New York Journal of Mathematics

New York J. Math. **4** (1998) 137–165.

The Shape of a Typical Boxed Plane Partition

Henry Cohn, Michael Larsen, and James Propp

Abstract. Using a calculus of variations approach, we determine the shape of a typical plane partition in a large box (i.e., a plane partition chosen at random according to the uniform distribution on all plane partitions whose solid Young diagrams fit inside the box). Equivalently, we describe the distribution of the three different orientations of lozenges in a random lozenge tiling of a large hexagon. We prove a generalization of the classical formula of MacMahon for the number of plane partitions in a box; for each of the possible ways in which the tilings of a region can behave when restricted to certain lines, our formula tells the number of tilings that behave in that way. When we take a suitable limit, this formula gives us a functional which we must maximize to determine the asymptotic behavior of a plane partition in a box. Once the variational problem has been set up, we analyze it using a modification of the methods employed by Logan and Shepp and by Vershik and Kerov in their studies of random Young tableaux.

Contents

Received January 6, 1997.

Mathematics Subject Classification. Primary 60C05, 05A16; Secondary 60K35, 82B20.

Key words and phrases. Plane partitions, rhombus tilings of hexagons, calculus of variations, random tilings, limit laws for random structures.

Cohn was supported by an NSF Graduate Research Fellowship. Larsen was supported by NSF grant DMS94-00833 and by the Sloan Foundation. Propp was supported by NSA grant MDA904-92-H-3060 and NSF grant DMS92-06374, and by a grant from the MIT Class of 1922.

ISSN 1076-9803/98

1. Introduction

In this paper we will show that almost all plane partitions that are constrained to lie within an $a \times b \times c$ box have a certain approximate shape, if a, b, and c are large. Moreover, this limiting shape depends only on the mutual ratios of a, b, and c. Our proof will make use of the equivalence between such plane partitions and tilings of hexagons by rhombuses.

Recall that plane partitions are a two-dimensional generalization of ordinary partitions. A *plane partition* π is a collection of non-negative integers $\pi_{x,y}$ indexed by ordered pairs of non-negative integers x and y, such that only finitely many of the integers $\pi_{x,y}$ are non-zero, and for all x and y we have $\pi_{x+1,y} \leq \pi_{x,y}$ and $\pi_{x,y+1} \leq \pi_{x,y}$. A more symmetrical way of looking at a plane partition is to examine the union of the unit cubes $[i, i+1] \times [j, j+1] \times [k, k+1]$ with i, j, and k non-negative integers satisfying $0 \leq k < \pi_{i,j}$. This region is called the *solid Young diagram* associated with the plane partition, and its volume is the sum of the $\pi_{x,y}$'s.

We say that a plane partition π *fits within an $a \times b \times c$ box* if its solid Young diagram fits inside the rectangular box $[0,a]\times[0,b]\times[0,c]$, or equivalently, if $\pi_{x,y} \leq c$ for all x and y, and $\pi_{x,y} = 0$ whenever $x \geq a$ or $y \geq b$. We call such a plane partition a *boxed plane partition*. Plane partitions in an $a \times b \times c$ box are in one-to-one correspondence with tilings of an equi-angular hexagon of side lengths a, b, c, a, b, c by rhombuses whose sides have length 1 and whose angles measure $\pi/3$ and $2\pi/3$. It is not hard to write down a bijection between the plane partitions and the tilings, but the correspondence is best understood informally, as follows. The faces of the unit cubes that constitute the solid Young diagram are unit squares. Imagine now that we augment the solid Young diagram by adjoining the three "lower walls" of the $a \times b \times c$ box that contains it (namely $\{0\} \times [0,b] \times [0,c]$, $[0,a] \times \{0\} \times [0,c]$, and $[0,a] \times [0,b] \times \{0\}$); imagine as well that each of these walls is divided into unit squares. If we look at this augmented Young diagram from a point on the line $x = y = z$, certain of the unit squares are visible (that is, unobstructed by cubes). These unit squares form a surface whose boundary is the non-planar hexagon whose vertices are the points $(a,0,c)$, $(0,0,c)$, $(0,b,c)$, $(0,b,0)$, $(a,b,0)$, $(a,0,0)$, and $(a,0,c)$, respectively. Moreover, these same unit squares, projected onto the plane $x + y + z = 0$ and scaled, become rhombuses which tile the aforementioned planar hexagon. For example, the plane partition π in a $2 \times 2 \times 2$ box defined by $\pi_{0,0} = \pi_{0,1} = \pi_{1,0} = 1$ and $\pi_{1,1} = 0$ corresponds to the tiling in Figure 1, where the points $(2,0,0)$, $(0,2,0)$, and $(0,0,2)$ are at the lower left, extreme right, and upper left corners of the hexagon. (The shading is meant as an aid for three-dimensional visualization, and is not necessary mathematically. The unshaded rhombuses represent part of the walls.)

We will use the term *a, b, c hexagon* to refer to an equi-angular hexagon of side lengths a, b, c, a, b, c (in clockwise order, with the horizontal sides having length b), and the term *lozenge* to refer to a unit rhombus with angles of $\pi/3$ and $2\pi/3$. We will focus, without loss of generality, on those lozenges whose major axes are vertical, which we call *vertical lozenges*. Although our method in this article is to reduce facts about plane partitions to facts about tilings, one can also go in the reverse direction. For example, one can see from the three-dimensional picture that in every lozenge tiling of an a, b, c hexagon, the number of vertical lozenges

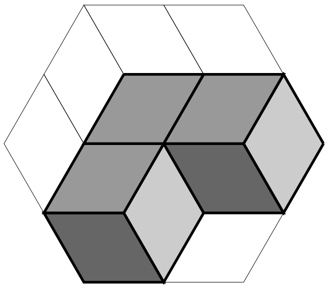

FIGURE 1. The tiling corresponding to a plane partition.

is exactly ac (with similar formulas for the other two orientations of lozenge); see [DT] for further discussion.

A classical formula of MacMahon [M] asserts that there are exactly

$$\prod_{i=0}^{a-1}\prod_{j=0}^{b-1}\prod_{k=0}^{c-1}\frac{i+j+k+2}{i+j+k+1}$$

plane partitions that fit in an $a \times b \times c$ box, or (equivalently) lozenge tilings of an a, b, c hexagon. In Theorem 2.2 of this article, we give a generalization that counts lozenge tilings with prescribed behavior on a given horizontal line.

Using Theorem 2.2, we will determine the shape of a typical plane partition in an $a\times b\times c$ box (Theorem 1.2). Specifically, that theorem implies that if a, b, c are large, then the solid Young diagram of a random plane partition in an $a\times b\times c$ box is almost certain to differ from a particular, "typical" shape by an amount that is negligible compared to abc (the total volume of the box). Equivalently, the visible squares in the augmented Young diagram of the random boxed plane partition form a surface whose maximum deviation from a particular, typical surface is almost certain to be $o(\min(a, b, c))$. Moreover, scaling a, b, c by some factor causes the typical shape of this bounding surface to scale by the same factor.

Before we say what the true state of affairs is, we invite the reader to come up with a guess for what this typical shape should be. One natural way to arrive at a guess is to consider the analogous problem for ordinary (rather than plane) partitions. If one considers all ordinary partitions that fit inside an $a \times b$ rectangle (in the sense that their Young diagrams fit inside $[0, a] \times [0, b]$), then it is not hard to show that the typical boundary of the diagram is the line $x/a + y/b = 1$. That is, almost all such partitions have roughly triangular Young diagrams. (One way to prove this is to apply Stirling's formula to binomial coefficients and employ direct counting. Another is to use probabilistic methods, aided by a verification that if we look at the boundary of the Young diagram of the partition as a lattice path, then the directions of different steps in the path are negatively correlated.) It therefore might seem plausible that the typical bounding surface for plane partitions would be a plane (except where it coincided with the sides of the box), as when a box is tilted on its corner and half-filled with fluid. However, Theorem 1.1 shows that that is in fact not the case.

To see what a typical boxed plane partition *does* look like, see Figure 2. This tiling was generated using the methods of [PW] and is truly random, to the extent

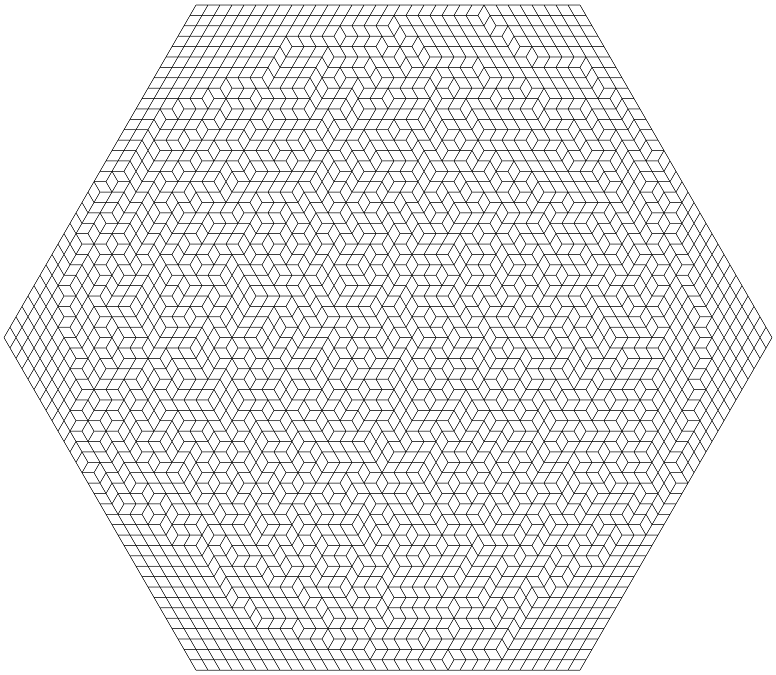

FIGURE 2. A random lozenge tiling of a $32, 32, 32$ hexagon.

that pseudo-random number generators can be trusted. Notice that near the corners of the hexagon, the lozenges are aligned with each other, while in the middle, lozenges of different orientations are mixed together. If the bounding surface of the Young diagram tended to be flat, then the central zone of mixed orientations would be an inscribed hexagon, and the densities of the three orientations of tiles would change discontinuously as one crossed into this central zone. In fact, what one observes is that the central zone is roughly circular, and that the tile densities vary continuously except near the midpoints of the sides of the original hexagon.

One can in theory use our results to obtain an explicit formula for the typical shape of the bounding surface, in which one specifies the distance from a point P on the surface to its image P' under projection onto the $x + y + z = 0$ plane, as a function of P'. However, the integral that turns up is quite messy (albeit evaluable in closed form), with the result that the explicit formula is extremely lengthy and unenlightening. Nevertheless, we can and do give a fairly simple formula for the tilt of the tangent plane at P as a function of the projection P', which would allow one to reconstruct the surface itself. In view of the correspondence between plane partitions and tilings, specifying the tilt of the tangent plane is equivalent to specifying the local densities of the three different orientations of lozenges for random tilings of an a, b, c hexagon.

Our result on local densities has as a particular corollary the assertion that, in an asymptotic sense, the zone of mixed orientation (defined as the region in which all three orientations of lozenge occur with positive density) is precisely the interior of the ellipse inscribed in the hexagon. This behavior is analogous to what has been proved concerning domino tilings of regions called Aztec diamonds (see [JPS] and [CEP]). These are roughly square regions, and if one tiles them randomly with dominos (1×2 rectangles), then the zone of mixed orientation tends in the limit to the inscribed disk. However, the known results for Aztec diamonds are stronger than the corresponding best known results for hexagons (see Conjecture 6.2 in Section 6).

To state our main theorem, we begin by setting up normalized coordinates. Suppose that we are dealing with an a, b, c hexagon, so that the side lengths are a, b, c, a, b, c in clockwise order (with the sides of length b horizontal). We let a, b, c tend to infinity together, in such a way that the three-term ratio $a : b : c$ (i.e., element of the projective plane $\mathbb{P}^2(\mathbb{R})$) converges to $\alpha : \beta : \gamma$ for some fixed positive numbers α, β, γ. Say $a + b + c = \sigma(\alpha + \beta + \gamma)$ for some scaling factor σ. Then we choose re-scaled coordinates for the a, b, c hexagon so that its sides are $a' = a/\sigma$, $b' = b/\sigma$, and $c' = c/\sigma$ (which by the hypothesis of our main theorem will be required to converge to α, β, and γ as a, b, and c get large). Note that a', b', c' in general are not integers. The origin of our coordinate system lies at the center of the hexagon. One can check that the sides of the hexagon lie on the lines $y = \frac{\sqrt{3}}{2}(2x + b' + c')$, $y = \frac{\sqrt{3}}{4}(a' + c')$, $y = \frac{\sqrt{3}}{2}(-2x + a' + b')$, $y = \frac{\sqrt{3}}{2}(2x - b' - c')$, $y = \frac{\sqrt{3}}{4}(-a' - c')$, and $y = \frac{\sqrt{3}}{2}(-2x - a' - b')$, and that the inscribed ellipse is described by

$$3(a'+c')^2x^2 - 2\sqrt{3}(a'+2b'+c')(a'-c')xy + ((a'+2b'+c')^2 - 4a'c')y^2 = 3a'b'c'(a'+b'+c').$$

Define $E_{\alpha,\beta,\gamma}(x, y)$ to be the polynomial

$$3\alpha\beta\gamma(\alpha+\beta+\gamma) - (3(\alpha+\gamma)^2x^2 - 2\sqrt{3}(\alpha+2\beta+\gamma)(\alpha-\gamma)xy + ((\alpha+2\beta+\gamma)^2 - 4\alpha\gamma)y^2),$$

whose zero-set is the ellipse inscribed in the α, β, γ hexagon. Also define $Q_{\alpha,\beta,\gamma}(x, y)$ to be the polynomial

$$\frac{\sqrt{3}}{2}\left(\frac{4}{3}y^2 - 4x^2 + \beta^2 + \alpha\beta + \beta\gamma - \alpha\gamma\right),$$

which will be useful shortly.

There are six points at which the ellipse inscribed in the α, β, γ hexagon meets the boundary of the hexagon. The four that occur on sides of length α or γ will be called *singular points*, for reasons that will be clear shortly. (Recall that we have already broken the underlying symmetry of the situation by agreeing to focus on vertical lozenges.) The points of the hexagon that lie outside the inscribed ellipse form six connected components. Let R_1 be the closure of the union of the two components containing the leftmost and rightmost points of the hexagon, minus the four singular points, and let R_0 be the closure of the union of the other four components, again minus the four singular points.

Finally, define the (normalized) coordinates of a vertical lozenge to be the (normalized) coordinates of its center. Then the following theorem holds:

Theorem 1.1. *Let U be the interior of a smooth simple closed curve contained inside the α, β, γ hexagon, with $\alpha, \beta, \gamma > 0$. In a random tiling of an a, b, c hexagon, as $a, b, c \to \infty$ with $a : b : c \to \alpha : \beta : \gamma$, the expected number of vertical lozenges whose normalized coordinates lie in U is $ab + bc + ac$ times*

$$\frac{1}{A}\iint_U \mathcal{P}_{\alpha,\beta,\gamma}(x, y)\, dx\, dy \;+\; o(1),$$

where $A = (\alpha\beta + \beta\gamma + \alpha\gamma)\sqrt{3}/2$ *is the area of the* α, β, γ *hexagon and*

$$\mathcal{P}_{\alpha,\beta,\gamma}(x,y) = \begin{cases} \dfrac{1}{\pi}\cot^{-1}\left(\dfrac{Q_{\alpha,\beta,\gamma}(x,y)}{\sqrt{E_{\alpha,\beta,\gamma}(x,y)}}\right) & \text{if } E_{\alpha,\beta,\gamma}(x,y) > 0, \\ 0 & \text{if } (x,y) \in R_0, \text{ and} \\ 1 & \text{if } (x,y) \in R_1. \end{cases}$$

In fact, our proof will give an even stronger version of Theorem 1.1, in which U is a horizontal line segment rather than an open set (and the double integral is replaced by a single integral). Since we can derive the expected number of vertical lozenges whose normalized coordinates lie in U by integrating over all horizontal cross sections (provided that the error term is uniform for all cross sections, as we will prove in Section 4), this variant of the claim implies the one stated above, though it is not obviously implied by it. We have stated the result in terms of open sets because that formulation seems more natural; the proof just happens to tell us more.

The intuition behind Theorem 1.1 is that $\mathcal{P}_{\alpha,\beta,\gamma}(x,y)$ gives the density of vertical lozenges in the normalized vicinity of (x,y); the factor of $ab+bc+ac$ arises simply because there are that many lozenges in a tiling of an a,b,c hexagon. One might be tempted to go farther and think of $\mathcal{P}_{\alpha,\beta,\gamma}(x,y)$ as the asymptotic probability that a random tiling of the a,b,c hexagon will have a vertical lozenge at any particular location in the normalized vicinity of (x,y) (see Conjecture 6.1 in Section 6). However, we cannot justify this interpretation rigorously, because it is conceivable that there are small-scale fluctuations in the probabilities that disappear in the double integral. (In Subsection 1.3 of [CEP], it is shown that the analogous probabilities for random domino tilings do exhibit such fluctuations, although the fluctuations disappear if one distinguishes between four classes of dominos, rather than just horizontal and vertical dominos.)

The formula for $\mathcal{P}_{\alpha,\beta,\gamma}$ is more natural than it might appear. Examination of random tilings such as the one shown in Figure 2 leads one to conjecture that the region in which all three orientations of lozenges occur with positive density is (asymptotically) the interior of the ellipse inscribed in the hexagon, and the known fact that an analogous claim holds for random domino tilings of Aztec diamonds (see [CEP]) lends further support to this hypothesis. This leads one to predict that $\mathcal{P}_{\alpha,\beta,\gamma}$ will be 0 in R_0 and 1 in R_1, and strictly between 0 and 1 in the interior of the ellipse. Comparison with the analogous theorem for domino tilings (Theorem 1 of [CEP]) suggests that within the inscribed ellipse, $\mathcal{P}_{\alpha,\beta,\gamma}(x,y)$ should be of the form

$$\frac{1}{\pi}\cot^{-1}\left(\frac{Q(x,y)}{\sqrt{E_{\alpha,\beta,\gamma}(x,y)}}\right)$$

for some quadratic polynomial $Q(x,y)$, and the only problem is actually finding $Q(x,y)$ in terms of α,β,γ. A simple description of the polynomial $Q(x,y)$ that actually works is that its zero-set is the unique hyperbola whose asymptotes are parallel to the non-horizontal sides of the hexagon and which goes through the four points on the boundary of R_1 where the inscribed ellipse is tangent to the hexagon. This determines $Q(x,y)$ up to a constant factor, and that constant factor

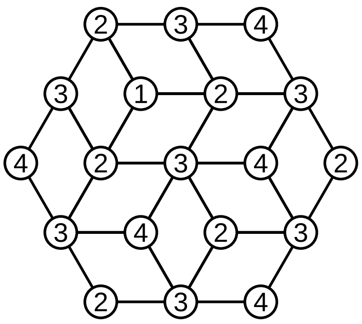

FIGURE 3. The height function corresponding to a tiling.

is determined (at least in theory—in practice the calculations would be cumbersome) by requiring that the average of $\mathcal{P}_{\alpha,\beta,\gamma}(x,y)$ over the entire hexagon must be $\alpha\gamma/(\alpha\beta+\beta\gamma+\alpha\gamma)$.

An alternative way to phrase Theorem 1.1 is in terms of height functions, which were introduced by Thurston in [Th] as a geometrical tool for understanding tilings of regions by lozenges or dominos. A height function is a numerical representation of an individual tiling of a specified region. In the case of lozenge tilings of a hexagon, the vertices of the lozenges occur at points of a certain triangular lattice that is independent of the particular tiling chosen, and the height function simply associates a certain integer to each such vertex so as to describe the shape of the plane partition that corresponds to the tiling. Given any lozenge tiling, one can assign heights $h(u)$ to the points u of the triangular lattice as follows. Give the leftmost vertex of the hexagon height $a+c$. (We choose this height so that the vertex of the bounding box farthest from the viewer, which is usually obscured from view, will have height 0.) Suppose that u and v are adjacent lattice points, such that the edge connecting them does not bisect a lozenge. If the edge from u to v points directly to the right, set $h(v)=h(u)+1$, and if it points up and to the right, or down and to the right, set $h(v)=h(u)-1$. (If it points left, change $+1$ to -1 and vice versa.) If one traces around each lozenge in the tiling and follows this rule, then every vertex is assigned a height. It is not hard to check that the heights are well-defined, so there is a unique *height function* associated to the tiling. For an example, see Figure 3. Conversely, every way of assigning integer heights to the vertices of the triangular grid that assigns height $a+c$ to the leftmost vertex of the hexagon and that satisfies the edge relation must be the height function associated to some unique tiling. If one views the tiling as a three-dimensional picture of the solid Young diagram of a plane partition, then the height function tells how far above a reference plane (of the form $x+y+z=\text{constant}$) each vertex lies. The values of the height function on the boundary of the hexagon are constrained, but all the values in the interior genuinely depend on the tiling. (It should be mentioned that height functions for lozenge tilings are implicit in the work of physicists Blöte and Hilhorst, in the context of the two-dimensional dimer model on a hexagonal lattice; see [BH].)

When dealing with normalized coordinates, it is convenient to use a *normalized height function*: if we scale the coordinates by dividing them by σ (as described above), then we also divide the height function values by σ. Also, we define the average height function to be the average of the height functions associated with all possible tilings. (Of course, it is not a height function itself.) We will show that almost every height function closely approximates the average height function, asymptotically:

Theorem 1.2. *Fix $\alpha, \beta, \gamma > 0$. As $a, b, c \to \infty$ with $a : b : c \to \alpha : \beta : \gamma$, the normalized average height function of the a, b, c hexagon converges uniformly to the function $\mathcal{H}_{\alpha,\beta,\gamma}$ with the appropriate (piecewise linear) boundary conditions such that*

$$\frac{\partial \mathcal{H}_{\alpha,\beta,\gamma}(x,y)}{\partial x} = 1 - 3\mathcal{P}_{\alpha,\beta,\gamma}(x,y).$$

For fixed $\varepsilon > 0$, the probability that the normalized height function of a random tiling will differ anywhere by more than ε from $\mathcal{H}_{\alpha,\beta,\gamma}$ is exponentially small in σ^2, where $\sigma = (a+b+c)/(\alpha+\beta+\gamma)$.

It is not hard to deduce Theorem 1.2 (with the exception of the claim made in the last sentence) from Theorem 1.1. In particular, the differential equation simply results from considering how the height changes as one crosses lozenges of each orientation. We will give the proof in detail in Section 4.

Unfortunately, although one can recover an explicit formula for $\mathcal{H}_{\alpha,\beta,\gamma}(x,y)$ from the boundary values and the knowledge of $\partial \mathcal{H}_{\alpha,\beta,\gamma}(x,y)/\partial x$, we cannot find any simple formula for it. By contrast, Proposition 17 of [CEP] gives a comparatively simple asymptotic formula for the average height function for domino tilings of an Aztec diamond.

Our methods also apply to the case of random domino tilings of Aztec diamonds. Formula (7) of Section 4 of [EKLP] is analogous to our Theorem 2.2, and can be used in the same way. It turns out that the functional arrived at by applying the methods of Section 3 of this paper to that formula is very similar to the one we will find in Section 3. After a simple change of variables, one ends up with essentially the same functional, but maximized over a slightly different class of functions. The methods of Section 4 apply almost without change, and the methods of Section 5 can be adapted to prove Proposition 17 of [CEP]. This proof is shorter than the one given in [CEP]. However, in [CEP], Proposition 17 comes as a corollary of a much stronger result (Theorem 1), which the methods of this paper do not prove.

2. The Product Formula

In this section, we will prove a refinement of MacMahon's formula, following methods first used by Elkies et al. in [EKLP]. This refinement (Proposition 2.1) is not strictly speaking new, since it is really nothing more than the Weyl dimension formula for finite-dimensional representations of $\mathrm{SL}(n)$ (we say a few words more about this connection below). However, we give our own proof of this result for two reasons: first, to make this part of the proof self-contained; and second, to illustrate an expeditious method of proof that has found applications elsewhere (see [JP] for related formulas derived by the same method).

Proposition 2.1 is stated in terms of Gelfand patterns, so we must first explain what Gelfand patterns are and what they have to do with lozenge tilings. It is not hard to see that a lozenge tiling of a hexagon is determined by the locations of its vertical lozenges. A semi-strict Gelfand pattern is a way to keep track of

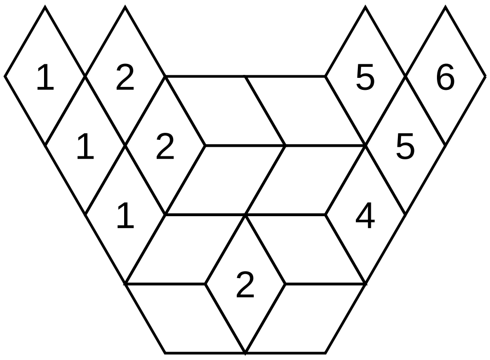

FIGURE 4. The semi-strict Gelfand pattern corresponding to a tiling.

these locations. Specifically, one augments the a, b, c hexagon by adding $a(a+1)/2$ vertical lozenges on the left and $c(c+1)/2$ vertical lozenges on the right, forming an approximate trapezoid of upper base $a+b+c$ and lower base b, with some triangular protrusions along its upper border, as in Figure 4. One then associates with each vertical lozenge in the tiling the horizontal distance from its right corner to the left border of the trapezoid, which we call its *trapezoidal position*. (When we want to use the left boundary of the hexagon instead of the left boundary of the approximate trapezoid as our bench mark, we will speak of the *hexagonal position* of a vertical lozenge.) For example, consider the tiling shown in Figure 1. We augment it by adding 6 vertical lozenges to form the tiling shown in Figure 4. The vertical lozenges form rows, and the only restriction on their placement is that given any two adjacent vertical lozenges in the same row, there must be exactly one vertical lozenge between them in the row immediately beneath them. If we index the locations of the vertical lozenges with their trapezoidal positions (the numbers shown in Figure 4), we arrive at the following semi-strict Gelfand pattern:

$$\begin{array}{ccccccc} 1 & & 2 & & 5 & & 6 \\ & 1 & & 2 & & 5 & \\ & & 1 & & 4 & & \\ & & & 2 & & & \end{array}$$

In general, a *semi-strict Gelfand pattern* is a triangular array of integers (such as this one), with the property that the entries increase weakly along diagonals running down and to the right, and the entries increase strictly along diagonals running up and to the right. As discussed above, there is a simple bijection between semi-strict Gelfand patterns with top row $1, 2, \ldots, a, a+b+1, a+b+2, \ldots, a+b+c$ and lozenge tilings of an equi-angular hexagon with side lengths a, b, c, a, b, c.

Moreover, consider the k-th horizontal line from the top in an a, b, c hexagon, where the top edge of the hexagon corresponds to $k = 0$, so that the trapezoidal positions of the vertical lozenges on the k-th line are precisely the entries in the $(k+1)$-st row of the associated semi-strict Gelfand pattern. If we discard all lozenges that lie above the line (but retain all vertical lozenges that are bisected by it and all lozenges that lie below it), then we get a tiling of a smaller approximate trapezoid, whose upper border as before consists of triangular protrusions alternating with straight edges, except that now the protrusions need not be concentrated at the left and right portions of the border. The trapezoidal positions of the vertical lozenges in this tiling are given by the entries in the semi-strict Gelfand pattern obtained from the original semi-strict Gelfand pattern by deleting the first k rows. In fact, if we limit ourselves to tilings of the a, b, c hexagon in which the locations of the vertical lozenges that are bisected by the k-th horizontal line are specified, then each individual tiling of this kind corresponds to a *pair* of semi-strict Gelfand patterns. We have already described one of these Gelfand patterns, which gives the behavior of the tiling below the cutting line; the other, which describes the tilings above the line, comes from reflecting the hexagon through the horizontal axis (and of course adjoining additional lozenges, as above). If $0 \leq k < \min(a, c)$, then one of the Gelfand patterns will include some of the augmenting vertical lozenges described above on both sides (a row of length $a - k$ on one side of the top of the pattern, and a row of length $c - k$ on the other) and the other pattern will contain augmenting vertical lozenges on neither side; if $\min(a, c) \leq k \leq (a + c)/2$, then one of the Gelfand patterns will contain a row of $|a - k|$ augmenting lozenges on one side and the other Gelfand pattern will contain $|c - k|$ on the other side. (The case $k > (a + c)/2$ is symmetric to the case $k < (a + c)/2$, so we do not treat it explicitly.)

In Theorem 2.2, we will use the following formula to determine how many tilings have a specified distribution of vertical lozenges along a horizontal line.

Proposition 2.1. *There are exactly*

$$\prod_{1 \leq i < j \leq n} \frac{x_j - x_i}{j - i}$$

semi-strict Gelfand patterns with top row consisting of integers $x_1, x_2, \ldots, x_n$ such that $x_1 < x_2 < \cdots < x_n$.

Proposition 2.1 has an explanation in terms of representation theory. Gelfand and Tsetlin [GT] show that semi-strict Gelfand patterns form bases of representations of $\mathrm{SL}(n)$, and one can deduce Proposition 2.1 from this fact using the Weyl dimension formula. (The Gelfand patterns in [GT] are not semi-strict, but there is an easy transformation that makes them so: Replace $m_{p,q}$ in equation (3) of [GT] with $m_{p,q} + (q - p)$ and then reflect the triangle through its vertical axis.) Thus, the novelty of our approach is not that one can count semi-strict Gelfand patterns, but rather that one can count tilings with prescribed behavior on a horizontal line (as in Theorem 2.2). Another proof of Proposition 2.1, and one that bypasses its representation-theoretic significance, is the article of Carlitz and Stanley [CS]. (That article does not deal directly with semi-strict Gelfand patterns, but it is easy to deduce Proposition 2.1 from the theorem proved there.)

Proof. Let $V(x_1, \ldots, x_n)$ be the number of semi-strict Gelfand patterns with top row $x_1, \ldots, x_n$. Given any such pattern, the second row must be of the form

$y_1, \ldots, y_{n-1}$ with $x_i \leq y_i < x_{i+1}$ for all i. Therefore,

$$V(x_1, \ldots, x_n) = \sum_{y_1=x_1}^{x_2}{}^{L} \sum_{y_2=x_2}^{x_3}{}^{L} \cdots \sum_{y_{n-1}=x_{n-1}}^{x_n}{}^{L} V(y_1, \ldots, y_{n-1}), \tag{2.1}$$

where the modified summation operator $\sum^L$ is defined by

$$\sum_{i=s}^{t}{}^{L} f(i) = \sum_{i=s}^{t-1} f(i).$$

The advantage to writing it this way is that

$$\sum_{i=r}^{s}{}^{L} f(i) + \sum_{i=s}^{t}{}^{L} f(i) = \sum_{i=r}^{t}{}^{L} f(i) \tag{2.2}$$

whenever $r < s < t$. There is a unique way to extend the the definition of

$$\sum_{i=s}^{t}{}^{L} f(i)$$

to the case $s > t$, if we require that (2.2) be true for all r, s, t. Then starting from the base relation $V(x_1) = 1$, we can use (2.1) to define $V(x_1, \ldots, x_n)$ for arbitrary integers $x_1, \ldots, x_n$ (not necessarily satisfying $x_1 < \cdots < x_n$).

We will prove the formula for $V(x_1, \ldots, x_n)$ by induction on n. It is clearly true for $n = 1$. Suppose that for all $y_1, \ldots, y_{n-1}$,

$$V(y_1, \ldots, y_{n-1}) = \prod_{1 \leq i < j \leq n-1} \frac{y_j - y_i}{j - i}. \tag{2.3}$$

Then $V(y_1, \ldots, y_{n-1})$ is a polynomial of total degree $(n-1)(n-2)/2$ in $y_1, \ldots, y_{n-1}$. When we put (2.3) into (2.1), we find that after each of the n summations in (2.1), the result is still a polynomial, and the degree increases by 1 each time. It is easy to check from (2.3) that the coefficient of $y_2^1 y_3^2 \ldots y_{n-1}^{n-2}$ in $V(y_1, \ldots, y_{n-1})$ is

$$\frac{1}{1!2! \ldots (n-2)!}.$$

From this and (2.1) it follows that the coefficient of $x_2^1 x_3^2 \ldots x_n^{n-1}$ in $V(x_1, \ldots, x_n)$ is

$$\frac{1}{1!2! \ldots (n-1)!}.$$

We have therefore shown that $V(x_1, \ldots, x_n)$ and

$$\prod_{1 \leq i < j \leq n} \frac{x_j - x_i}{j - i} \tag{2.4}$$

are polynomials in $x_1, \ldots, x_n$ of the same total degree, and with the same coefficient of $x_2^1 x_3^2 \ldots x_n^{n-1}$. If we can show that $V(x_1, \ldots, x_n)$ is divisible by (2.4), then they must be equal. Equivalently, we just need to show that $V(x_1, \ldots, x_n)$ is skew-symmetric in $x_1, \ldots, x_n$.

For convenience, let $(s_1, t_1; s_2, t_2; \ldots; s_{n-1}, t_{n-1})$ denote

$$\sum_{y_1=s_1}^{t_1}{}^{L} \sum_{y_2=s_2}^{t_2}{}^{L} \cdots \sum_{y_{n-1}=s_{n-1}}^{t_{n-1}}{}^{L} V(y_1, y_2, \ldots, y_{n-1}).$$

We want to show that $(x_1, x_2; x_2, x_3; \dots; x_{n-1}, x_n)$ is skew-symmetric in $x_1, \dots, x_n$.

To start off, notice that for all i,

$$(\dots; s_i, u_i; \dots) = (\dots; s_i, t_i; \dots) + (\dots; t_i, u_i; \dots). \tag{2.5}$$

Also, since $V(y_1, \dots, y_{n-1})$ is a skew-symmetric function of $y_1, \dots, y_{n-1}$ by (2.3), for $i < j$ we must have

$$(\dots; s_i, t_i; \dots; s_j, t_j; \dots) = -(\dots; s_j, t_j; \dots; s_i, t_i; \dots).$$

The relation $(\dots; s_i, t_i; \dots) = -(\dots; t_i, s_i; \dots)$ follows easily from the definition of $(\dots; s_i, t_i; \dots)$. From the last two relations, we see that $(\dots; s_i, t_i; \dots; s_j, t_j; \dots)$ vanishes if $s_i = s_j$ and $t_i = t_j$, or if $s_i = t_j$ and $s_j = t_i$.

To verify that $(x_1, x_2; \dots; x_{n-1}, x_n)$ is skew-symmetric in $x_1, \dots, x_n$, it suffices to check that it changes sign under transpositions of adjacent x_i's. We check the effect of exchanging x_{i+1} with x_{i+2} as follows. If we write $x'_k = x_{i+k}$ (to simplify the subscripts), we have

$$\begin{aligned}(\dots; x'_0, x'_2; x'_2, x'_1; x'_1, x'_3; \dots) \quad = \quad & (\dots; x'_0, x'_1; x'_2, x'_1; x'_1, x'_2; \dots)\\ &+ (\dots; x'_0, x'_1; x'_2, x'_1; x'_2, x'_3; \dots)\\ &+ (\dots; x'_1, x'_2; x'_2, x'_1; x'_1, x'_2; \dots)\\ &+ (\dots; x'_1, x'_2; x'_2, x'_1; x'_2, x'_3; \dots).\end{aligned}$$

by several applications of (2.5). All terms on the right except the second are 0, so

$$\begin{aligned}(\dots; x'_0, x'_2; x'_2, x'_1; x'_1, x'_3; \dots) &= (\dots; x'_0, x'_1; x'_2, x'_1; x'_2, x'_3; \dots)\\ &= -(\dots; x'_0, x'_1; x'_1, x'_2; x'_2, x'_3; \dots).\end{aligned}$$

Thus, exchanging x_{i+1} with x_{i+2} introduces a minus sign whenever $i \geq 1$ and $i + 3 \leq n$. The other cases (exchanging x_1 with x_2 or x_{n-1} with x_n) are easily dealt with in a similar way. Therefore, $V(x_1, \dots, x_n)$ is skew-symmetric in $x_1, \dots, x_n$, so as discussed above we must have

$$V(x_1, \dots, x_n) = \prod_{1 \leq i < j \leq n} \frac{x_j - x_i}{j - i},$$

as desired. □

Notice that after some manipulation of the product

$$V(1, 2, \dots, a, a + b + 1, a + b + 2, \dots, a + b + c),$$

Proposition 2.1 implies MacMahon's formula. However, our main application will be to counting tilings with prescribed behavior on horizontal lines.

Consider the k-th horizontal line from the top in an a, b, c hexagon. If $k < \min(a, c)$, then in every tiling there must be k vertical lozenges on the k-th line; if $\min(a, c) \leq k \leq (a + c)/2$, then there must be $\min(a, c)$ vertical lozenges on it. (As mentioned earlier, symmetry frees us from needing to treat the case $k > (a + c)/2$, so we will routinely assume $k \leq (a + c)/2$.) In either case, note that the number of vertical lozenges on the k-th row is $\min(k, a, c)$.

Define the function

$$V(x_1, \dots, x_n) = \prod_{1 \leq i < j \leq n} \frac{x_j - x_i}{j - i},$$

as in the proof of Proposition 2.1. Then we have the following formulas:

Theorem 2.2. *Suppose we require that the vertical lozenges bisected by the k-th horizontal line from the top in an a, b, c hexagon occur at hexagonal positions $1 \leq a_1 < a_2 < \cdots < a_\ell \leq b + \min(k, a)$ (and nowhere else), with $\ell = \min(k, a, c)$. If $k < \min(a, c)$, there are*

$$V(a_1, a_2, \ldots, a_\ell)V(1, 2, \ldots, a-k, a-k+a_1, \ldots, a-k+a_\ell, a+b+1, \ldots, a+b+c-k)$$

such tilings. If $a \leq c$ and $a \leq k \leq (a+c)/2$, there are

$$V(1, 2, \ldots, k-a, k-a+a_1, \ldots, k-a+a_\ell)V(a_1, a_2, \ldots, a_\ell, a+b+1, \ldots, a+b+c-k)$$

such tilings (and a similar formula applies if $c < a$ and $c \leq k \leq (a+c)/2$).

For the proof, simply notice that tilings of the parts of the hexagon above and below the k-th line correspond naturally to semi-strict Gelfand patterns with certain top rows, and then apply Proposition 2.1 to count them. In both cases, the two factors correspond to the parts of the tiling that lie above and below the cutting line.

3. Setting up the Functional

We now turn to the proofs of our main theorems. As is usually done in situations such as ours, where one seeks to establish a law of large numbers for some class of combinatorial objects, we approach the problem by trying to find the individually most likely outcome (in this case, the individually most likely behavior of the height function on a fixed horizontal line), and showing that outcomes that differ substantially from it are very unlikely, even considered in aggregate. We will begin in this section by setting up a functional to be maximized; the function that maximizes it will be a simple transformation of the average height function.

Our method is to focus on the locations of the vertical lozenges rather than the height function per se. The two are intimately related, because, as we move across the tiling from left to right, the (unnormalized) height decreases by 2 when we cross a vertical lozenge and increases by 1 when we fail to cross a vertical lozenge. Thus, in determining the likely locations of vertical lozenges, we will in effect be solving for the average height function. Theorem 2.2 tells us that we can count tilings with prescribed behavior on horizontal lines, so we will start off by taking the logarithm of the product formula in Theorem 2.2 and interpreting it as a Riemann sum for a double integral.

In fact, it will be convenient to look first at a more general product, and then apply our analysis of it to the product appearing in Theorem 2.2. Fix positive integers ℓ and n satisfying $\ell \leq n$, and non-negative integers n_L and n_R. We will try to determine the distribution of $(a_1, a_2, \ldots, a_\ell)$ satisfying $1 \leq a_1 < a_2 < \cdots < a_\ell \leq n$ that maximizes

$$V(1, 2, \ldots, n_L, n_L+a_1, n_L+a_2, \ldots, n_L+a_\ell, n_L+n+1, n_L+n+2, \ldots, n_L+n+n_R).$$

For convenience, let b_i denote the i-th element of the sequence $1, 2, \ldots, n_L, n_L + a_1, \ldots, n_L + a_\ell, n_L + n + 1, \ldots, n_L + n + n_R$.

Set $\rho_L = n_L/n$, $\lambda = \ell/n$, and $\rho_R = n_R/n$. We will work in the limit as $n \to \infty$, with ρ_L, λ, and ρ_R tending toward definite limits. For a more convenient way to keep track of $a_1, \ldots, a_\ell$ as we pass to the continuum limit, we define a weakly increasing function $A : [0, 1] \to [0, 1]$ as follows. Set $A(i/n) = \#\{j : a_j \leq i\}/n$ for $0 \leq i \leq n$, and then interpolate by straight lines between these points. Similarly, set $B(i/N) = \#\{j : b_j \le i\}/N$ for $0 \le i \le N$, where $N \overset{\text{def}}{=} n_L + n + n_R$, and then interpolate by straight lines. To simplify the notation later, set $\rho = \rho_L + \rho_R$, so that $N = (1+\rho)n$, $A(0) = 0$, $A(1) = \lambda$, $B(0) = 0$, and $B(1) = (\lambda+\rho)/(1+\rho)$.

The functions A and B satisfy the Lipschitz condition with constant 1; that is, $|A(s) - A(t)|$ and $|B(s) - B(t)|$ are bounded by $|s - t|$. Note that the derivatives A' and B' are not defined everywhere, but they are undefined only at isolated points, and where they are defined they equal either 0 or 1; when we make statements about A' and B', we will typically ignore the points of non-differentiability. We can also derive a simple relation between A' and B'. To do so, notice that for $0 \le i \le n$, we have

$$B\left(\frac{i+n_L}{N}\right) = \frac{\#\{j : b_j \le i + n_L\}}{N} = \frac{\#\{j : a_j \le i\}}{N} + \frac{n_L}{N} = A\left(\frac{i}{N} \cdot \frac{N}{n}\right)\frac{n}{N} + \frac{n_L}{N};$$

if we set $t = i/N$, we find that this equation becomes

$$B\left(t + \frac{\rho_L}{1+\rho}\right) = \frac{A\left((1+\rho)t\right)}{1+\rho} + \frac{\rho_L}{1+\rho}. \tag{3.1}$$

For $0 \le t \le 1/(1+\rho)$, the values of A and B occurring in (3.1) are defined by interpolating linearly between points at which we have just shown that (3.1) holds, so it must hold for all such t. Therefore, for $0 \le t \le 1/(1+\rho)$, we have

$$B'\left(t + \frac{\rho_L}{1+\rho}\right) = A'((1+\rho)t),$$

except at isolated points of non-differentiability. All other values of B' are 1, since it follows immediately from the definition of B that for $0 < t < \rho_L/(1+\rho)$ or $1 > t > (1+\rho_L)/(1+\rho)$, we have $B'(t) = 1$.

We have $a_i = nA^{-1}(i/n)$, $b_i = NB^{-1}(i/N)$. (Whenever we refer to $A^{-1}(t)$, we consider it to take the smallest value possible, to avoid ambiguity, and we interpret $B^{-1}(t)$ analogously.) When we take the logarithm and then multiply by n^{-2}, the double product

$$V(b_1, \dots, b_{n_L+\ell+n_R}) = \prod_{1 \le i < j \le n_L+\ell+n_R} \frac{b_j - b_i}{j - i}$$

ought to approach an integral like

$$(1+\rho)^2 \iint_{0 \le s < t \le \frac{\lambda+\rho}{1+\rho}} \log \frac{B^{-1}(t) - B^{-1}(s)}{t - s}\, ds\, dt.$$

(The factor of $(1+\rho)^2$ appears because $N = (1+\rho)n$ and we rescaled by n^{-2} instead of N^{-2}.) In the appendix, we will justify this claim rigorously, except with the function B replaced by a nicer function C. (The justification is not very difficult, but it is long enough that here it would be a distraction.)

The conclusion from the appendix is that

$$\frac{\log V(b_1, \dots, b_{n_L+\ell+n_R})}{n^2} = (1+\rho)^2 \iint_{0 \le s < t \le \frac{\lambda+\rho}{1+\rho}} \log \frac{C^{-1}(t) - C^{-1}(s)}{t - s}\, ds\, dt + o(1),$$

where C is a certain strictly increasing, continuous, piecewise linear function satisfying $C(0) = 0$, $C(1) = (\lambda+\rho)/(1+\rho)$, and $|B' - C'| = O(1/N)$. Note that C has a continuous inverse (unlike B, which is only weakly increasing and thus may not even have an inverse).

By symmetry, the integral equals

$$\frac{(1+\rho)^2}{2}\int_0^{\frac{\lambda+\rho}{1+\rho}}\int_0^{\frac{\lambda+\rho}{1+\rho}}\log\frac{C^{-1}(t)-C^{-1}(s)}{t-s}\,ds\,dt.$$

We can write the integral as

$$\begin{aligned}\frac{(1+\rho)^2}{2}\int_0^{\frac{\lambda+\rho}{1+\rho}}\int_0^{\frac{\lambda+\rho}{1+\rho}}\log\frac{|C^{-1}(t)-C^{-1}(s)|}{|t-s|}\,ds\,dt&\\ =\frac{(1+\rho)^2}{2}\int_0^{\frac{\lambda+\rho}{1+\rho}}\int_0^{\frac{\lambda+\rho}{1+\rho}}&\log|C^{-1}(t)-C^{-1}(s)|\,ds\,dt\\ -\frac{(1+\rho)^2}{2}\int_0^{\frac{\lambda+\rho}{1+\rho}}\int_0^{\frac{\lambda+\rho}{1+\rho}}&\log|t-s|\,ds\,dt;\end{aligned}$$

since the integral being subtracted is a constant, we can ignore it. (The individual integrands are unbounded, but since the singularities are merely logarithmic they do not interfere with integrability.) Letting $u=C^{-1}(s)$ and $v=C^{-1}(t)$, we can rewrite the part that matters as

$$\frac{(1+\rho)^2}{2}\int_0^1\int_0^1 C'(u)C'(v)\log|u-v|\,du\,dv.$$

Since $|C'-B'|=O(1/N)$, this integral differs by $o(1)$ from

$$\frac{(1+\rho)^2}{2}\int_0^1\int_0^1 B'(u)B'(v)\log|u-v|\,du\,dv. \tag{3.2}$$

We will now use our formula expressing B' in terms of A'. Recall that

$$B'\left(\frac{t+\rho_L}{1+\rho}\right)=A'(t)$$

for $0\le t\le 1$, and that $B'(t)=1$ for $t\in(-\rho_L,0)\cup(1,1+\rho_R)$. To take advantage of this, we change variables to s and t (which are different from the s and t used earlier in the article) with $u=(s+\rho_L)/(1+\rho)$ and $v=(t+\rho_L)/(1+\rho)$. Then when $0\le s,t\le 1$ we have $B'(u)=A'(s)$ and $B'(v)=A'(t)$; elsewhere B' is 1. Thus, (3.2) is equal to

$$\frac{1}{2}\int_{-\rho_L}^{1+\rho_R}\int_{-\rho_L}^{1+\rho_R}(A'(s)+I(s))(A'(t)+I(t))\log\left|\frac{s-t}{1+\rho}\right|\,ds\,dt, \tag{3.3}$$

where I is the characteristic function of $[-\rho_L,0]\cup[1,1+\rho_R]$, and where we set $A'=0$ outside $[0,1]$.

Removing the $1+\rho$ from the denominator of the argument of the logarithm simply adds

$$\frac{1}{2}\int_{-\rho_L}^{1+\rho_R}\int_{-\rho_L}^{1+\rho_R}(A'(s)+I(s))(A'(t)+I(t))\log|1+\rho|\,ds\,dt$$

to the integral; this quantity is a constant because the only occurrence of A in it is through the integral

$$\int_{-\rho_L}^{1+\rho_R}A'(s)\,ds=A(1)-A(0)=\lambda$$

(and the square of this integral). We can also change the range of integration in (3.3) to the entire plane (since the integrand has support only in the rectangle $[-\rho_L, 1+\rho_R] \times [-\rho_L, 1+\rho_R]$). Thus, we have arrived at the result that, for some irrelevant constant K, $n^{-2}\log V(b_1, \dots, b_{n_L+\ell+n_R})$ equals

$$\frac{1}{2}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}(A'(s)+I(s))(A'(t)+I(t))\log|s-t|\,ds\,dt + K + o(1). \tag{3.4}$$

We can now apply this to Theorem 2.2. Suppose that on the k-th line from the top in the a, b, c hexagon (with $k \le (a+c)/2$), the vertical lozenges have hexagonal positions $1 \le a_1 < a_2 < \cdots < a_\ell \le n$, where $n \stackrel{\text{def}}{=} b + \min(k, a, c)$. Define the function A as above. Then our analysis so far, combined with Theorem 2.2, shows that if we take the logarithm of the number of tilings with the given behavior on the k-th line, and divide by n^2, then we get the sum of two terms of the form (3.4); for, Theorem 2.2 gives us a product of two V-expressions (whose exact nature depends on whether k lies between a and c), and when we take logarithms and divide by n^2, we get two terms, each of which is half of a double integral (plus negligibly small terms and irrelevant constants).

To put this into an appropriately general context, define the bilinear form $\langle\, ,\rangle$ by

$$\langle f, g\rangle = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f'(x)g'(y)\log|x-y|\,dx\,dy \tag{3.5}$$

for suitable functions f and g (for our purposes, functions such that their derivatives exist almost everywhere, are bounded, are integrable, and have compact support). We will now use this notation to continue the analysis begun in the previous paragraph.

In this paragraph we will systematically omit additive constants and $o(1)$ terms, since they would be a distraction. If $k < \min(a, c)$, then one of the two terms derived from Theorem 2.2 is $\frac{1}{2}\langle A, A\rangle$, and the other is $\frac{1}{2}\langle A + J_0, A + J_0\rangle$, where J_0 is a continuous function with derivative equal to the characteristic function of $[-|a-k|/n, 0] \cup [1, 1 + |c-k|/n]$. If $\min(a, c) \le k \le (a+c)/2$, then one term is $\frac{1}{2}\langle A + J_1, A + J_1\rangle$ and the other is $\frac{1}{2}\langle A + J_2, A + J_2\rangle$, where the derivative of J_1 is the characteristic function of $[-|a-k|/n, 0]$ and that of J_2 is the characteristic function of $[1, 1 + |c-k|/n]$.

Now a few simple algebraic manipulations bring these results to the following form.

Proposition 3.1. *Let $n = b + \min(k, a, c)$, $\rho_L = |a-k|/n$, and $\rho_R = |c-k|/n$. Then the logarithm of the number of tilings of an a, b, c hexagon with vertical lozenges at hexagonal positions $a_1, \dots, a_\ell$ (and nowhere else) along the k-th line (where $\ell = \min(k, a, c)$ and $1 \le a_1 < \cdots < a_\ell \le n$), when divided by n^2, equals*

$$\langle A + J, A + J\rangle + \text{constant} + o(1),$$

where J is any continuous function whose derivative is half the characteristic function of $[-\rho_L, 0] \cup [1, 1 + \rho_R]$, and A is defined (as earlier) by interpolating linearly between the values $A(i/n) = \#\{j : a_j \le i\}/n$ for $0 \le i \le n$.

Recall that we are working in the limit as $n \to \infty$ with ρ_L, ρ_R, and λ converging towards fixed limits. It is easy to check that the unspecified constant in Proposition 3.1 converges as well. We would like everything to be stated in terms of the limiting values of ρ_L, ρ_R, and λ. Replacing ρ_L and ρ_R in the definition of J by their limits changes $\langle A+J, A+J\rangle$ by only $o(1)$, and we can increase A' by $o(1)$ in such a way that $A(0)$ remains 0, $A(1)$ becomes the limiting value of λ, and A still satisfies the Lipschitz condition. Thus, we can let ρ_L, ρ_R, and λ denote their limiting values from now on.

We have now re-framed our problem. We must find a function A that maximizes $\mathcal{V}(A) \stackrel{\text{def}}{=} \langle A+J, A+J\rangle$, subject to certain conditions. We will look at real-valued functions A on $[0,1]$ that are continuous, weakly increasing, and subject to the following constraints: $A(0)=0$, $A(1)=\lambda$, and A must satisfy a Lipschitz condition with constant 1 (so $0 \le A' \le 1$ where A' is defined). For convenience, define $A(t)=0$ for $t<0$ and $A(t)=\lambda$ for $t>1$. Call a function A that satisfies these conditions *admissible*. Clearly, the functions A considered in this section are admissible. We will show in the next section that there is a unique admissible function A that maximizes $\mathcal{V}(A)$. (Notice that every admissible function A is differentiable almost everywhere, and A' is integrable and has compact support; for a proof of the necessary facts from real analysis, see Theorem 7.18 of [R]. Thus, $\mathcal{V}(A)$ makes sense for every admissible A.)

4. Analyzing the Functional

Let $\mathcal{F}$ be the set of admissible functions. We can topologize $\mathcal{F}$ using the sup norm, L^1 norm, or L^2 norm on $[0,1]$; it is straightforward to show that they all give the same topology, and that $\mathcal{F}$ is compact. In this section, we will show that $\mathcal{V}$ is a continuous function on $\mathcal{F}$, so it must attain a maximum. We will show furthermore that there is a unique function $A \in \mathcal{F}$ such that $\mathcal{V}(A)$ is maximal.

The proof will use several useful formulas for the bilinear form $\langle\,,\rangle$ (formulas (4.1) to (4.3)). These formulas are derived in [LS]; we repeat the derivations here for completeness. One can find similar analysis in [VK1] and [VK2].

The formulas are stated in terms of the Fourier and Hilbert transforms. For sufficiently well-behaved functions f, define the Fourier transform $\widehat{f}$ of f by

$$\widehat{f}(x) = \int_{-\infty}^{\infty} f(t)e^{-ixt}\,dt$$

and the Hilbert transform $\widetilde{f}$ by

$$\widetilde{f}(x) = \frac{1}{\pi}\int_{-\infty}^{\infty} \frac{f(t)}{x-t}\,dt$$

(which we make sense of by taking the Cauchy principal value). Note that the Fourier transform of $\widetilde{f}$ is $x \mapsto i(\operatorname{sgn} x)\widehat{f}(x)$, and that of f' is $x \mapsto ix\widehat{f}(x)$; multiplying by $i(\operatorname{sgn} x)$ and multiplying by ix commute with each other, so differentiation commutes with the Hilbert transform.

Integration by parts with respect to y in (3.5) shows that

$$\langle f, g\rangle = \pi \int_{-\infty}^{\infty} f'(x)\widetilde{g}(x)\,dx. \tag{4.1}$$

When done with respect to x, it also shows that

$$\langle f, g\rangle = -\pi \int_{-\infty}^{\infty} f(x)\widetilde{g}'(x)\,dx. \tag{4.2}$$

Unfortunately, Hilbert transforms are not always defined. For our purposes, it is enough to note that (4.1) makes sense and is true when f' and g have compact support, and similarly that (4.2) holds when f and g' have compact support.

If we set $g = f$ and apply Parseval's identity to (4.1), we find that when f has compact support,

$$\langle f, f\rangle = -\frac{1}{2}\int_{-\infty}^{\infty} |\widehat{f}(x)|^2 |x|\, dx. \tag{4.3}$$

Thus, the bilinear form $\langle\, ,\rangle$ is negative definite (on functions of compact support).

We can now prove easily that $\mathcal{V}$ is a continuous function on $\mathcal{F}$. To do so, notice that the definition of $\mathcal{V}$ and (4.1) imply that

$$\mathcal{V}(A_1) - \mathcal{V}(A_2) = \langle A_1 + A_2 + 2J, A_1 - A_2\rangle = \pi \int_{-\infty}^{\infty} f_1'(x)\widetilde{f}_2(x)\, dx,$$

where $f_1 = A_1 + A_2 + 2J$ and $f_2 = A_1 - A_2$. Thus, since $|f_1'(x)| \le 2$ for all x and $f_1' = 0$ outside some interval I not depending on A_1 and A_2,

$$|\mathcal{V}(A_1) - \mathcal{V}(A_2)| \le 2\pi \int_I |\widetilde{f}_2| \ll \left(\int_I |\widetilde{f}_2|^2\right)^{1/2} = ||\widetilde{f}_2||_2.$$

(The second bound follows from applying the Cauchy-Schwartz inequality to $|\widetilde{f}_2|\cdot 1$.)

It is known (see Theorem 90 of [Ti]) and easy to prove (combine Parseval's identity with the formula for the Fourier transform of a Hilbert transform) that $||\widetilde{f}_2||_2 = ||\widehat{f}_2||_2 = ||f_2||_2$. Thus, $|\mathcal{V}(A_1) - \mathcal{V}(A_2)| = O(||A_1 - A_2||_2)$, so the function $\mathcal{V}$ is continuous on $\mathcal{F}$.

Because $\mathcal{F}$ is compact, $\mathcal{V}$ must attain a maximum on $\mathcal{F}$. Now we apply the identity

$$\frac{\mathcal{V}(A_1) + \mathcal{V}(A_2)}{2} = \mathcal{V}\left(\frac{A_1 + A_2}{2}\right) + \left\langle \frac{A_1 - A_2}{2}, \frac{A_1 - A_2}{2}\right\rangle,$$

which is a form of the polarization identity for quadratic forms. Because $(A_1 - A_2)/2$ has compact support, (4.3) implies that

$$\frac{\mathcal{V}(A_1) + \mathcal{V}(A_2)}{2} \le \mathcal{V}\left(\frac{A_1 + A_2}{2}\right),$$

with equality if and only if $A_1 = A_2$. Thus, two different admissible functions could not both maximize $\mathcal{V}$, since then their average would give an even larger value. Therefore, there is a unique admissible function that maximizes $\mathcal{V}$. Let $\mathbb{A}$ be that function. (Notice that $\mathbb{A}$ depends on λ, ρ_L, and ρ_R, and hence on α, β, and γ, although our notation does not reflect this dependence.)

We are now almost at the point of being able to prove that there is a function $\mathcal{H}_{\alpha,\beta,\gamma}$ such that Theorem 1.2 holds (except for the part relating $\mathcal{H}_{\alpha,\beta,\gamma}$ to the explicitly given function $\mathcal{P}_{\alpha,\beta,\gamma}$). First, we need to relate A to the normalized average height function.

Assume that $k/(a + c) \to \kappa$ as $a, b, c \to \infty$ for some κ satisfying $0 \le \kappa \le 1$. Choose normalized coordinates for the a, b, c hexagon so that the k-th horizontal line from the top has normalized length 1, and in particular coordinatize that line so that its left endpoint is 0 and its right endpoint is 1; equivalently, coordinatize the α, β, γ hexagon so that the horizontal line that cuts it proportionately κ of the way from its upper border has length 1. (The truth or falsity of Theorem 1.2 is clearly unaffected by our choice of coordinates.) We can then identify the scaling factor σ with n, to within a factor of $1 + o(1)$. Given a tiling of the a, b, c hexagon, if we scan to the right along this line, whenever we cross a vertical lozenge the normalized height function decreases by $2/n$, and whenever we cross a location that could hold a vertical lozenge but does not the normalized height function increases by $1/n$. It follows that the normalized height function at location t is given by

$$-2A(t) + (t - A(t)) = t - 3A(t)$$

plus the value at $t = 0$, since this function changes by the same amount as the normalized height function does as one moves to the right.

Let $\varepsilon > 0$. Then there exists a $\delta > 0$ such that $|\mathcal{V}(\mathbb{A}) - \mathcal{V}(A)| < \delta$ implies $\sup_t |\mathbb{A}(t) - A(t)| < \varepsilon$. (This claim holds for every continuous function on a compact space that takes its maximal value at a unique point.) For n sufficiently large, Proposition 3.1 implies that if $\sup_t |\mathbb{A}(t) - A(t)| \geq \varepsilon$, then in a random tiling, every behavior within $o(1)$ of $\mathbb{A}$ is at least $e^{n^2(\delta + o(1))}$ times as likely to occur along the k-th line as the behavior A is. Since the number of possibilities for A is only exponential in n, the probability that $\sup_t |\mathbb{A}(t) - A(t)| \geq \varepsilon$ is exponentially small in n^2 (and hence in σ^2). In other words, the probability that a random height function differs along this line from the height function obtained from $\mathbb{A}$ is exponentially small in σ^2. It follows that the same is true without the restriction to the particular horizontal line, because of the Lipschitz constraint on height functions: if we show that large deviations are unlikely on a sufficiently dense (but finite) set of horizontal lines, then the same holds even between them. Thus, we have proved Theorem 1.2, except for the connection between $\mathcal{H}_{\alpha,\beta,\gamma}$ and $\mathcal{P}_{\alpha,\beta,\gamma}$.

Furthermore, the density of vertical lozenges near location t along a given horizontal line is almost always approximately equal to $\mathbb{A}'(t)$. We can make this claim precise and justify it as follows. Given a random tiling, $A(t)$ gives the number of vertical lozenges to the left of the location t, divided by n (plus $O(n^{-1})$ if t is not at a vertex of the underlying triangular lattice). Thus, the number of vertical lozenges in an interval $[a, b]$ is $n(A(b) - A(a)) + O(1)$. We have seen that the probability that this quantity will differ by more than εn from $n(\mathbb{A}(b) - \mathbb{A}(a))$ is exponentially small in n^2. Therefore, as $n \to \infty$ (equivalently, $\sigma \to \infty$), the expected value of $A(b) - A(a)$ is $\mathbb{A}(b) - \mathbb{A}(a) + o(1)$. Thus, we get the expected number of vertical lozenges in $[a, b]$, which is also equal to the expected number of vertical lozenges in (a, b) (up to a negligible error). Notice that the $o(1)$ error term is uniform in the choice of a, b, and the horizontal line, because the probability of a large deviation in height anywhere in the hexagon is small.

If we take the result we have just proved for horizontal line segments and integrate it over the horizontal line segments that constitute the interior of any smooth simple closed curve, then we can conclude that Theorem 1.1 holds, except for the explicit determination of $\mathcal{P}_{\alpha,\beta,\gamma}$ (which is equivalent to the explicit determination of $\mathbb{A}$, since $\mathcal{P} = \mathbb{A}'$). Also, notice that our method of proof implies that $\mathcal{H}_{\alpha,\beta,\gamma}$ and $\mathcal{P}_{\alpha,\beta,\gamma}$ must satisfy

$$\frac{\partial \mathcal{H}_{\alpha,\beta,\gamma}(x, y)}{\partial x} = 1 - 3\mathcal{P}_{\alpha,\beta,\gamma}(x, y),$$

as desired (although it does not yet determine them explicitly). Thus, all that remains to be done is to determine the maximizing function $\mathbb{A}$ explicitly. We will do so in Section 5.

5. The Typical Height Function

Unfortunately, it is not clear how to find the admissible function $\mathbb{A}$ that maximizes $\mathcal{V}(\mathbb{A})$. Ordinary calculus of variations techniques will not produce admissible solutions. However, we will see that techniques similar to those used in [LS] and [VK1, VK2] can be used to verify that a function $\mathbb{A}$ maximizes $\mathcal{V}(\mathbb{A})$, if we can guess $\mathbb{A}$. (It is not clear a priori that the techniques will work, but fortunately everything works out just as one would hope.)

As we saw in Section 4, for the cases that are needed in the proof of Theorem 1.1, an explicit formula for $\mathbb{A}$ is equivalent to one for $\mathcal{P}_{\alpha,\beta,\gamma}$; since we know already what the answer should be, guessing it will not present a problem. In Section 1, we tried to give some motivation by presenting a slightly simpler description of $\mathcal{P}_{\alpha,\beta,\gamma}$ than the explicit formula. However, we do not know of any straightforward way to guess the answer from scratch. We arrived at it partly by analogy with the arctangent formula for random domino tilings of Aztec diamonds (Theorem 1 of [CEP]), partly on the basis of symmetry and simplicity, and partly on the basis of numerical evidence.

To avoid unnecessarily complicated notation, we will solve the problem in greater generality than is needed simply for Theorem 1.1. We will deal with the case of arbitrary λ satisfying $0 < \lambda < 1$, and arbitrary non-negative ρ_L and ρ_R (which we assume for simplicity are strictly greater than 0). We will use the same notation as earlier in the paper; for example, we set $\rho = \rho_L + \rho_R$. Of course, guessing the admissible function that maximizes the functional in general requires additional effort, but the symmetry and elegance of the general formulas are a helpful guide.

We will express the maximizing function $\mathbb{A}$ in terms of auxiliary functions f_1 and f_2. Define

$$f_1(t) = 2t(1-t) - (\lambda^2 + \rho\lambda - \rho_R)t - (\lambda^2 + \rho\lambda - \rho_L)(1-t)$$

and

$$f_2(t) = (\rho + 2\lambda)^2 t(1-t) - (\lambda^2 + \rho\lambda - \rho_R)^2 t - (\lambda^2 + \rho\lambda - \rho_L)^2(1-t).$$

(Note that both expressions are invariant under $(t, \rho_L, \rho_R) \leftrightarrow (1-t, \rho_R, \rho_L)$. This observation reduces some of the work involved in verifying the claims that follow.) Since the discriminant of $f_2(t)$ is

$$16\lambda(1-\lambda)(\lambda + \rho_R)(\lambda + \rho_L)(\lambda + \rho)(\lambda + \rho + 1),$$

$f_2(t)$ has distinct real roots $r_1 < r_2$. We can show that both roots are in $[0, 1]$ as follows. Since $f_2(0)$ and $f_2(1)$ are at most 0, both roots of f_2 lie in $[0, 1]$ if the point at which f_2 achieves its maximum does. The maximum occurs when

$$t = \frac{(1-\lambda)\rho_L^2 + \lambda\rho_R^2 + 2\lambda^2 + (2-\lambda)\lambda\rho_L + (2+\lambda)\lambda\rho_R + \rho_L\rho_R}{(\rho_L + \rho_R + 2\lambda)^2},$$

and this point is easily seen to lie in $[0, 1]$.

We will specify the function $\mathbb{A}$ by specifying its derivative $\mathbb{A}'$, which together with the condition $\mathbb{A}(0) = 0$ uniquely determines the function. (We will then check that the newly defined $\mathbb{A}$ maximizes $\mathcal{V}$, and is thus the same as the previous $\mathbb{A}$.) For $t \in (r_1, r_2)$ define

$$\mathbb{A}'(t) = \frac{1}{\pi}\cot^{-1}\left(\frac{f_1(t)}{\sqrt{f_2(t)}}\right).$$

For $t \in [0, r_1]$ define $\mathbb{A}'(t) = \lim_{t \to r_1+} \mathbb{A}'(t)$, and similarly for $t \in [r_2, 1]$ define $\mathbb{A}'(t) = \lim_{t \to r_2-} \mathbb{A}'(t)$. We can show that the first limit will be 0 or 1 if $r_1 \in (0,1)$, and the second will be 0 or 1 if $r_2 \in (0,1)$; to verify this, it suffices to check (using resultants, for example) that f_1 and f_2 cannot have a common root in $(0,1)$, from which it follows that at r_1 or r_2 the denominator of the argument of the arccotangent vanishes without the numerator vanishing. Notice that $\mathbb{A}'(0) = 0$ if $f_1(r_1) > 0$, and $\mathbb{A}'(0) = 1$ if $f_1(r_1) < 0$. Similarly, $\mathbb{A}'(1) = 0$ if $f_1(r_2) > 0$, and $\mathbb{A}'(1) = 1$ if $f_1(r_2) < 0$. Also, if $r_1 = 0$, then f_1 and f_2 both vanish at 0, and it follows that $\lim_{t \to 0+} \mathbb{A}'(t) = \frac{1}{2}$; similarly, if $r_2 = 1$, then $\lim_{t \to 1-} \mathbb{A}'(t) = \frac{1}{2}$.

Let $\mathbb{A} : [0,1] \to [0,1]$ be the unique function satisfying $\mathbb{A}(0) = 0$ with derivative $\mathbb{A}'$. We will show later in this section that $\mathbb{A}(1) = \lambda$, from which it follows that $\mathbb{A}$ is an admissible function (since the other conditions are clearly satisfied). (We then extend $\mathbb{A}$ to a function on all of $\mathbb{R}$ in the usual way, so that $\mathbb{A}(t) = \lambda$ for $t > 1$ and $\mathbb{A}(t) = 0$ for $t < 0$.) We will prove that $\mathbb{A}$ is the unique admissible function such that $\mathcal{V}(\mathbb{A})$ is maximal.

Before beginning the proof, it is helpful for motivation to examine what the calculus of variations tells us about how the maximizing function should behave. For every admissible function A we have

$$\mathcal{V}(A) = \langle A + J, A + J \rangle,$$

where J is any continuous function whose derivative is half the characteristic function of $[-\rho_L, 0] \cup [1, 1 + \rho_R]$. Suppose we perturb our function $\mathbb{A}$ by adding to it a small bump centered at t, which we write as $\varepsilon \delta_t$ where δ_t is a delta function. (We should actually perturb by a bump rather than a spike, in order to try to maintain the Lipschitz constraint as long as ε is small enough. However, treating δ_t as a delta function gives the right answer.) Because

$$\langle \mathbb{A} + J + \varepsilon\delta_t, \mathbb{A} + J + \varepsilon\delta_t \rangle = \langle \mathbb{A} + J, \mathbb{A} + J \rangle + 2\varepsilon \langle \delta_t, \mathbb{A} + J \rangle + O(\varepsilon^2),$$

the first variation of $\mathcal{V}$ is $2\langle \delta_t, \mathbb{A} + J \rangle$. By (4.2),

$$\langle \delta_t, \mathbb{A} + J \rangle = -\pi \int_{-\infty}^{\infty} \delta_t(x) \left(\widetilde{\mathbb{A}'}(x) + \widetilde{J}'(x) \right) dx = -\pi \left(\widetilde{\mathbb{A}'}(t) + \widetilde{J}'(t) \right).$$

Thus, in order for the first variation to vanish, we must have $\widetilde{\mathbb{A}'}(t) + \widetilde{J}'(t) = 0$. When $\mathbb{A}'(t) \in (0,1)$ this must be the case, assuming $\mathbb{A}$ maximizes $\mathcal{V}(\mathbb{A})$. However, when $\mathbb{A}'(t) \in \{0,1\}$, every perturbation violates the Lipschitz constraint, and we can conclude nothing.

Our strategy for proving that $\mathbb{A}$ maximizes $\mathcal{V}(\mathbb{A})$ will take advantage of the vanishing of $\widetilde{\mathbb{A}'}(t) + \widetilde{J}'(t)$ (which we will prove directly later in this section). To begin, for any admissible A we write

$$\mathcal{V}(A) - \mathcal{V}(\mathbb{A}) = \langle A - \mathbb{A}, A - \mathbb{A} \rangle + 2\langle A - \mathbb{A}, \mathbb{A} + J \rangle.$$

Because $\langle A - \mathbb{A}, A - \mathbb{A} \rangle \le 0$ with equality if and only if $A = \mathbb{A}$, in order to prove that $\mathcal{V}(A) \le \mathcal{V}(\mathbb{A})$, we need only show that

$$2\langle A - \mathbb{A}, \mathbb{A} + J \rangle \le 0.$$

To show that this is the case, we start by using (4.2), which tells us that

$$\langle A - \mathbb{A}, \mathbb{A} + J \rangle = -\pi \int_{-\infty}^{\infty} (A(t) - \mathbb{A}(t))(\widetilde{\mathbb{A}'}(t) + \widetilde{J}'(t))\, dt.$$

Thus, we want to show that

$$\int_{-\infty}^{\infty}(A(t)-\mathbb{A}(t))(\widetilde{\mathbb{A}'}(t)+\widetilde{J'}(t))\,dt\geq 0. \tag{5.1}$$

We will prove that the integrand is everywhere non-negative, by showing that $\widetilde{\mathbb{A}'}(t)+\widetilde{J'}(t)=0$ when $t\in(r_1,r_2)$ (the interval where $\mathbb{A}'(t)\in(0,1)$), and that in the rest of $(0,1)$ its sign is the same as that of $A(t)-\mathbb{A}(t)$. (Outside $(0,1)$, $A(t)=\mathbb{A}(t)$.)

In order to prove these facts, we will apply the following theorem. For a proof, see Theorem 93 of [Ti, p. 125].

Theorem 5.1. *Let Φ be a holomorphic function on the upper half plane, such that the integrals*

$$\int_{-\infty}^{\infty}|\Phi(x+iy)|^2\,dx$$

exist for all $y>0$ and are bounded. Then there exists a function F on the real line such that for almost all x, $\Phi(x+iy)\to F(x)$ as $y\to 0+$, and the imaginary part of F is the Hilbert transform of the real part of F.

We will apply this theorem to determine the Hilbert transform of $\mathbb{A}'+J'$. To prepare for the application of the theorem, we begin by defining, for $t\in(r_1,r_2)$,

$$g(t)=\frac{f_1(t)}{\sqrt{f_2(t)}}$$

and

$$\Phi(t)=\frac{1}{\pi}\cot^{-1}g(t). \tag{5.2}$$

(Of course, $\Phi(t)=\mathbb{A}'(t)$ on (r_1,r_2), but this new notation will help avoid confusion soon.) Then g extends to a unique holomorphic function on the (open) upper half plane. The function Φ extends as well to a unique holomorphic function on the upper half plane, together with all of $\mathbb{R}$ except the points $-\rho_L$, 0, r_1, r_2, 1, and $1+\rho_R$. To see why, notice that $g(t)^2+1$ has only four roots, in particular, simple roots at each of $-\rho_L$, 0, 1, and $1+\rho_R$. There is always a holomorphic branch of the arccotangent of a holomorphic function on a simply-connected domain, as long as that function does not take on the values $\pm i$; this fact gives us the analytic continuation of Φ. Of course, $\cot\pi\Phi(t)=g(t)$ for all t in the upper half plane.

For real t (except $-\rho_L$, 0, r_1, r_2, 1, and $1+\rho_R$), define

$$F(t)=\lim_{s\to 0+}\Phi(t+is)=\Phi(t).$$

(We will use this notation to distinguish between the function F on the real line and the function Φ on the upper half plane.) In order to apply Theorem 5.1, we must determine the real and imaginary parts of $F(t)$. Outside of (r_1,r_2), $\operatorname{Re}F(t)$ is piecewise constant (in particular, constant between the points where F is undefined) since $F'(t)$ is imaginary there. The integrability of $F'(t)$ at $t=r_1$ and $t=r_2$ implies that $F(t)$ is continuous there, which implies that $\operatorname{Re}F(t)=\mathbb{A}'(t)$ for all $t\in(0,1)$.

To determine the behavior of $\operatorname{Re}F(t)$ for $t\notin(0,1)$, we just have to see how much it changes by at $-\rho_L$, 0, 1, and ρ_R, since it is constant on $(-\infty,-\rho_L)$, $(-\rho_L,0)$, $(1,1+\rho_R)$, and $(1+\rho_R,\infty)$. To do so, notice that if Φ' has a pole with residue r at a point on the real axis, then F changes by $-r\pi i$ as one moves from the left of that point to its right. (To see this, integrate over a small semi-circle in the upper half plane, centered at the point.) If $g(u) = \pm i$, then

$$\begin{aligned}\mathrm{Res}_{t=u}\Phi'(t) &= \frac{1}{\pi}\lim_{t\to u} -\frac{(t-u)g'(t)}{1+g(t)^2}\\ &= \frac{1}{\pi}\lim_{t\to u} -\frac{(t-u)g'(t)}{(g(t)+g(u))(g(t)-g(u))}\\ &= -\frac{1}{2\pi g(u)}.\end{aligned}$$

Therefore, if $g(u) = \pm i$, then $\mathrm{Re}\, F$ changes by $\pm\frac{1}{2}$ from the left of u to its right.

To determine the precise sign of $g(u)$ when $g(u) = \pm i$, we will need to know how $1/\sqrt{f_2(t)}$ behaves when analytically continued through the upper half plane. We know that it is positive on (r_1, r_2). If one analytically continues it along any path through the upper half plane that starts in (r_1, r_2) and ends on the real axis to the left of r_1, then the result is a negative imaginary number (i.e., one with argument $-\pi i/2$). Similarly, if the path ends to the right of r_2, then the result is a positive imaginary number.

Thus, $g(-\rho_L) = -i \,\mathrm{sgn}\, f_1(-\rho_L)$. In fact, $g(-\rho_L) = i$, because

$$f_1(-\rho_L) = -\lambda^2 - \rho\lambda - \rho_L\rho_R - \rho_L - \rho_L^2 < 0.$$

It follows that $\mathrm{Re}\, F$ increases by $\frac{1}{2}$ at $-\rho_L$. Similarly, $\mathrm{Re}\, F$ decreases by $\frac{1}{2}$ at $1+\rho_R$.

The analysis at 0 and 1 is slightly more subtle. We have $g(0) = -i\,\mathrm{sgn}\, f_1(0)$, and $g(1) = i\,\mathrm{sgn}\, f_1(1)$. It turns out that $\mathrm{sgn}\, f_1(0) = \mathrm{sgn}\, f_1(r_1)$ and $\mathrm{sgn}\, f_1(1) = \mathrm{sgn}\, f_1(r_2)$. To prove this claim, we will deal with $\mathrm{Im}\, F$. (The results about $\mathrm{Im}\, F$ will be needed later, so this approach is worthwhile even though one might wish for a direct proof.)

It is impossible for $\mathrm{Im}\, F(t)$ to vanish for $t \in (0, r_1) \cup (r_2, 1)$, since $\mathrm{Re}\, F(t)$ is 0 or 1 for such t, but $F(t)$ cannot be 0 or 1 (since otherwise g would have a singularity at t, as one can see from (5.2)). Thus, the sign of $\mathrm{Im}\, F(t)$ is constant for t in each of $(0, r_1)$ and $(r_2, 1)$.

The imaginary part of the arccotangent does not depend on the branch used (since the values of the arccotangent always differ by a multiple of π). To determine the sign of the imaginary part of $F(t)$, we will use the fact that for real u with $|u| > 1$,

$$\mathrm{sgn}\,\mathrm{Im}\,\cot^{-1}(ui) = -\mathrm{sgn}\, u. \tag{5.3}$$

In order to apply this formula to F, we need to determine the sign of g on the axis. We determined above how $1/\sqrt{f_2(t)}$ behaves. Since

$$\Phi(t) = \frac{1}{\pi}\cot^{-1}\left(\frac{f_1(t)}{\sqrt{f_2(t)}}\right),$$

we find, by combining the facts about $1/\sqrt{f_2(t)}$ with (5.3), that $\mathrm{sgn}\,\mathrm{Im}\, F(t) = \mathrm{sgn}\, f_1(t)$ for $t \in (0, r_1)$, and $\mathrm{sgn}\,\mathrm{Im}\, F(t) = -\mathrm{sgn}\, f_1(t)$ for $t \in (r_2, 1)$.

Because $\mathrm{sgn}\,\mathrm{Im}\, F$ is constant, we can deduce two important facts. First, we see that $\mathrm{sgn}\, f_1$ must be constant on $(0, r_1)$ and $(r_2, 1)$. Notice that in fact it is constant on $[0, r_1]$ and $[r_2, 1]$, because we showed earlier that it cannot vanish at one of the endpoints of one of these intervals unless the interval consists only of one point. Second, we see that $\operatorname{sgn}\operatorname{Im} F(t) = \operatorname{sgn} f_1(r_1)$ for $t \in (0, r_1)$ and $\operatorname{sgn}\operatorname{Im} F(t) = -\operatorname{sgn} f_1(r_2)$ for $t \in (r_2, 1)$.

Thus, having taken a short detour, we can see that $g(0) = -i \operatorname{sgn} f_1(r_1)$, and $g(1) = i \operatorname{sgn} f_1(r_2)$. It follows that $\operatorname{Re} F$ increases by $\frac{1}{2}$ at 0 iff $f_1(r_1) < 0$, and decreases by $\frac{1}{2}$ at 0 iff $f_1(r_1) > 0$. Notice that these are exactly the conditions under which $\mathbb{A}'(0)$ is 1 or 0, respectively. Similarly, $\operatorname{Re} F$ increases by $\frac{1}{2}$ at 1 iff $f_1(r_2) > 0$, and decreases by $\frac{1}{2}$ at 1 iff $f_1(r_2) < 0$, and these are exactly the conditions under which $\mathbb{A}'(1)$ is 0 or 1, respectively.

The information that we have obtained determines $\operatorname{Re} F$, and in fact shows that

$$\operatorname{Re} F(t) = \begin{cases} 0 & \text{if } t < -\rho_L, \\ \frac{1}{2} & \text{if } -\rho_L < t < 0, \\ \mathbb{A}'(t) & \text{if } 0 < t < 1 \text{ (and } t \neq r_1, r_2), \\ \frac{1}{2} & \text{if } 1 < t < 1 + \rho_R, \text{ and} \\ 0 & \text{if } 1 + \rho_R < t. \end{cases}$$

In other words, $\operatorname{Re} F = \mathbb{A}' + J'$.

We would like to apply Theorem 5.1 to conclude that the Hilbert transform of $\mathbb{A}' + J'$ must be $\operatorname{Im} F$. For the hypotheses of the theorem to be satisfied, we need some integral estimates, in particular that the integrals

$$\int_{-\infty}^{\infty} |\Phi(r + is)|^2 \, dr$$

exist for all $s > 0$ and are bounded. To prove existence of the integral, we use the estimate $\Phi(t) = O(t^{-1})$ as $t \to \infty$. To verify this estimate, notice that as $t \to \infty$ in the upper half plane, $g(t) = -(\rho/2 + \lambda)^{-1} it + O(1)$, and for such t we have

$$\Phi(t) = \frac{1}{\pi} \cot^{-1}(-(\rho/2 + \lambda)^{-1} it + O(1)) = \frac{i(\frac{\rho}{2} + \lambda)}{\pi t} + O(t^{-2}) + k,$$

for some integer k depending on t and the branch of the arccotangent. For large t, continuity implies that k must be constant, and our knowledge of the behavior of Φ on the real axis tells us that $k = 0$. It follows that $\Phi(t) = O(t^{-1})$, so the integrals must converge. To prove boundedness, we need only show that the integrals remain bounded as $s \to 0$. To see that they do, notice that the limiting integrand has singularities, but they are only logarithmic singularities (since Φ' has poles of order 1 there), so it is still integrable.

We can now verify that $\mathbb{A}(1) = \lambda$. (This fact, which is necessary for $\mathbb{A}$ to be admissible, was stated earlier, but the proof was postponed.) To determine $\mathbb{A}(1)$, we need to integrate $\mathbb{A}'$ from 0 to 1. If $\mathcal{C}$ denotes a semi-circle of radius R centered at 0, lying in the upper half plane, and oriented clockwise, then for $R > \max(\rho_L, 1 + \rho_R)$, Cauchy's theorem implies that

$$\begin{aligned} \operatorname{Re} \int_{\mathcal{C}} \Phi(t) \, dt &= \operatorname{Re} \int_{-R}^{R} F(t) \, dt \\ &= \int_{-\rho_L}^{1+\rho_R} \mathbb{A}'(t) + J'(t) \, dt \\ &= \mathbb{A}(1) + \frac{\rho}{2}. \end{aligned}$$

Since $\Phi(t) = \frac{i(\rho/2+\lambda)}{\pi t} + O(t^{-2})$,

$$\begin{aligned}\int_{\mathcal{C}} \Phi(t)\,dt &= \int_{\mathcal{C}} \frac{i(\frac{\rho}{2}+\lambda)}{\pi t}\,dt + O(R^{-1})\\ &= (-\pi i)\frac{i(\frac{\rho}{2}+\lambda)}{\pi} + O(R^{-1})\\ &= \lambda + \frac{\rho}{2} + O(R^{-1}).\end{aligned}$$

Hence, $\mathbb{A}(1) = \lambda$.

Now Theorem 5.1 tells us that because

$$F(t) = \lim_{s\to 0+} \Phi(t+is)$$

(except where undefined), the imaginary part of F is the Hilbert transform of the real part, i.e.,

$$\widetilde{\mathbb{A}'} + \widetilde{J'} = \operatorname{Im} F.$$

To complete the proof of (5.1), we need more information about how $\operatorname{Im} F(t)$ behaves for $t \in [0,1]$. We know that $\operatorname{Im} F(t) = 0$ for $t \in (r_1, r_2)$, by the definition of Φ, so for such t,

$$(A(t) - \mathbb{A}(t))(\widetilde{\mathbb{A}'}(t) + \widetilde{J'}(t)) = 0.$$

If we can ensure that

$$(A(t) - \mathbb{A}(t))(\widetilde{\mathbb{A}'}(t) + \widetilde{J'}(t)) \geq 0 \tag{5.4}$$

for all $t \in (0,r_1) \cup (r_2,1)$, then we will have proved (5.1).

We will deal first with the sign of $A(t) - \mathbb{A}(t)$. Recall that $\mathbb{A}'$ is constant on $(0,r_1)$, and is either 0 or 1 (assuming $r_1 > 0$). Because of the Lipschitz condition $0 \leq A' \leq 1$, it follows that either $A'(t) - \mathbb{A}'(t) \geq 0$ for all $t \in (0,r_1)$, or $A'(t) - \mathbb{A}'(t) \leq 0$ for all such t, according as $\mathbb{A}'$ is 0 or 1 on that interval. Integrating and using $A(0) = \mathbb{A}(0) = 0$ implies that $A(t) - \mathbb{A}(t) \geq 0$ for $t \in (0,r_1)$ in the first case (where $\mathbb{A}'(0) = 0$), and $A(t) - \mathbb{A}(t) \leq 0$ in the second (where $\mathbb{A}'(0) = 1$). Similarly, if $\mathbb{A}'(1) = 0$ then $A(t) - \mathbb{A}(t) \leq 0$ for $t \in (r_2,1)$, and if $\mathbb{A}'(1) = 1$ then $A(t) - \mathbb{A}(t) \geq 0$ for $t \in (r_2,1)$. Therefore, to prove (5.4), we need only prove the same inequalities as here, with $A(t) - \mathbb{A}(t)$ replaced by $\operatorname{Im} F(t)$.

We have already shown that $\operatorname{sgn} \operatorname{Im} F(t) = \operatorname{sgn} f_1(r_1)$ for $t \in (0,r_1)$, and that $\operatorname{sgn} \operatorname{Im} F(t) = -\operatorname{sgn} f_1(r_2)$ for $t \in (r_2,1)$. We know as well that $\mathbb{A}'(0) = 0$ if $f_1(r_1) > 0$ and $\mathbb{A}'(0) = 1$ if $f_1(r_1) < 0$, and that similarly, $\mathbb{A}'(1) = 0$ if $f_1(r_2) > 0$ and $\mathbb{A}'(1) = 1$ if $f_1(r_2) < 0$. (Note that the only possible conditions under which r_1 or r_2 are roots of f_1 are $r_1 = 0$ and $r_2 = 1$, respectively.) These conditions, when combined with those derived in the previous paragraph, give us what we need. We conclude that for $t \in (0,r_1) \cup (r_2,1)$, we have $A(t) - \mathbb{A}(t) \geq 0$ iff $\operatorname{Im} F(t) \geq 0$. Since $\operatorname{Im} F(t) = \widetilde{\mathbb{A}'}(t) + \widetilde{J'}(t)$, and is 0 for $t \in (r_1,r_2)$, we see that for all $t \in (0,1)$ (and trivially for $t \notin (0,1)$ since then $A(t) = \mathbb{A}(t)$),

$$(A(t) - \mathbb{A}(t))(\widetilde{\mathbb{A}'}(t) + \widetilde{J'}(t)) \geq 0.$$

This inequality implies (5.1), which completes our proof.

Thus, $\mathbb{A}$ is indeed the unique admissible function such that $\mathcal{V}(\mathbb{A})$ is maximal. We leave to the reader the task of checking that applying this result to the specific functional arrived at in Proposition 3.1 leads to the explicit formula for $\mathcal{P}_{\alpha,\beta,\gamma}$ given in Theorem 1.1.

6. Conjectures and Open Questions

The theorems we have proved do not answer all the natural questions about the typical plane partition in a box, or about random lozenge tilings of hexagons.

Given a location (x, y) in the normalized α, β, γ hexagon, we can ask whether the probability of finding a vertical lozenge near (x, y) is given by $\mathcal{P}_{\alpha,\beta,\gamma}(x, y)$. Theorem 1.1 tells us that this is true if we average over all (x, y) in some macroscopic region. However, it is conceivable that there might be small-scale fluctuations in the probabilities that would even out on a large scale. We believe that that is not the case.

Conjecture 6.1. Let V be any open set in the α, β, γ hexagon containing the four points at which $\mathcal{P}_{\alpha,\beta,\gamma}$ is discontinuous. As $\sigma \to \infty$, the probability of finding a vertical lozenge at normalized location $(x, y) \notin V$ is given by $\mathcal{P}_{\alpha,\beta,\gamma}(x, y) + o(1)$, where the $o(1)$ error bound is uniform in (x, y) for $(x, y) \notin V$.

There is numerical evidence that Conjecture 6.1 is true. Also, the analogous result for random domino tilings of Aztec diamonds has been proved in [CEP], and it is not hard to prove that the local statistics for the one-dimensional case described in Section 1 do in fact converge to i.i.d. statistics, so it is plausible that Conjecture 6.1 is true. A similar result should also hold for higher-order statistics (probabilities of finding configurations of several dominos); one can deduce explicit hypothetical formulas for these probabilities from the theorems and conjectures in [CKP].

Of course, we do not need to restrict our attention to hexagons, but can look at tilings with lozenges of any regions. Hexagons do seem to involve the most elegant combinatorics and analysis, but one can prove in general that almost all tilings of large regions have normalized height functions that cluster around the solution to the problem of maximizing a certain entropy functional; see [CKP] for the details.

We also conjecture an analogue of the arctic circle theorem of Jockusch, Propp, and Shor. (See [JPS] for the original proof, or [CEP] for the proof of a stronger version on which our conjecture is based.) Define the *arctic region* of a lozenge tiling to be the set of lozenges connected to the boundary by sequences of adjacent lozenges of the same orientation (where a lozenge is said to be adjacent to another lozenge, or to the boundary, if they share an edge).

Conjecture 6.2 (Arctic Ellipse Conjecture)**.** Fix $\varepsilon > 0$. The probability that the boundary of the arctic region is more than a distance ε (in normalized coordinates) from the inscribed ellipse is exponentially small in the scaling factor σ.

There are also several questions for which we do not even have conjectural answers.

Open Question 6.3. Is there a simple way to derive the results of Section 5 without having to guess any formulas?

Such a method would be much more pleasant than our approach. A good test case would be the following open question.

Open Question 6.4. Is there a q-analogue of Theorem 1.1?

Of course, the non-trivial situation is when $q \to 1$ as $\sigma \to \infty$ (although we do not know the precise relationship between q and σ that will lead to interesting limiting behavior). There is a simple q-analogue of MacMahon's enumeration of boxed plane partitions (also due to MacMahon), and a q-analogue of Proposition 2.1 (which can be established by a proof similar to that of Proposition 2.1). We believe that the answer to Open Question 6.4 is yes, and that the same approach should work, but the guess work required is likely to be tedious. We hope that further development of these techniques will someday let one answer such questions more easily.

Open Question 6.5. Is there an analogue of Theorem 1.1 for "space partitions" (the natural generalization of plane partitions from the plane to space)?

The answer to Open Question 6.5 may well be yes, but it is extremely unlikely that similar techniques can be used to prove it. (For example, no analogue of MacMahon's formula is known, and there is no reason to believe that one exists.)

Appendix. Converting the Sum to an Integral

In Section 3, we had to convert a sum to an integral. The sum was

$$\sum_{1 \le i < j \le n_L + \ell + n_R} \log\left(\frac{B^{-1}(j/N) - B^{-1}(i/N)}{j/N - i/N}\right)\left(\frac{1}{N}\right)^2, \tag{6.1}$$

and we interpreted it as a Riemann sum for the double integral

$$\iint_{0 \le s < t \le \frac{\lambda+\rho}{1+\rho}} \log \frac{C^{-1}(t) - C^{-1}(s)}{t-s}\, ds\, dt, \tag{6.2}$$

for some function C which we have not yet specified. In Section 3, we claimed that the difference between the sum and the integral is $o(1)$, and that C can be chosen so that B' and C' nowhere differ by more than $O(1/N)$. In this appendix, we will define C and justify these claims.

The main obstacle is that $(B^{-1})'$ can be quite large (infinite, in fact, when $B' = 0$). We will now define a modification C of B designed to keep $(C^{-1})'$ from being too large. We put $C(0) = 0$ and $C(b_i/N) = i/N$ for $1 \le i \le n_L + \ell + n_R$ (so that $C(1) = (\lambda + \rho)/(1+\rho)$). Between these points, we will define C so that it is a continuous, strictly increasing, piecewise linear function on $[0,1]$ such that C' is constant on intervals $(i/N, (i+1)/N)$, C' is never smaller than $1/N^2$ or greater than 1, and $|C' - B'| = O(1/N)$. There is no canonical way to do this; one way that works is as follows. If $b_{i+1} > b_i + 1$, then set $C((b_{i+1} - 1)/N) = i/N + 1/N^2$ (and similarly if $b_1 > 1$ set $C((b_1 - 1)/N) = 1/N^2$). Then interpolate linearly to define C in between the points at which we have defined it so far. Notice that changing B to C does not change the sum (6.1), since $C^{-1}(i/N) = b_i/N = B^{-1}(i/N)$.

To begin, for $0 \le s < t \le \frac{\lambda+\rho}{1+\rho}$, we define

$$f(s,t) = \log \frac{C^{-1}(t) - C^{-1}(s)}{t-s}.$$

Because $(C^{-1})' = O(N^2)$, the mean value theorem implies that $f(s,t) = O(\log N)$.

Consider small squares of side length N^{-1}, with their sides aligned with the s- and t-axes and their upper right corners at $(s,t) = (i/N, j/N)$, for $1 \le i < j \le$

$n_L + \ell + n_R$. Each square has area N^{-2}, and f is bounded by $O(\log N)$, so we can safely remove up to $o(N^2/\log N)$ squares from the domain of integration and the corresponding terms from the sum without changing either by more than $o(1)$. We will do so in order to restrict our attention to squares on which we can bound the partial derivatives of f.

We first remove the $O(N^{3/2})$ squares containing some point (s,t) with $|s-t| \leq N^{-1/2}$.

Next, we remove all squares containing some point (s,t) satisfying $(C^{-1})'(s) \geq N^{1/3}$ or $(C^{-1})'(t) \geq N^{1/3}$. We can check as follows that there are at most $O(N^{5/3})$ such squares. Since C^{-1} is increasing and has range contained in $[0,1]$, the set of all t with $(C^{-1})'(t) \geq N^{1/3}$ has measure $O(N^{-1/3})$. Also, as (s,t) varies over each small square, $(C^{-1})'(s)$ and $(C^{-1})'(t)$ are constant. Hence, in this step we are removing at most $O(N^{-1/3}/N^{-2}) = O(N^{5/3})$ squares.

Thus, we can restrict our attention to squares containing only points (s,t) with $|s-t| \geq N^{-1/2}$, $(C^{-1})'(s) \leq N^{1/3}$, and $(C^{-1})'(t) \leq N^{1/3}$.

Now we will estimate the difference between the sum and the integral. If we can show that on each square, f varies by at most $o(1)$ (uniformly for all squares), then we will be done. To determine how much f can vary over a square of side length N^{-1}, we compute

$$\frac{\partial f}{\partial t} = \frac{(C^{-1})'(t)}{C^{-1}(t) - C^{-1}(s)} - \frac{1}{t-s}.$$

The second term has absolute value at most $N^{1/2}$ (since by assumption $|s-t| \geq N^{-1/2}$). To bound the first term, we start with the denominator. Because $C' \leq 1$ everywhere, we have $(C^{-1})' \geq 1$, so $|C^{-1}(t) - C^{-1}(s)| \geq |s-t|$, and thus

$$\left|\frac{(C^{-1})'(t)}{C^{-1}(t) - C^{-1}(s)}\right| \leq \frac{|(C^{-1})'(t)|}{|s-t|} \leq (C^{-1})'(t)N^{1/2}.$$

Finally, since $(C^{-1})'(t) \leq N^{1/3}$, we have

$$\frac{\partial f}{\partial t} = O(N^{5/6}) = o(N).$$

The same holds for $\partial f/\partial s$, of course.

Therefore, over one of the small squares of side length N^{-1}, f can vary by at most $o(1)$. Thus, the sum (6.1) differs from the integral (6.2) by at most $o(1)$.

Acknowledgements

We thank Tom Brennan, Persi Diaconis, and Noam Elkies for useful discussions, the anonymous referee for pointing out an oversight in our proof, and Valerie Samn for helping with the figures.

References

[BH] H. W. J. Blöte and H. J. Hilhorst, *Roughening transitions and the zero-temperature triangular Ising antiferromagnet*, J. Phys. A **15** (1982), L631–L637.

[CS] L. Carlitz and R. Stanley, *Branchings and partitions*, Proc. Amer. Math. Soc. **53** (1975), 246–249.

[CEP] H. Cohn, N. Elkies, and J. Propp, *Local statistics for random domino tilings of the Aztec diamond*, Duke Math. J. **85** (1996), 117–166, MR 97k:52026.

[CKP] H. Cohn, R. Kenyon, and J. Propp, *A variational principle for domino tilings*, preprint, 1997.

[DT] G. David and C. Tomei, *The problem of the calissons*, Amer. Math. Monthly **96** (1989), 429–431, MR 90c:51024.

[EKLP] N. Elkies, G. Kuperberg, M. Larsen, and J. Propp, *Alternating sign matrices and domino tilings*, J. Algebraic Combin. **1** (1992), 111–132, MR 94f:52035, and 219–234, MR 94f:52036.

[GT] I. M. Gelfand and M. L. Tsetlin, *Finite-dimensional representations of the group of unimodular matrices* (Russian), Dokl. Akad. Nauk SSSR **71** (1950), 825–828 (reprinted in English in *Izrail M. Gelfand: Collected Papers*, Springer-Verlag, Berlin, 1988, vol. 2, pp. 653–656).

[JP] W. Jockusch and J. Propp, *Antisymmetric monotone triangles and domino tilings of quartered Aztec diamonds*, to appear in the Journal of Algebraic Combinatorics.

[JPS] W. Jockusch, J. Propp, and P. Shor, *Random domino tilings and the arctic circle theorem*, preprint, 1995.

[LS] B. F. Logan and L. A. Shepp, *A variational problem for random Young tableaux*, Adv. in Math. **26** (1977), 206–222, MR 98e:05108.

[M] P. A. MacMahon, *Combinatory Analysis*, Cambridge University Press, 1915–16 (reprinted by Chelsea Publishing Company, New York, 1960).

[PW] J. Propp and D. Wilson, *Exact sampling with coupled Markov chains and applications to statistical mechanics*, Random Structures and Algorithms **9** (1996), 223–252.

[R] W. Rudin, *Real and Complex Analysis*, McGraw-Hill, New York, 1987.

[Th] W. P. Thurston, *Conway's tiling groups*, Amer. Math. Monthly **97** (1990), 757–773, MR 91k:52028.

[Ti] E. C. Titchmarsh, *Introduction to the Theory of Fourier Integrals*, Oxford University Press, London, 1937.

[VK1] A. M. Vershik and S. V. Kerov, *Asymptotics of the Plancherel measure of the symmetric group and the limiting form of Young tables*, Soviet Math. Dokl. **18** (1977), 527–531.

[VK2] A. M. Vershik and S. V. Kerov, *Asymptotics of the largest and the typical dimensions of irreducible representations of a symmetric group*, Functional Anal. Appl. **19** (1985), 21–31.

Department of Mathematics, Harvard University, Cambridge, MA 02138
cohn@math.harvard.edu http://www.math.harvard.edu/~cohn

Department of Mathematics, Indiana University, Bloomington, IN 47405
larsen@math.indiana.edu

Department of Mathematics, University of Wisconsin, Madison, WI 53706
propp@math.wisc.edu http://www.math.wisc.edu/~propp

This paper is available via http://nyjm.albany.edu:8000/j/1998/4-10.html.