts missed ≥ 34.
For $81 \leq N \leq 100$, number of contacts missed ≥ 36.

Biplane Configuration. Consider for instance the biplane configuration for N bipoles. According to this configuration the bipoles are placed such that the hydrophobic residues form two opposing rectangles that are of sizes equal, or differing by one. The dimensions of the rectangles are the closest to square. That is, given $N = 2k + 1$ the two rectangles (call them *top* and *bottom*) have sizes k, and $k + 1$ respectively. Then, for size k find minimum l such that $l * (l - 1) \geq k$, or $l * l \geq k$. Then the top rectangle has dimensions $l, l - 1$, or l, l, respectively. Similarly for the bottom rectangle. It is a trivial exercise to place these two rectangles of hydrophobic residues against each other so as to maximize hydrophobic contacts.

These lower bounds can be used to guarantee an approximation ratio for the number of contacts achieved by the biplane configuration. For $N = 50$ for instance, this configuration achieves a total of $5 \times 50 - 40 = 210$ contacts (double counted as above). This is easy to calculate, and we omit the calculation. Compared with the upper bound of 222 contacts, the biplane configuration is guaranteed to be within .9459 of optimal. For $N = 100$, the biplane achieves $5 \times 100 - 60 = 440$ contacts, similarly providing a $440/464 = .9483$-approximation.

Part IV
TORTILLA Protein Folding Software

6 The TORTILLA Protein Folding Package

The Tortilla Software Package is a graphical UNIX/Motif tool for designing protein-related algorithms and experiments. Tortilla was designed with flexibility and expandability in mind. It is fully object-oriented and uses state-of-the-art software patterns. Additional algorithms and data types may be added by the end user at compile time, at link time and at run time.

Tortilla features many built-in capabilities, including:

- Support for any crystallographic lattice. Built in definitions for 2D/3D cubic, face-centered cubic, body-centered cubic, diamond, and hexagonal close packed.
- A conformation editor (for on-lattice structures) – to create or refine an on-lattice protein conformation (Figure 10).
- A conformation display ("3d" interactive graphical display) as either ball-and-stick, space-filling spheres, peptide planes, or as a cartoon "ribbon".
- Ab-initio on-lattice protein folding algorithms.
- Protein structure comparison.
- Contact map display.
- Protein sequence alignment.
- On-lattice energy landscape experiments, including determining the lowest energy conformation of a single sequence and determining the set of sequences with a single lowest energy conformation from all sequences of a specified length.

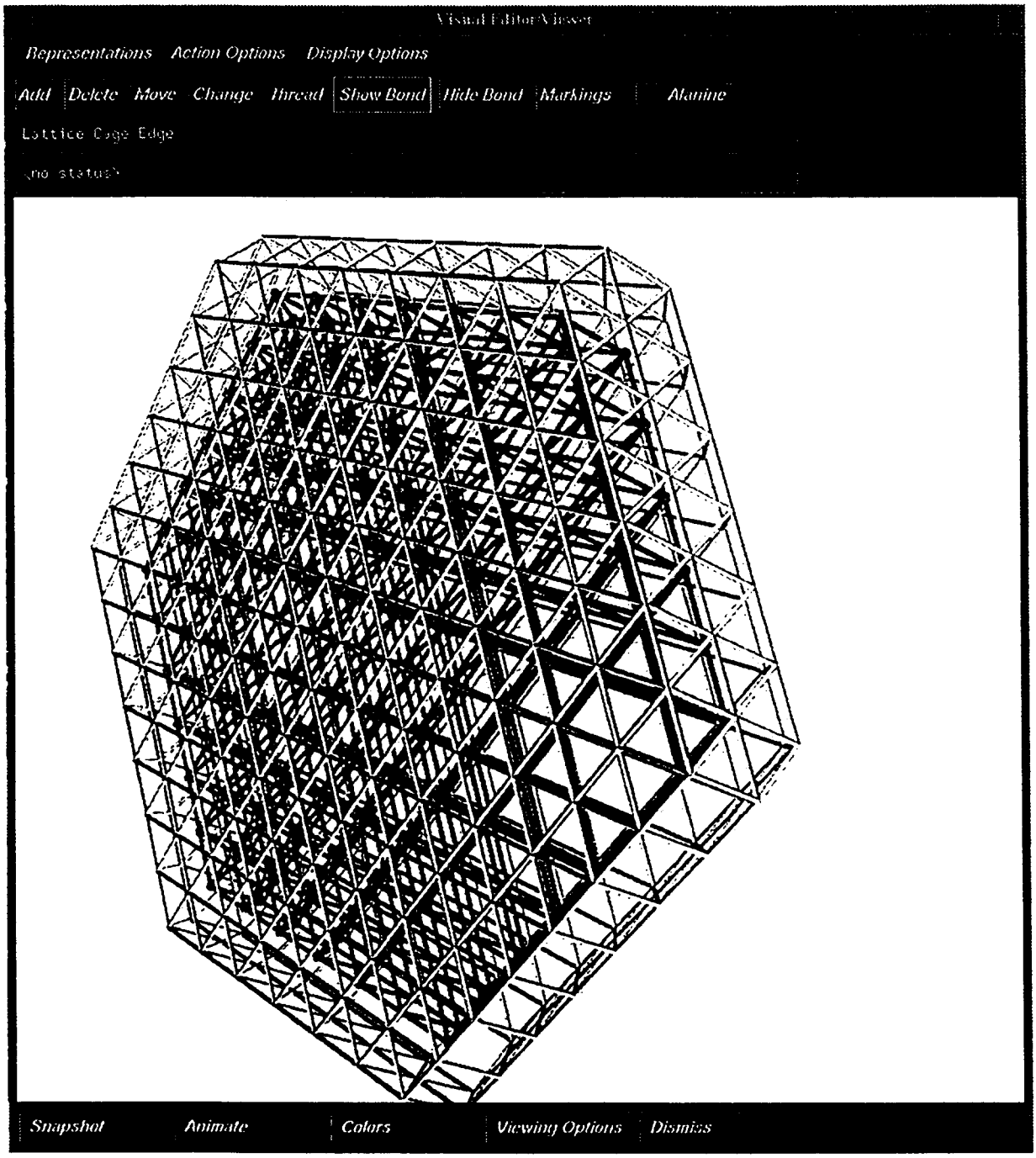

Figure 10: The conformation editor with a biplane-conjuecture structure.