Markov Chain Convergence and the Efficiency of Some Self-Assembly Models

Dana Randall*

College of Computing
Georgia Institute of Technology

Abstract. Algorithms based on Markov chains are ubiquitous across scientific disciplines as they provide a method for extracting statistical information about large, complicated systems. For some self-assembly models, Markov chains can be used to predict both equilibrium and non-equilibrium dynamics. In fact, the efficiency of these self-assembly algorithms can be related to the rate of convergence of simple chains. We give an overview of the theory of Markov chains and show how many natural chains, including some relevant in the context of self-assembly, undergo a phase transition as a parameter of the model representing temperature is varied.

1 Introduction

Markov chain Monte Carlo methods are used in many areas of science as a computational tool for studying large, combinatorial sets. A Markov chain is an algorithm that simulates a random walk moving among configurations in the large set, just like shuffling a deck of cards. Even though each configuration might have only a relatively small number of nearest neighbors, a Markov chain is designed so that it converges to desirable distributions over the entire space of configurations. The rate of convergence to this equilibrium distribution, known as the *mixing rate*, determines whether a particular Markov chain provides an efficient tool for sampling. For a state space that is exponentially large in the size of the input, we require that the chain comes close to the equilibrium distribution in only polynomial time in order for this sampling method to be effective.

Recently there has been great success in providing rigorous proofs that certain Markov chains are efficient, and insights from these proofs have guided the design of creative alternative algorithms for sampling. Surveys of some of these proof techniques can be found in [5, 7, 8, 12, 17]. On the other hand, understanding when various Markov chains fail to converge in polynomial time can shed light on the limitations of many natural sampling algorithms, including many tile based DNA self-assembly protocols. We illustrate such a dichotomy in the convergence time of a Markov chain by considering the well-known Ising model from statistical physics, as well as a variant arising in the context of self-assembly.

* Supported in part by NSF grants CCR-0515105 and DMS-0505505.

This abstract is meant to give a brief introduction to some of the main concepts underlying the design and analysis of efficient Markov chains. In section 2 we provide formalization for the pairwise-influence models that arise in both statistical physics and DNA-based self-assembly, in particular the Ising model and a variant. In section 3 we review the fundamentals of Markov chains and in section 4 we discuss the behavior of a natural chain for sampling Ising configurations. In particular, we show that this chain will be efficient at high temperatures and inefficient at low temperatures because of a phase transition in the model. Finally, in the last section we discuss possible ways to remedy this by designing alternative chains for sampling Ising configurations.

2 Self-assembly and pairwise influence models

2.1 DNA-based self-assembly

Self-assembly is a process in which large numbers of simple objects aggregate into larger structures in predictable ways. One exciting approach that has received much attention is tile-based self-assembly models. In tiling models, tiles are designed with markings on each side so that two tiles are more likely to join together along an edge if they have identical markings. Wang studied such tiling systems and showed that they form a universal model of computation [22], making them an appealing object of study.

The primary challenge, then, is to define a set of marked tiles so that tiles are likely to assemble into large aggregates and so that markings determine which tiles are allowed to line up. The approach taken by Seeman, Winfree and others is to use DNA double-crossover molecules [4] to construct marked tiles with DNA sequences on each side [20]. Tiles we would like to have next to each other are defined to have large numbers of complementary base pairs along their matched edges, with the likelihood of their joining determined by hybridization energies. See, e.g., [4, 15, 16, 20, 21] for more details.

Here we are interested in a theoretical abstraction that captures the fundamental features of this model. We imagine we have a large (infinite) supply of rectangular tiles of various types. These tiles have markings on each of their four sides, and there are well-defined energies saying how likely it is for pair to line up along any of their edges. For example, imagine we have two types of tiles, A and B. If the bottom and left sides of A are complementary to the right and top sides of A, and B has similar characteristics but is not similar to A, then tiles of the same type will tend to cluster together. Moreover, this preference is amplified at low temperatures and dampened at high temperatures. See figure 1.

2.2 General pairwise influence models

The tile-based self-assembly algorithms are an example of a broad class of models studied in statistical physics that we can describe as *pairwise-influence models*. Each nearest-neighbor pair of tiles is given a weight representing how likely they

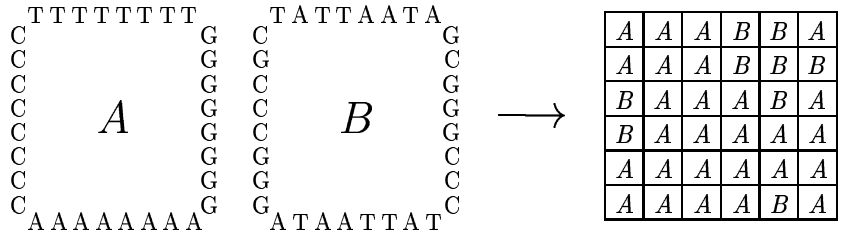

Fig. 1. Tiles A and B designed so that each tile prefers to be adjacent to a tile of the same type.

are to bond to each other. This abstraction is most appropriate when the tiles form a regular lattice so we can think of their placement in the lattice as an assignment to the vertices. Then the "energy" of an edge is precisely the bond strength between the two tiles occupying the endpoints of the edge.

More precisely, suppose we are given any n-vertex graph $G = (V, E)$, say the $n \times n$ grid. Let $\Omega = \{1, ..., q\}^n$ be the state space, where $f : V \to \{1, \dots q\}$ assigns a value from the set $\{1, ..., q\}$ to each vertex in the graph. We define a symmetric set of weights $\{X_{i,j} = X_{j,i}\}$ for each pair $i, j \in \{1, \dots, q\}$ and give each configuration $\sigma \in \Omega$ a weight

$$w(\sigma) = \prod_{u,v:(u,v)\in E} X_{f(u),f(v)}.$$

Normalizing, we get a probability distribution

$$\pi(\sigma) = \frac{w(\sigma)}{\sum_{\tau\in\Omega} w(\tau)}.$$

By adjusting the values for $X_{i,j}$ we can favor certain pairs of nearest neighbors and discourage others. For example, letting $X_{i,j} = 1$ for all $i \neq j$ and letting $X_{i,j} = 0$ whenever $i = j$, the probability distribution arising from the pairwise influence model is precisely the uniform distribution on the set of proper q-colorings.

These models were originally defined to represent simple physical systems. An *energy function* on the space of configurations is defined by a *Hamiltonian* $H(\sigma)$. For models where the energy is determined solely from nearest-neighbor interactions, $H(\sigma) = \sum_{(u,v)\in E} f(\sigma(u), \sigma(v))$, for some function f. Just like a spring relaxing, systems tend to favor configurations that minimize energy, where this preference is controlled by temperature. Each configuration in Ω is given a weight

$$w(\sigma) = e^{-\beta H(\sigma)},$$

where $\beta = 1/T$ is inverse temperature. Thus, for low values of β the differences between the energy of configurations are dampened, while at large β these differences are magnified. Taking $X_{i,j} = e^{f(i,j)}$ reconciles these two definitions.

The likelihood of each configuration is then given by

$$\pi(\sigma) = w(\sigma)/Z,$$

where $Z = \sum_\tau w(\tau)$ is the normalizing constant known as the *partition function.* This is known as the *Gibbs (or Boltzmann) distribution.* Taking derivatives of the generating function Z (or $\ln Z$) with respect to the appropriate variables allows us to calculate many of the interesting thermodynamic properties of the system, such as the specific heat and the free energy.

We can use this formalization to represent many familiar models by adjusting the values for $X_{i,j}$ so as to favor certain pairs of nearest neighbors and discourage others. For example, treating all edges uniformly and letting $X_{i,j} = 1$ for all $i \neq j$ and letting $X_{i,j} = 0$ whenever $i = j$, the probability distribution arising from the pairwise influence model is precisely the uniform distribution on the set of proper q-colorings.

Moreover, we can easily modify the pairwise influence model to make it more amenable to problems arising in the context of self-assembly. Think of G as a directed graph by letting E be a set of ordered pairs. Then, for each directed edge $(u, v) \in E$, we let $\{X_{i,j}^{(u,v)}\}$ be the energy on the edge for an assignment σin which $\sigma(u) = i$ and $\sigma(v) = j$, and each configuration $\sigma \in \Omega$ is given a weight

$$w(\sigma) = \prod_{e=(u,v)\in E} X_{f(u),f(v)}^{(u,v)}.$$

Again normalizing, we get a probability distribution:

$$\pi(\sigma) = \frac{w(\sigma)}{\sum_{\tau\in\Omega} w(\tau)}.$$

In most physical models $X_{i,j}^{(u,v)}$ is taken to be uniform over all edges $(u, v) \in E$, so the superscript is suppressed. We include in here because it allows us to treat energies over directed edges individually as is necessary for tile models. In particular, we can allow one set of energies for tiles that lie next to each other vertically and another if they are next to each other horizontally.

2.3 Saturated and non-saturated Ising models

The *Ising model* is a standard model of ferromagnetism studied in statistical physics. Given a graph G on n vertices, our state space is defined by the 2^n ways of assigning *spins* $+1$ or -1 to each of the vertices. In the *ferromagnetic* Ising model, the Hamiltonian is defined so as to favor configurations which tend to have equal spins on the endpoints of its edges; the *antiferromagnetic* Ising model prefers unequal spins on the endpoints. For $\sigma \in \{\pm\}^n$, the Hamiltonian is given by

$$\pi(\sigma) = \frac{e^{\sum_{(u,v)\in E} \delta(\sigma(u)=\sigma(v))}}{Z},$$

where Z is the normalizing constant and $\delta(a = b) = 1$ if $a = b$ and is 0 otherwise. For an informal introduction to the Ising model, see [3].

This can be thought of as a fully-packed, or *saturated*, self-assembly model with two tiles A and B where each tile prefers to be next to others of the same type (regardless of direction). We can construct this by designing two tiles such that opposite sides of each tile contains complementary sequences, and yet each side is far from complementary to any side of the other type of tile. If the sequences on each side are long enough, then we can arrange it so that the bond strengths satisfy $X_{A,A} = X_{B,B} > X_{A,B} = X_(B, A) = 1$ over all edges, which is precisely the Ising model. See figure 1 for an example of an Ising configuration.

A simple modification of the standard Ising model makes it even more appropriate for the tile-based self-assembly models in which configurations often include empty spaces where there are no tiles present. In the *unsaturated* model, empty spaces are modeled by introducing a third tile type (or spin) C and letting $X_{A,A} = X_{B,B} > X_{A,B} > X_{A,C} = X_{B,C} = X_{C,C} = 1$ be the pairwise interactions. This generalization helps capture the way that tiles will assemble and pull apart at equilibrium according to the temperature. Thus, a tile of either type most prefers to be next to another tile of the same type, but it also prefers to have each of its neighbors occupied rather than unoccupied, even if it is with the other tile. This is demonstrated in figure 2.

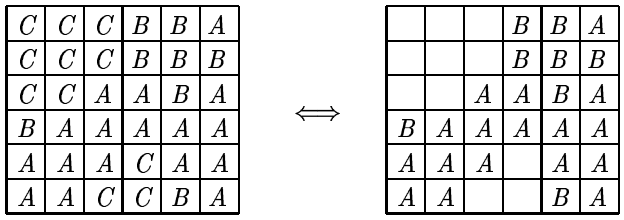

Fig. 2. The unsaturated Ising model.

3 The basics of sampling

A popular method for sampling is to define a Markov chain whose states correspond to the elements of the set. We first define a graph H that connects the states space by allowing us to move between pairs of configurations which are close under some metric. For the Ising model, local or *Glauber dynamics* connect pairs of configurations that have Hamming distance one. This means that in a given step, the chain can pick any position and change the type of tile that is there. For the unsaturated Ising model it might make the most sense to connect two configurations of Hamming distance one only if one of the configurations has a tile of type C in the position of disagreement — recall that C represents an

empty position, so these moves correspond to a tiles of type A or B attaching or detaching from the larger configuration. Notice that the Markov chain connects the state space even if we disallow directly changing a tile from type A to type B.

To define the transition probabilities of the Markov chain on the edges of H, we refer to the desired stationary distribution π. In the case of the Ising model, we want to sample from the *Gibbs distribution.* The *Metropolis-Hastings algorithm* dictates transition probabilies which will force the Markov chain to converge to this distibution [10]:

$$P(x,y) = \frac{1}{2\Delta} \min\left(1, \frac{\pi(y)}{\pi(x)}\right),$$

for all x, y, neighbors in H, where Δ is the maximum degree of G. To implement Glauber dynamics for the Ising model, e.g., starting at configuration x, we pick a vertex v in the lattice uniformly at random as well as a new candidate spin i for that vertex; we then change the spin at v to i with probability $\min\left(1, \frac{\pi(y)}{\pi(x)}\right)$, if this is an allowable move (see figure 3).

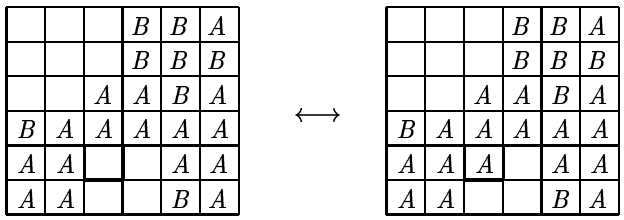

Fig. 3. A move in the local Markov chain for the unsaturated Ising model.

It is easy to verify that if the kernel is ergodic (connected) and aperiodic, then π is the unique stationary distribution. This means that if we start at any vertex in H and perform a random walk according to the transition probabilities P, and we walk long enough, we will converge to the desired distribution π. For this to be useful, we need that we are converging rapidly to π so that after a small, polynomial number of steps, our samples will be chosen from a distibution which is provably arbitrarily close to π. A Markov chain with this property is *rapidly mixing.*

Let $P^t(x,y)$ denote the t-step transition probability from x to y.

Definition 1. *The* total variation distance *at time t is*

$$\|P^t, \pi\| = \max_{x \in \Omega} \frac{1}{2} \sum_{y \in \Omega} |P^t(x,y) - \pi(y)|.$$

Definition 2. *Let $\varepsilon > 0$, then the* mixing time $\tau(\varepsilon)$ *is*

$$\tau(\varepsilon) = \min\{t : \|P^{t'}, \pi\| \leq \varepsilon, \forall t' \geq t\}.$$

$\mathcal{M}$ is *rapidly mixing* if the mixing time is bounded above by a polynomial in n and $\log \frac{1}{\varepsilon}$, where n is the size of each configuration in the state space. When the mixing time is exponential in n, we say the chain is *slowly mixing*.

4 Mixing rates for pairwise influence models

The physical interpretation of pairwise influence models reveals insights into their equilibrium structure, as well as the efficiency of various Markov chain dynamics. Typically, local chains are rapidly mixing at high temperature and slowly mixing at low temperature. For instance, consider the Ising model on the two-dimensional Cartesian lattice. At sufficiently high temperature, Ising configurations will tend to be half + and half -, and we can use a coupling argument to show that Glauber dynamics (the local chain) is rapidly mixing (see, e.g., [1]). At sufficiently low temperature, however, Ising configurations will be predominantly one spin with very high probability. To move from a mostly + configuration to a mostly - one requires moving through a configuration that is half + and half -, but these are exponentially unlikely. This reveals an exponentially small cut in the state space that indicates the chain will require exponential time to converge to equilibrium [5].

Similar results can be shown for the unsaturated Ising model with empty sites, as follows.

Theorem 1. *The mixing time of Glauber dynamics for the unsaturated Ising model on an $n \times n$ square lattice region is at most polynomial in $n, \log(1/\epsilon)$ if β is small enough.*

Theorem 1 follows from a simple coupling argument (see, e.g., [1]). The idea behind coupling is to show that there is a Markov chain on the space $\Omega \times \Omega$ that evolves pairs of configurations simultaneously so that each of the marginals are evolving with probabilities given by the Markov chain under consideration. To show Theorem 1, we define a coupling that updates any pair of configurations by choosing the same position and trying to change the tile at that position to the same new tile. The *coupling time*, or time it takes for the pair of configurations to become identical under these coupled moves, is an upper bound on the mixing time. We can show that at sufficiently high temperature (low β) the Hamming distance between any two unequal configurations will decrease in expectation, and this yields a polynomial upper bound on the mixing time.

At low temperature we see very different behavior.

Theorem 2. *The mixing time of Glauber dynamics for the unsaturated Ising model on an $n \times n$ square lattice region is at least $exp(\psi(\beta)n)$ where $\psi(\beta) > 0$ if β is large enough.*

Theorem 2 can be shown by demonstrating that there is a "bad cut" in the state space, so it will take exponential time in expectation to get from one side of the cut to the other. This indicates that the mixing time will be exponential (see,

e.g., [18]). We can show this cut exists using the idea of topological obstructions introduced in [13] in the constext of independent sets. Both of these proofs follow from arguments used previously the verify analagous theorems for the standard (saturated) Ising model ([19, 11]).

5 Conclusions

In addition to helping us appreciate the dichotomy underlying the behavior of Markov chains based on local updates, the physical models also aid our design of better sampling algorithms. One approach that has received a lot of attention is *simulated annealing* and its sampling analogue *simulated tempering*. Madras and Zhong [9] showed that although sampling Ising configurations at low temperatures is known to be slow, a version of tempering in which the temperature is raised and lowered during the simulation can be shown to fast and can provide samples at all temperatures; they prove this when the underlying graph is the complete graph instead of the lattice. However, it has been shown that for the ferromagnetic Potts model, a closely related pairwise influence model in which there are a larger number of spins, tempering can also require exponential time [2]. This shows that annealing algorithms must be very carefully analyzed to determine whether they can be used for sampling.

Since it is known that locally defined chains cannot be efficient for sampling Ising configurations at low temperature, it suggests that we should look for alternate algorithms based on nonlocal chains. Jerrum and Sinclair found such an algorithm for estimating the partition function of any Ising model at any temperature [6], and Randall and Wilson showed that this can always be used to sample Ising configurations [14]. These results demonstrate that the interplay between statistical physics and computer science can bring to light new ideas in the world of sampling. It is not known whether such results can be extended to the unsaturated model. Moreover, the sampling algorithm of [14] is very indirect and does not provide a natural algorithm in the context of self-assembly. It seems worthwhile to continue to look for alternative approaches that would be more useful in that context.

References

1. D. Aldous. Random walks on finite groups and rapidly mixing Markov chains. *Séminaire de Probabilités* XVII, *Springer Lecture Notes in Mathematics* **986**:243–297, 1981/82.
2. N. Bhatnagar and D. Randall. Torpid mixing of simulated tempering on the Potts model. *Proc. 15th ACM/SIAM Symposium on Discrete Algorithms (SODA)*, 478-487, 2004.
3. B.A. Cipra. An introduction to the Ising model. *American Mathematical Monthly* **94**, pp. 937–959, 1987.
4. T.-J. Fu and N.C. Seeman. DNA double-crossover molecules. *Biochemistry,* **32:** 3211–3220, 1993.

5. M.R. Jerrum and A.J. Sinclair. The Markov chain Monte Carlo method: an approach to approximate counting and integration. In *Approximation Algorithms for NP-Hard Problems,* D.S. Hochbaum, ed., PWS Publishing, Boston: 482–520, 1997.
6. M.R. Jerrum and A.J. Sinclair. Polynomial-time approximation algorithms for the Ising model. *SIAM J. on Computing,* **22**: 1087–1116, 1993.
7. R. Kannan. Markov chains and polynomial time algorithms. *Proc. 35th IEEE Symp. on Foundations of Computer Science (FOCS),* 656–671, 1994.
8. L. Lovász and P. Winkler. Mixing times. *Microsurveys in Discrete Probability,* D. Aldous and J. Propp, eds., *DIMACS Series in Discrete Math. and Theoretical Computer Science,* **41**: 85-134, 1998.
9. N. Madras and Z. Zheng. On the swapping algorithm. *Random Structures and Algorithms*, 22: 6697, 2003.
10. N. Metropolis, A. W. Rosenbluth, M.N. Rosenbluth, A.H. Teller, E. Teller. Equation of state calculations by fast computing machines. *Journal of Chemical Physics*, **21**: 1087–1092, 1953.
11. Y. Peres. On the lower bound for mixing time of Glauber dynamics for the ising model. Unpublished manuscript, 2006.
12. D. Randall. Mixing. *Proc. 44th IEEE Symp. on Foundations of Computer Science (FOCS),* 4–15, 2003.
13. D. Randall. Slow mixing of Glauber dynamics via topological obstructions. *Proc. 17th ACM/SIAM Symposium on Discrete algorithms (SODA)*, 870–879, 2006.
14. D. Randall and D.B. Wilson. Sampling Spin Configurations of an Ising System. *Proc. 10th ACM/SIAM Symposium on Discrete Algorithms (SODA)*, S959–960, 1999.
15. J.H. Reif, S. Sahu and P. Yin Compact Error-Resilient Computational DNA Tiling Assemblies. In *Proc. 10th International Meeting on DNA Computing*, 293–307, 2004
16. N.C. Seeman. DNA in a material world. *Nature,* **421:** 427–431, 2003.
17. A.J. Sinclair. Convergence rates for Monte Carlo experiments. *Numerical Methods for Polymeric Systems,* S.G. Whittington, ed., *IMA Volumes in Mathematics and its Applications,* 1-18, 1997.
18. A.J. Sinclair. *Algorithms for random generation & counting: a Markov chain approach.* Birkhäuser, Boston, 1993.
19. L.E. Thomas. Bound on the mass gap for finite volume stochastic ising models at low temperature. *Communications in Mathematical Physics* **126**, 1–11, 1989.
20. E. Winfree. Simulations of Computing by Self-Assembly. *Proc. 4th DIMACS Meeting on DNA Based Computers*, University of Pennsylvania, 1998.
21. E. Winfree, X. Yang and N.C. Seeman. Universal compuation via self-assembly of DNA: Some theory and experiments. In *DNA Based Computers II,* L.F. Landweber and E.B. Baum, editors, *DIMACS series* **44:** 191–213, 1996.
22. H. Wang. Dominoes and the AEA case of the decision problem. In *Proc. Symp. Math. Theory of Automata* 23–55, 1963.