Exponential Metrics

Lecture by Sarah Cannon and Emma Cohen

November 18th, 2014

1 Background

The coupling theorem we proved in class:

Theorem 1.1. *Let ϕ_t be a random variable [distance metric] satisfying:*

- $0 \leq \phi_t \leq B$ *for some* $B \geq 0$
- $\phi_t = 0$ *implies* $\phi_{t+1} = 0$
- $\mathbb{E}[\Delta\phi_t] \leq 0$
- *If* $\phi_t > 0$ *then* $\mathbb{E}[(\Delta\phi_t)^2] > V$ *for some* $V > 0$

Then the coupling time $T^{x,y}$ *from starting states* x *and* y *satisfies*

$$\mathbb{E}[T^{x,y}] \leq \frac{\phi_0(2B - \phi_0)}{V}$$

In general, you want to show some distance metric is decreasing (or at least non-increasing) in expectation. If there's an actual decrease, then its easy to get a good bound, but even if its just non-increasing then enough variance guarantees it'll get to 0 within polynomial time.

Note that for this to be useful, the upper bound B on the distance must be polynomial.

2 An Interesting Tiling Problem

We will work with a simply connected subset $R \subseteq \mathbb{Z}^d$ (thought of as a union of unit cubes). Usually we will think of R as being a hyper-rectangle, but it doesn't have to be. Let u_i denote the unit vector in the ith coordinate and let $u^* = (1, 1, \ldots, 1)$.

Let $R_L = \{x \in R \mid \exists i \in [k],\ x - u_i \notin R\}$ denote the lower boundary of R.

A subset $S \subseteq R$ is a *downset* if $R_L \subseteq S$ and

$$x \in S \text{ and } x - u_i \in R \implies x - u_i \in S$$

for all i. The *boundary* of a downset $S \subseteq R$ is

$$\partial S = \{x \in S \mid x + u^* \notin S\}.$$

Note that ∂S contains exactly one element along the ray $\{x + ku^* | k \in Z\}$ for each $x \in R_L$. In particular, $|\partial S|$ is an invariant of R common to all downsets $S \subseteq R$, and for any downsets S, T there is a natural bijection between ∂S and ∂T by $v \leftrightarrow v + ku^*$ (where k is the unique integer such that $v + ku^* \in \partial T$).

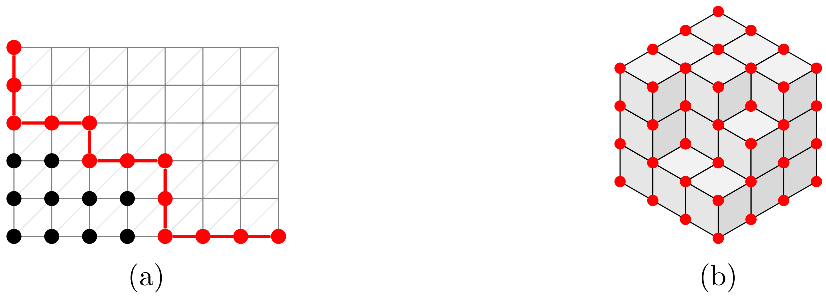


Figure 1: Downsets S in hyperrectangular subsets of $\mathbb{Z}^2$ and $\mathbb{Z}^3$. The red points are the boundary of S. Note that there is exactly one boundary point on each diagonal.

We have talked in the context of lozenge tilings about the 3-dimensional version of the following Markov chain on downsets of R:

- From any state S, pick $(v, b) \in \partial S \times \{\pm 1\}$ uniformly.
- If $b = +1$ move to $S' = S \cup \{v + u^*\}$ if S' is monotone.
- If $b = -1$ move to $S' = S \setminus \{v\}$ if S' is monotone.
- Otherwise remain at S.

The above chain converges to the uniform distribution over downsets of R (the transition matrix is symmetric). What if we want to bias towards larger or smaller subsets? Consider the following *biased* chain with bias $\lambda > 1$:

- From any state S, pick $(x, b, p) \in \partial S \times \{\pm 1\} \times (0, 1)$ uniformly.
- If $b = +1$ move to $S' = S \cup \{x + u^*\}$ if S' is monotone.
- If $b = -1$ $\underline{\text{and } p \leq 1/\lambda}$ move to $S' = S \setminus \{x\}$ if S' is monotone.
- Otherwise remain at S.

Exercise. Use the detailed balance equations to check that in the stationary distribution of this chain each configuration S has weight proportional to $\lambda^{|S|}$.

Question 2.1. When can we show that the biased chain is fast-mixing?

2.1 An attempt at a coupling argument

Let's try the natural coupling using the Hamming distance $d(S,T) = |S \oplus T|$: in each step pick (x, b, p) to update S, and update T according to (x', b, p), where x' is the unique corresponding element of ∂T along the same diagonal ray (i.e., with $x' = x + ku^*$).

The issue is that while (in two dimensions) there may always be at least *as many* good moves as bad moves, the bad moves may be more likely (recall that increase moves are more likely than decreases).

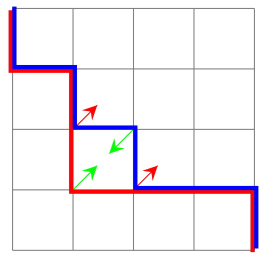


Figure 2: There are two good moves and two bad moves, but increases are more likely than decreases, so overall the bad moves are more likely to occur.

The same situation from Figure 2 in $d > 2$ dimensions is even worse: there are still only two good moves (one up, one down) but now d bad moves (all up).

Remark 2.2. It's only to be expected that this would fail in higher dimensions, as the unweighted version doesn't work well there either (with 2 good moves and up to d bad moves). Istead in class we discussed a modified chain for $d = 3$ where we added or removed whole towers, and it is good to note that the argument for towers also fails in the biased case, for similar reasons.

2.2 A suggested fix

Note that the bad moves are all higher up than the good moves, so let's make the distances of lower moves greater, giving more weight to the expected improvement from good moves. In particular, we'll define a new distance metric

$$\phi(S,T) = \sum_{x \in S \oplus T} \lambda^{(M - x \cdot u^*)/2},$$

where $M = \max_{x \in R} x \cdot u^*$ is the height of the highest point in R. This distance puts more weight on the lower differences between S and T than it does on the higher differences.

Note that (unlike the Hamming distance) this distance $\phi(S,T)$ need not always be integral, so it is conceivable that the distance could keep decreasing indefinitely without ever actually coupling. Fortunately, if $S \neq T$ we have some $x_0 \in S \oplus T$ and so $\phi(S,T) \geq \lambda^{(M - x \cdot u^*)/2} \geq \lambda^0 = 1$. This means that we can still get fast coupling if we can show that the expected distance *decreases by a multiplicative factor* after each step.

In particular, if we have

$$\mathbb{E}[\phi(S_{t+1}, T_{t+1})] \leq \beta \phi(S_t, T_t)$$

then since unequal states are always at distance at least 1 we still have

$$\mathbb{P}[S_t \neq T_t] \leq \mathbb{E}[\phi(S_t, T_t)] \leq \beta^t B \leq e^{-(1-\beta)t} B,$$

where B is the maximum distance between any two states. Setting this $\leq \epsilon$ gives the result

$$\tau(\epsilon) \leq \frac{\ln B/\epsilon}{1-\beta},$$

which is polynomial as long as $1-\beta$ is at least inverse polynomial (assuming B is at most exponential).

But what if we cannot get $\beta < 1$, or $1-\beta$ is exponentially small? Theorem 1.1 may still apply, but we do not get a polynomial bound for this problem since B is not necessarily polynomial for our new metric. In fact, we can see that

$$B \sim \sum_{x \in R} \lambda^{(M - x \cdot u^*)/2} \leq \lambda^{(M-m)/2} |R|,$$

where $m = \min_{x \in R} x \cdot u^*$ is the height of the lowest point in R. This is exponential in the "width" $M - m$ of R.

3 Exponential Metrics

A statement of the full exponential metric theorem:

Theorem 3.1 (Greenberg, Pascoe, Randall, 2009)**.** *Let $\phi : \Omega \times \Omega \to \mathbb{R}^+ \cup \{0\}$ be a metric that takes on finitely many values in $\{0\} \cup [1, B]$. Let $U \subseteq \Omega \times \Omega$ be such that for all $(X_t, Y_t) \in \Omega \times \Omega$, there exists a path $X_t = Z_0, Z_1, ..., Z_r = Y_t$ such that $(Z_i, Z_{i+1}) \in U$ for $0 \leq i < r$ and $\sum_{i=0}^{r-1} \phi(Z_i, Z_{i+1}) = \phi(X_t, Y_t)$.*

Let $\mathcal{M}$ be a lazy Markov chain on Ω and let (X_t, Y_t) be a coupling of $\mathcal{M}$, with $\phi_t := \phi(X_t, Y_t)$. Suppose there exists a $\beta \leq 1$ such that, for all $(X_t, Y_t) \in U$,

$$\mathbb{E}[\phi_{t+1}] \leq \beta \phi_t.$$

1. If $\beta < 1$, the mixing time satisfies

$$\tau(\varepsilon) \leq \frac{\ln(B\varepsilon^{-1})}{1-\beta}.$$

2. If there exists $\kappa, \eta \in (0, 1)$ such that $\mathbb{P}[|\phi_{t+1} - \phi_t| \geq \eta\phi_t] \geq \kappa$ for all t provided that $X_t \neq Y_t$, then

$$\tau(\varepsilon) \leq \left\lceil \frac{e \ln^2(B)}{\ln^2(1+\eta)\kappa} \right\rceil \left\lceil \ln \varepsilon^{-1} \right\rceil$$

Note here distances don't have to be integer, as long as none are between 0 and 1. Additionally, the mixing time is bounded by $\ln(B)$ instead of B.

Proof Sketch. Define new distance

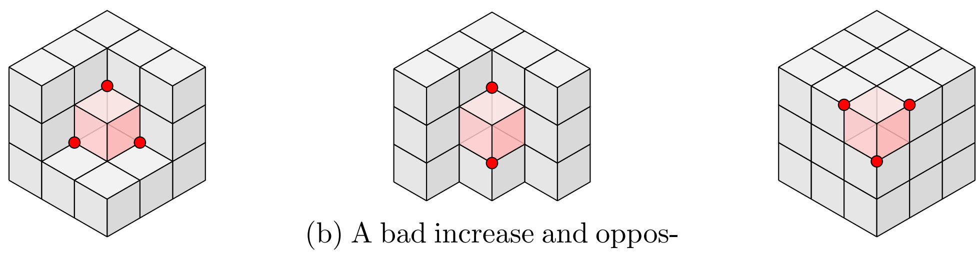
$$\psi_t = \begin{cases} \ln(\phi_t) & : \phi_t > 0 \\ -2\ln 2 & : \phi_t = 0 \end{cases}$$

Lemmas relate the expected changes in ϕ and ψ, and the probabilities that these changes are large. One can then bound the coupling time of ψ, similar to how we proved the coupling theorem in class, by appealing to the Optional Stopping Theorem.

4 Applying the Exponential Metric Theorem

4.1 Analysis of Biased Monotone Surfaces

First note that if $T = S \cup \{x\}$ then the bad moves from (S,T) are those which try to add some $x + u_i$ (which may succeed in T but fail in S) or remove some $x - u_i$ (which may succeed in S but fail in T). Note that at most d such moves can be simultaneously possible, as any bad increase and bad decrease must be in the same dimension. To see this, note that adding $x + u_i$ is only possible in T if all of the elements $x + u_i - u_i'$ are in T, but removing $x - u_j$ is only possible in S if none of the elements $x - u_j + u_j'$ are in S. In particular, if both moves are possible then $x + u_i - u_j \in T \setminus S$, so $u_i = u_j$. That is, either all bad moves are increases (at most d), all are decreases (again at most d) or all bad moves are in the same dimension (at most 2).



(a) Bad increases

(b) A bad increase and opposing decrease

(c) Bad decreases

Figure 3: Possible sets of bad moves. S is given by grey cubes and x is red. Every bad increase must be opposite every bad decrease.

Now we are ready to analyze the expected change in distance after one coupled move of the chain, starting from $(S, T = S \cup \{x\})$ at distance $\phi(S,T) = \lambda^{(M - x\cdot u^*)/2}$. Write $\alpha = |\partial R|$ for the number of possible locations at which to attempt an increase or decrease.

Note that a bad move adding $x + u_i$ increases the distance to

$$\phi(S,T') = \lambda^{(M - x\cdot u^*)/2} + \lambda^{(M-(x+u_i)\cdot u^*)/2} = (1 + 1/\sqrt{\lambda})\phi(S,T),$$

while a bad move removing $x - u_i$ increases the distance to

$$\phi(S',T) = \lambda^{(M - x\cdot u^*)/2} + \lambda^{(M-(x-u_i)\cdot u^*)/2} = (1 + \sqrt{\lambda})\phi(S,T).$$

However, the bad removal only occurs with probability $1/\lambda$ (given that we pick the correct location and direction), so in either case the expected increase in distance (conditioned on picking a bad location and direction, of which there are at most d) is $\frac{1}{\sqrt{\lambda}}\phi(S,T)$.

There are two good moves, occuring with probabilities 1 and $1/\lambda$, each decreasing the distance by $\phi(S,T)$, so the expected change in distance after moving from S,T is at most

$$\Delta\phi \le \left(\frac{d}{2\alpha\sqrt{\lambda}} - \frac{1 + 1/\lambda}{2\alpha}\right)\phi(S,T).$$

The multiplicative factor $\beta = \frac{d}{2\alpha\sqrt{\lambda}} - \frac{1+1/\lambda}{2\alpha}$ is nonpositive when $\sqrt{\lambda} > \frac{2}{d - \sqrt{d^2-4}}$. If λ is significantly larger than this then we can apply the usual coupling noted above for multiplicative decreases. But using the new exponential coupling theorem we can get polynomial mixing right up to this boundary.

(a) (b)

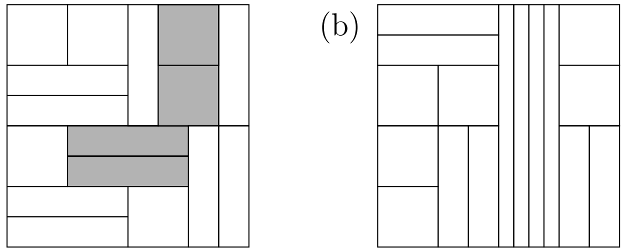


Figure 4: Two tilings of a 16×16 square by 16 rectangles, each of area 16. The shaded rectangles in (a) are not dyadic, but tiling (b) is dyadic.

Theorem 4.1. *The biased monotone surface chain is fast mixing when* $\sqrt{\lambda} > \frac{2}{d-\sqrt{d^2-4}}$.

Remark 4.2. For $d = 2$ this is good for all $\lambda \geq 1$. For larger d it suffices to take $\lambda \geq d^2$. It is an open question whether these chains are fast mixing for smaller λ.

Proof. From Theorem 3.1 it suffices to find $\kappa, \eta > 0$ such that

$$\mathbb{P}[|\phi(S_{t+1}, T_{t+1}) - \phi(S_t, T_t)| > \eta\phi(S_t, T_t)] \geq \kappa$$

whenever $|S_t \oplus T_t| = 1$. Indeed, we always have at least one good increase move, which occurs with probability $1/2\alpha$ and decreases the distance by $\phi(S_t, T_t)$, so $\eta = 1$ and $\kappa = 1/2\alpha$ suffice. $\square$

4.2 Other uses of the exponential metric coupling bound

The exponential metric theorem, above, shows fast mixing for a different problem in "Phase Transitions in Random Dyadic Tilings and Rectangular Dissections," by Cannon, Miracle, and Randall; to appear, 26th Symposium on Discrete Algorithms, 2015.

The Model: An $n \times n$ square in the plane, where $n = 2^k$ for some integer k, tiled by n rectangles, each of area n, as in Figure 4. A tiling is *dyadic* if each rectangle can be written in the form

$$[s2^u, (s+1)2^u] \times [t2^v, (t+1)2^v],$$

for some nonnegative integers s, t, u, v. Intuitively, dyadic tilings are formed by dividing the square in half, either horizontally or vertically; dividing each of those halves in half again, either horizontally or vertically; and repeating until rectangles are of area n. See Figure 5.

The Unweighted Chain: An *edge flip* removes a common side of two adjacent rectangles, and replaces it in the perpendicular direction, as long as the result is another dyadic tiling; see Figure 6. The unweighted Markov chain picks an edge uniformly at random, and the flip occurs (if the result is a dyadic tiling) with probability 1/2.

Exercise. Edge flips connect the state space of all $n \times n$ dyadic tilings.

When all possible edge flips are equally likely to occur, it's not known if this Markov chain is rapidly mixing. A standard path coupling argument, where the distance between two tilings is the number of edge flips needed to get from one to the other, doesn't work, because there could be more bad moves that good moves; see Figure 7, which shows subsets

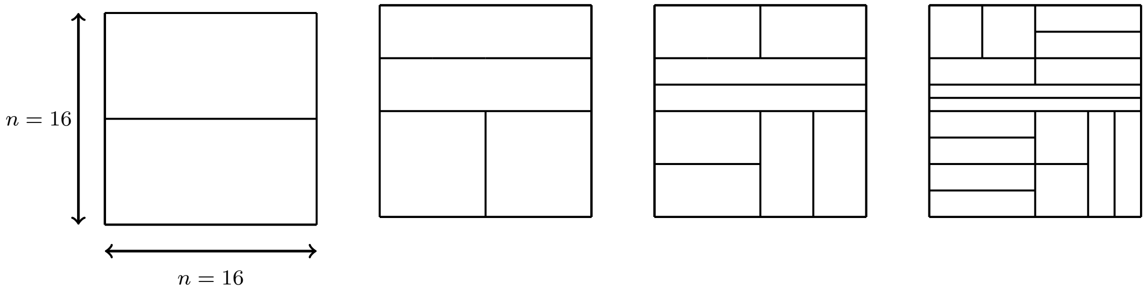


Figure 5: Dyadic tilings are formed by successively dividing existing rectangles in half, horizontally or vertically, until there are $n = 2^k$ rectangles.

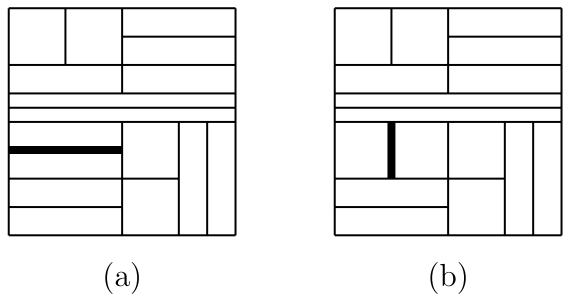


Figure 6: Some dyadic tilings for $n = 16$. Tilings (a) and (b) differ by an edge flip, shown in bold

of two tilings that differ by one edge flip. In fact, it's not obvious that even an exponential metric could be applied.

Notice that the bad moves involve flipping edges of different lengths than the good moves, so it might be possible to introduce some weights (varying with edge lengths) and corresponding distances (from an exponential metric) so that good moves get more weight compared to bad moves.

The Weighted Chain: For a weighted version of the chain, let $\lambda < 1$, where the closer λ is to 1 the closer to uniform the distribution that we're sampling from is. Consider the Markov chain that picks an edge uniformly at random. If it is flippable, call the edge it flips to e', and flip to e' with probability $\min\{1, \lambda^{|e'|-|e|}\}$. This means a longer edge always flips to a shorter edge, but a shorter edge flips to a longer edge with some probability less than one.

Exercise. For a tiling σ, let $|\sigma|$ denote the sum of all edges in σ. Show the weighted Markov chain above has stationary distribution π, where for all dyadic tilings σ, $\pi(\sigma) = \lambda^{|\sigma|}/Z$ for some normalizing constant Z.

This weighted chain can be coupled: Pick uniformly at random an integer point in the $n \times n$ square, an orientation $o \in \{vertical, horizontal\}$, and and a random value $r \in (0, 1)$. If the point is the midpoint of a flippable edge e in orientation o, then e is flipped if r is less than the probability of the flip occurring. Note this coupling may have more stationary probability at each vertex than described above, but the relative probabilities of moves occurring stays the same.

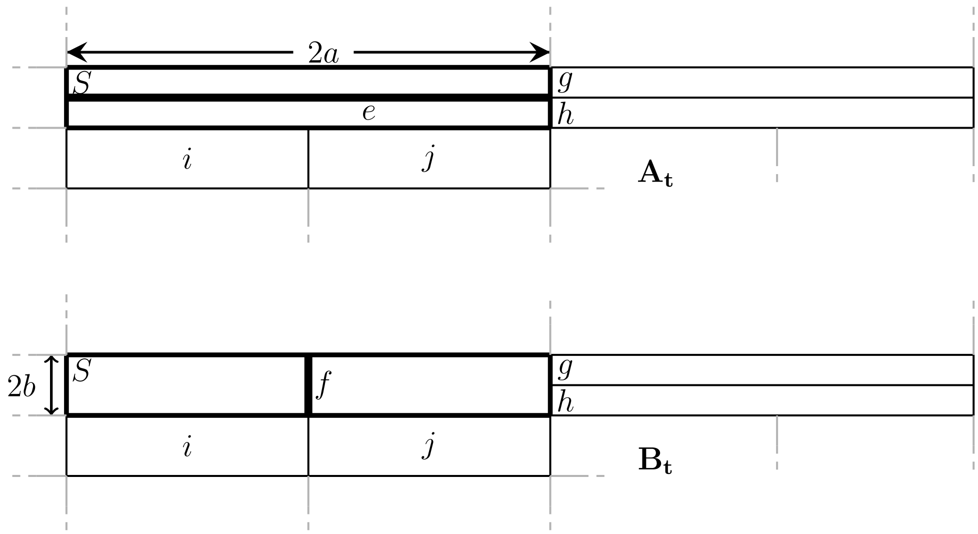


Figure 7: A subset of two coupled tilings that differ by an edge flip between e and f. In A_t flipping e is a good move but flipping g or h is a bad move. In B_t, flipping f is a good move, but flipping i or j is a bad move.

Theorem 4.3. *The weighted chain is rapidly mixing for all $\lambda < 1$.*

This follows from a path coupling argument, with an appropriate exponential metric.

The Metric: Two tilings that differ by a flip between e and e', where $|e| > |e'|$, are at distance $\lambda^{|e'|-|e|} \geq 1$. Distances between any two tilings differing by more than a flip are just the sum of distances along a shortest path of flips between them. Distances could certainly be exponential and are not necessarily integer, but are all at least one.

Proof of Theorem 4.3: Let q be the probability that a given edge is selected to be flipped, the same for all flippable edges; the flip then actually occurs with some probability depending on the lengths of e and e'.

Consider any two tilings differing by a flip between edge e and edge f; we'll just consider the case where $|e| \geq 8|f|$. As in Figure 7, there can be bad flips to the left or the right of e but not both, because of dyadic properties; similarly, there can be bad flips above or below f but not both. Looking at Figure 7, where e has length $2a$ and f has length $2b$:

- Bad edge flips g and h are each selected with probability q, flips occur with probability λ^{4a-b}, and the change in distance is λ^{b-4a}.
- Bad edge flips i and j are each selected with probability q, flips occur with probability 1, and the change in distance is λ^{4b-a}.
- Good move flipping e to f is selected with probability q, occurs with probability 1, and the change in distance is $-\lambda^{2b-2a}$.
- Good move flipping f to e is selected with probability q, occurs with probability λ^{2a-2b}, and the change in distance is $-\lambda^{2b-2a}$.

Altogether, assuming n is large enough so that $\lambda < 3^{-1/\sqrt{n}}$, if $\phi_t = \lambda^{2b-2a}$ is the distance between the two coupled chains that differ by one step at time t ,

$$\begin{aligned}
\mathbb{E}[\phi_{t+1} - \phi_t] &\leq -q - q\lambda^{2b-2a} + 2q + 2q\lambda^{4b-a} \\
&= -q\lambda^{2b-2a}(\lambda^{2a-2b} + 1 - 2\lambda^{2a-2b} - 2\lambda^{2b+a}) \\
&= -q\phi_t(1 - \lambda^{2a-2b} - 2\lambda^{2b+a}) \\
\therefore \mathbb{E}[\phi_{t+1}] &\leq (1 - qc)\phi_t.
\end{aligned}$$

The constant c above just depends on how close λ is to the bound given above. This, along with the other hypotheses of the exponential metric theorem, show that the weighted edge flip chain for dyadic tilings is rapidly mixing for all $\lambda < 1$, in time $\frac{\ln(B\varepsilon^{-1})}{qc} = O(n^2 \log(n/\varepsilon))$. For full details and calculations, see Section 3 of "Phase transitions in Random Dyadic Tilings and Rectangular Dissections."

References

- S. Greenberg, A. Pascoe, and D. Randall. "Sampling biased lattice configurations using exponential metrics." In 20th Symposium on Discrete Algorithms, volume 21, pages 225-251, 2009.

- S. Cannon, S. Miracle, and D. Randall. "Phase transitions in random dyadic tilings and rectangular dissections." Symposium on Discrete Algorithms, 2015, to appear.