A Survey of Computational Models of Self-Assembly

Algorithmic self-assembly is a mechanism used in all designs of DNA-based computers and based on the self-assembling nature of DNA by hybridization. In this talk, we will present an introduction to the field of algorithmic self-assembly that was largely pioneered by Winfree in 1998. Specifically, we will to motivate the study of the field of algorithmic self-assembly, provide descriptions of the Abstract Tile Assembly Model (aTAM) and the Kinetic Tile Assembly Model (kTAM), and give a simulated demonstration of computation using Winfree et al.'s Xgrow software.

Proof That Theoretical Computer Science Isn't Dead

Report on Erik Winfree's 1998 dissertation, "Algorithmic Self-Assembly of DNA"

COT6410 Spring 2015

Abstract

Algorithmic self-assembly is a mechanism used in all designs of DNA-based computers and based on the self-assembling nature of DNA by hybridization. In this short report, we seek to do the following: (1) motivate the study of self-assembly, (2) summarize some of the key contributions of Winfree's 1998 dissertation on computational self-assembly, (3) give our take on the challenges of presenting this work to a general computer science audience, and (4) present some of our speculations on the future direction of this work. We affectionately refer to our report as "Proof that Theoretical Computer Science isn't Dead."

1 Introduction

Time and space complexity boundaries have presented limitations to the capabilities of traditional computers since their introduction, motivating the exploration of alternative computational models. While the most popularly known work on such an alternative computational model is quantum computing, in this report we focus our attention on DNA computing, pioneered by Adleman in 1994 [1]. DNA sequences can be utilized to encode information such as numbers and answers to mathematical problems; Adleman was able to show that DNA can be used to solve the seven-point Hamiltonian path problem.

Perhaps to the dismay of early speculators of using DNA computation to solve NP-complete problems, DNA computation still fails to fundamentally move the boundary of computational intractability. DNA molecules still take up finite amounts of space, and exponential space solutions can quickly exceed the magnitude of the number of atoms in the universe—certainly larger than the number of possible usable DNA molecules for a solution.

The rest of the report is organized as follows: *Section 2: Motivation* will provide the reasoning behind the exploration of DNA self-assembly with relation to computation. *Section 3: Summary* will provide a brief overview of Winfree's contributions to the study of DNA computation in his 1998 dissertation. *Section 4: Presentation Challenges* will detail some of the creative efforts we had to undertake to present this work in a way palatable by a general computer science student audience. *Section 5: Speculations on Future Directions* will present some of our hypotheses on what the future may hold for the field of DNA computation as outsiders to the research field.

2 Motivation

DNA computation may prove to have several notable advantages over traditional computers for certain applications. Unlike traditional computers, DNA computers have extremely low levels of power consumption, high information density on the order of 10^7 gigabits per square inch, and high parallelization potential due to having many molecules of DNA to try different solutions at once [2] [3].

Beyond simply being able to theoretically encode information in DNA molecules, several problems need to be solved in order to make DNA computation practically feasible. Some major subproblems in 1998 were showing that DNA computation was theoretically universal, demonstrating that DNA computation was a priori plausible when factoring in certain important kinetics, and that the basic mechanisms behind DNA computation were experimentally plausible. Winfree provided proof of theoretic computation universality of a restricted set of DNA computation and provided affirmative positive evidence of the ability to overcome the latter two issues in his dissertation.

3 Summary

Winfree has made a number of contributions to the field of DNA computation [4], however we focus on his landmark dissertation publication in 1998 [2] where he showed that ligation and annealing are sufficient mechanisms for universal computation, and that two abstract models based on Wang computation could be used to understand DNA computation in a simulated environment.

Demonstrating that ligation and annealing alone can theoretically be used for universal computation is the principle conclusion of Winfree's dissertation. Winfree showed this by proving that a restricted class (and thus the whole class) of two-dimensional self-assembly is equivalent to one-dimensional cellular automata, which is known to be capable of universal computation. Two-dimensional self-assembly is done via a class of molecule called a double-crossover (DX), which can be theoretically codified into rules and then assembled into a lattice without outside interaction. An example of DX construction and structure can be seen in Figure 1, reproduced from figure 3.1.H of Winfree's dissertation [2]. An example of DX molecules being assembled into a lattice structure can be seen in Figure 2, reproduced from figure 3.14 of the same paper [2].

A notable shortcoming of Winfree's results was that universal computation via two-dimensional self-assembly was not demonstrated experimentally. However, some

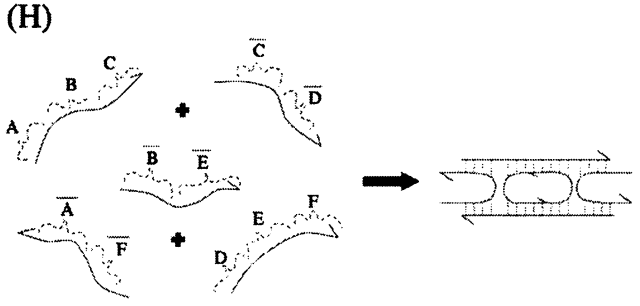

Figure 1: Double Crossover (DX) DNA construction. Reproduced from figure 3.1.H of [2]

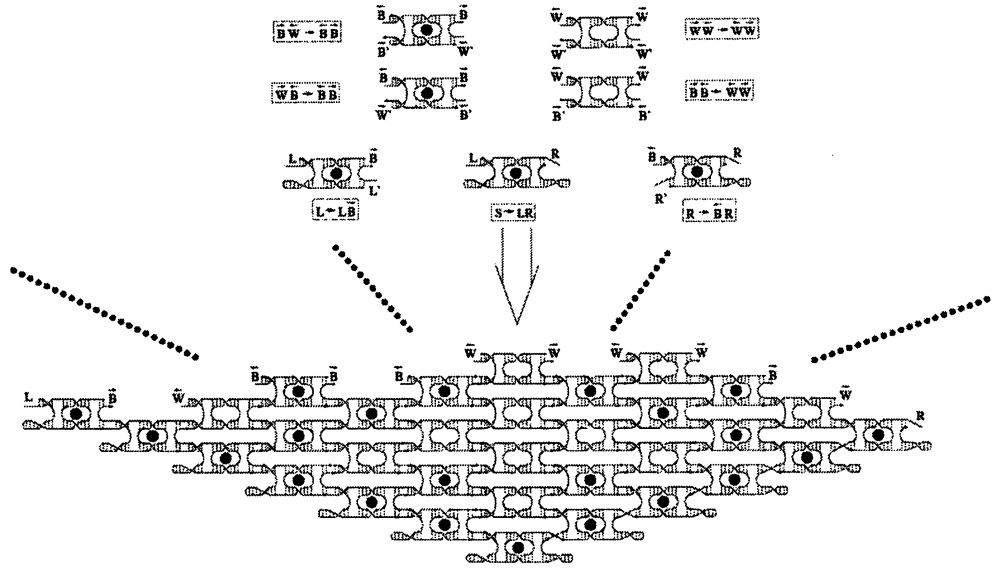

Figure 2: An algorithmic pattern in a self-assembled lattice. Reproduced from figure 3.14 of [2]

preliminary steps towards self-assembly of DX molecules as a lattice were explored. Winfree noted two practical subproblems that remained in the way of real-world lattice computation: (1) whether the geometric structure of the lattice will indeed self-assemble at all, and (2) whether the correct DX molecules would hybridize into the correct parts of the lattice with sufficient accuracy (e.g. whether there would be a high rate of units hybridizing into the wrong spots on the lattice, creating propagated errors in the computation as the lattice grew). Winfree explores the latter issue by doing lab experiments of a simplest case—whether a single DX molecule would hybridize into the correct spot of a simple structure. Winfree's experimental results showed this to be possible.

Winfree introduced two tiling models for abstractly analyzing the factors that lead to successful or unsuccessful latticing by DNA molecules. The first of these is Abstract Tile Assembly Model (aTAM), originally simply called Tile Assembly Model in his dissertation. Although we stray from providing a full formal description of aTAM in this summary, the model consists of the following:

- Tiles that correspond to DX molecules
- Four edge labels per tile that correspond to the four sticky ends of DX molecules (see figure 3.1.H. of Winfree's dissertation for figure of the four ends)
- Strengths of edge labels that correspond to, for instance, how long sticky ends of DNA molecules are
- A temperature parameter that influences the propensity for tiles to associate and bind. High temperatures, for instance, make a structure unstable

The second model introduced is called Kinetic Tile Assembly Model (kTAM). kTAM features the addition of two additional parameters to aTAM: (1) entropic cost of fixing a monomer unit to another, and (2) the free energy cost of breaking a sticky end bond. kTAM remains a simplified assumption of real world molecule interaction, but provides the substantial improvement over aTAM that it takes into account the rates at which molecules join to and free from each other in a lattice structure. The purpose of this model was to establish better credibility of the possibility of computation by self-assembly of DNA without concrete experimental evidence.

Though kTAM is a more realistic representation of wet-lab assembly of DX molecules, Winfree noted several shortcomings of the model due to simplification including:

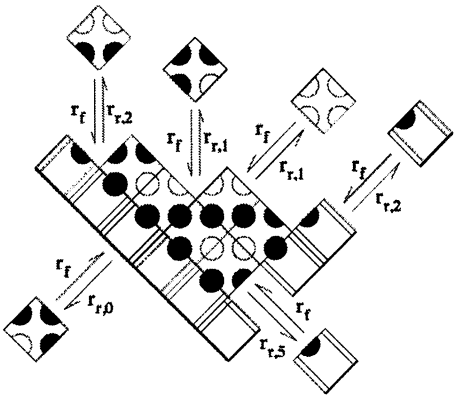

Figure 3:Pictorial representation of kTAM model with forward and reverse rates indicated by r variables. Reproduced from figure 3.22 of [2]

- Monomer (i.e. a square tile in the model) concentrations are held constant throughout simulation, which doesn't perfectly model realistic wet-lab conditions
- Aggregates (i.e. a set square tiles that has reached a stable state in the model) do not react with each other; a drawback because some more subtle reactions between aggregates may be important for understanding self-assembly kinetics in practice
- The forward rates of different types of monomers—basically the rate at which a tile will be added to a structure in the simulation—is assumed to be the same, which is only an approximation of real world scenarios
- The reverse rates of monomers—basically the rate at which a tile will leave a structure in simulation—is assumed to be affected by the number of base-pair bonds that must be broken, but not to be affected ever by mismatched sticky ends. This is another approximation

4 Presentation Challenges

Several of the lessons we learned when developing this project for the class had to do with the intrinsic challenges of bringing highly specialized research such as DNA computation to a more general computer science student audience.

A shortcoming of some, but not all, presentations we saw before ours was that they delved too far into technical details before people could even understand what they were looking at in the first place. We decided that in order to give classmates the most value from watching our presentation, we needed to focus more on helping them grasp the high-level details of the topic at hand rather than grinding through technical formalizations by the second slide. After all, the idea of alternative computational means is not a standard topic in undergraduate computer science courses. Our solution was to replace summarization of Winfree's dissertation from the inside-out standpoint of the dissertation's research itself with an outsider-looking-in viewpoint of DNA computation with some emphasis on Winfree's work.

Our implementation of the solution to this challenge was to build an intuition for self-assembly in general, interactively demonstrate the topic's connection to computational complexity, motivate why any outsider would care about the work in the first place, and finally demonstrate what self-assembly "looks like" using a simulator called Xgrow [5]. We felt that all of these outside components helped to show how Winfree's 1998 work fit into the broader picture of self-assembly.

5 Speculations on Future Directions

A brief look at recent work on the Internet shows that there is current research on improving certain aspects of DNA computation. Work in 2012 has shown that DNA can

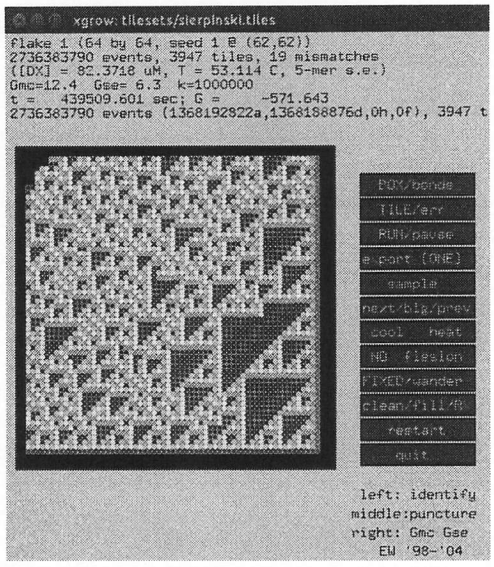

Figure 4: Screenshot of Xgrow doing a kTAM tile simulation of the Sierpinski triangle

be a practical means for data storage [6], and work as recent as February 2015 shows efforts to further improve the efficacy of such a mechanism [7]. There is a belief among some researchers that DNA storage is an option for long-term archival of information due to its high resilience to deterioration under certain conditions [7], and thus we believe this branch of research is likely to continue to be explored for its high potential to produce practical applications.

Another branch of research efforts seems to focus on broadening the scope of problems that can be solved using DNA computation and bringing biological computing closer to practicality. For instance, recent work has been done by Qian and Winfree has been to build more capable DNA computation circuits [8], and work done by Weitz et al. has been towards solving engineering subproblems that remain in the way of practical biological computation [9].

6 Conclusion

In this report we described a number of our efforts in learning about Winfree's 1998 research on DNA self-assembly [2]. We gently introduced and motivated the work of DNA computation by self-assembly, provided a high-level summary of Winfree's 1998 dissertation, presented some of our perceived challenges in helping others to learn the work, and provided our speculations on the future of the field.

Works Cited

[1] L. Adleman, "Molecular computation of solutions to combinatorial problems," *Science,* pp. 1021-1024, 1994.

[2] E. Winfree, Algorithmic Self-Assembly of DNA, California Institute of Technology, 1998.

[3] D. Lewin, "DNA Computing," *Computing in Science & Engineering,* May 2002.

[4] "Google Scholar Page of Erik Winfree," [Online]. Available: https://scholar.google.com/citations?user=5GQF2FwAAAAJ&hl=en&oi=sra. [Accessed April 2015].

[5] The DNA and Natural Algorithms Group, "The Xgrow Simulator," [Online]. Available: http://www.dna.caltech.edu/Xgrow/. [Accessed April 2015].

[6] G. Church, Y. Gao and S. Kosuri, "Next-generation digital information storage in DNA," *Science,* vol. 337, p. 1628, 2012.

[7] R. Grass, R. Heckel, M. Puddu, D. Paunescu and W. Stark, "Robust Chemical Preservation of Digital Information on DNA in Silica with Error-Correcting Codes," *Angewandte Chemie International Edition,* vol. 54, no. 8, pp. 2552-2555, 2015.

[8] L. Qian and E. Winfree, "Scaling up digital circuit computation with DNA strand displacement cascades," *Science,* vol. 332, no. 6034, pp. 1196-1201, 2011.

[9] M. Weitz, J. Kim, K. Kapsner, E. Winfree, E. Franco and F. Simmel, "Diversity in the dynamical behaviour of a compartmentalized programmable biochemical oscillator," *Nature chemistry,* 2014.