Reconstructing random pictures

Bhargav Narayanan and Corrine Yap

Abstract. Given a random binary picture P_n of size n, i.e., an $n \times n$ grid filled with zeros and ones uniformly at random, when is it possible to reconstruct P_n from its k-deck, i.e., the multiset of all its $k \times k$ subgrids? We demonstrate very sharp 'two-point' concentration for the reconstruction threshold by showing that there is an integer $k_c(n) \sim (2 \log n)^{1/2}$ such that if $k > k_c$, then P_n is reconstructible from its k-deck with high probability, and if $k < k_c$, then with high probability, it is impossible to reconstruct P_n from its k-deck. The proof of this result uses a combination of interface-exploration arguments and entropic arguments.

1. Introduction

Reconstruction problems, at a very high level, ask the following general question: is it possible to uniquely reconstruct a discrete structure from the 'deck' of all its substructures of some fixed size? The study of such problems dates back to the graph reconstruction conjectures of Kelly and Ulam [7, 8, 20], and analogous questions for various other families of discrete structures have since been studied; see [1, 14, 15, 16] for examples concerning other objects such as hypergraphs, abelian groups, and subsets of Euclidean space. The line of inquiry that we pursue here concerns reconstructing typical, as opposed to arbitrary, structures in a family of discrete structures. Such questions, best phrased in the language of probabilistic combinatorics, often have substantially different answers compared to their extremal counterparts; see [4, 19], for instance.

Our aim in this paper is to investigate a two-dimensional reconstruction problem, namely that of reconstructing random pictures. Before we describe the precise question we study, let us motivate the problem at hand. Perhaps the most basic one-dimensional reconstruction problem concerns reconstructing a random binary string from the multiset of its substrings of some fixed size, a problem intimately connected to that of shotgun-sequencing DNA sequences; on account of its wide applicability, this question has been investigated in great detail, as in [2, 12] for instance. A natural analogue of the aforementioned one-dimensional problem concerns reconstructing a random binary grid (or picture, for short) from the multiset of its subgrids of some fixed size; this is the question that will be our focus here. While shotgun-reconstruction of random strings is

Date: 17 October, 2022.

2010 *Mathematics Subject Classification.* Primary 60C05; Secondary 60K35, 68R15.

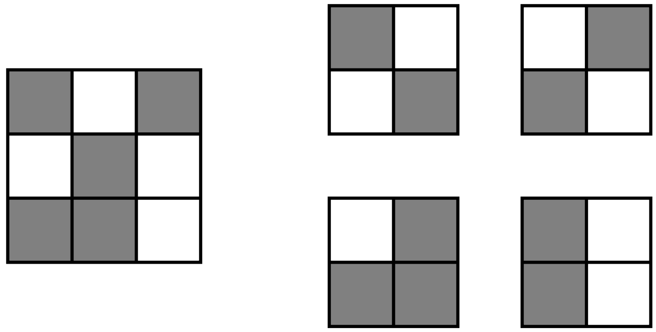


FIGURE 1. A picture of size 3 and its 2-deck.

a well-studied problem, there has been renewed interest — originating from the work of Mossel and Ross [10] — in generalisations of this problem like the one considered here; see [6, 11, 18] for some recent examples.

Writing $[n]$ for the set $\{1, 2, \dots, n\}$, a *picture of size* n is an element of $\{0,1\}^{[n]^2}$ viewed as a two-colouring of an $n \times n$ grid using the colours 0 and 1. The k*-deck* of a picture P of size n, denoted $\mathcal{D}_k(P)$, is the multiset of its $k \times k$ coloured subgrids of which there are precisely $(n-k+1)^2$; see Figure 1 for an illustration. We say that a picture P is *reconstructible* from its k-deck if $\mathcal{D}_k(P') = \mathcal{D}_k(P)$ implies that $P' = P$. Writing P_n for a random picture of size n chosen uniformly from the set of all pictures of size n, our primary concern is then the following question: when is P_n reconstructible from its k-deck with high probability? We shall give a nearly complete answer to this question. Define a function $f : \mathbb{R} \to \mathbb{Z}$ by

$$f([m - 3/4, m + 1/4)) = \{m\}$$

for all $m \in \mathbb{Z}$, and set

$$k_c(n) = f\big((2 \log_2 n)^{1/2}\big) \tag{1}$$

for all $n \in \mathbb{N}$. Writing $\mathcal{R}(n,k)$ for the event that the random picture P_n is reconstructible from its k-deck, our main result is as follows.

Theorem 1.1. *As $n \to \infty$, we have*

$$\mathbb{P}(\mathcal{R}(n,k)) \to \begin{cases} 1 & \text{if } k > k_c(n), \text{ and} \\ 0 & \text{if } k < k_c(n). \end{cases}$$

In other words, Theorem 1.1 shows that the 'reconstruction threshold' of a random picture is concentrated on at most two consecutive integers. The two results contained in the statement of Theorem 1.1 are proved by rather disparate methods: the '0-statement' follows from entropic considerations, while the '1-statement' is proved by an interface-exploration argument, in the spirit of the Peierls argument from percolation theory.

Let us briefly mention that a related, but somewhat different, two-dimensional question about reconstructing 'jigsaws' was raised by Mossel and Ross [10]. Following work of Bordenave, Feige, and Mossel [5] and Nenadov, Pfister, and Steger [13], the coarse asymptotics of the reconstruction threshold in this setting were independently established by Balister, Bollobás, and the first author [3] and by Martinsson [9]. In contrast, our main result establishes fine-grained asymptotics of the reconstruction threshold in the (somewhat different) setting considered here.

This paper is organized as follows. We give the short proof of the 0-statement in Theorem 1.1 in Section 2. The bulk of the work in this paper is in the proof of the 1-statement in Theorem 1.1 which follows in Section 3. We conclude with some discussion in Section 4.

2. Proof of the 0-statement

In this short section, we prove the 0-statement in Theorem 1.1 which, as mentioned earlier, follows from considerations of entropy.

Proof of the 0-statement in Theorem 1.1. An easy calculation shows that the definition of $k_c(n)$, see (1), ensures that

$$n^2/2^{k^2} \to \infty$$

as $n \to \infty$ for every $k < k_c(n)$. Under this condition, we show that with high probability, it impossible to reconstruct a random picture P_n of size n from its k-deck. The reason is simple: under this assumption on k, the k-deck does not contain enough entropy to allow reconstruction; for simplicity, we phrase this argument in the language of counting.

First, the number of pictures of size n is 2^{n^2}. Next, the number of such pictures that are reconstructible from their k-decks is at most the number of distinct k-decks, which is itself at most the number of solutions to the equation

$$x_1 + x_2 + \cdots + x_{2^{k^2}} = (n-k+1)^2$$

over the non-negative integers. It follows that

$$\mathbb{P}[\mathcal{R}(n,k)] \leq \binom{(n-k+1)^2 + 2^{k^2} - 1}{2^{k^2} - 1} 2^{-n^2} \leq \left(\frac{10n^2}{2^{k^2}}\right)^{2^{k^2}} 2^{-n^2}.$$

It is now easy to check that $\mathbb{P}[\mathcal{R}(n,k)] \to 0$ if $n^2/2^{k^2} \to \infty$. This proves the 0-statement in Theorem 1.1. $\square$

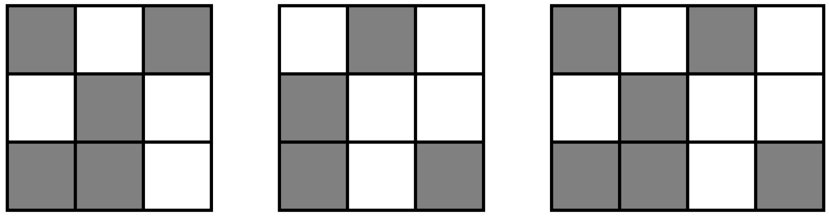


FIGURE 2. On the left is a grid S, in the center is a grid T, and on the right is the extension of S to the right by T.

3. Proof of the 1-statement

A simple calculation shows that the definition of $k_c(n)$ as given by (1) ensures that $kn^2/2^{k^2-k} \to 0$ as $n \to \infty$ for all $k > k_c(n)$. We shall show that if $k, n \to \infty$ and satisfy

$$kn^2/2^{k^2-k} \to 0,$$

then P_n is reconstructible from its k-deck with high probability. To accomplish this, we provide an algorithm for reconstruction and bound the probability that P_n is not the output when this algorithm is run on $\mathcal{D}_k(P_n)$. Our proof uses an interface argument reminiscent of the Peierls argument in percolation [17].

We first need the notion of 'extending' a picture. Given a $k \times \ell$ coloured grid S with columns $s_1, s_2, \dots, s_\ell \in \{0,1\}^k$ and a $k \times m$ coloured grid T with columns $t_1, \dots, t_m \in \{0,1\}^k$ such that $m \le \ell$, we say T *extends S to the right* if $t_{m-i-1} = s_{\ell-i}$ for all $i \in [m-2]$. We call the $k \times (\ell+1)$ subgrid with columns $s_1, s_2, \dots, s_\ell, t_m$ an *extension of S to the right*. See Figure 2 for an illustration; extensions of a grid to the left, upwards, and downwards are defined analogously.

Given an $\ell \times m$ coloured grid S, an *(a,b)-extension of S to the right* for $a \ge \ell$ is an $a \times b$ coloured grid such that the k-grid in the upper-left corner extends the top k-rows of S to the right. An (a,b)-extension to the left is defined analogously, as are (a,b)-extensions upwards or downwards for $b \ge m$. A small example is provided in Figure 3.

3.1. **Reconstruction algorithm.** We now give an algorithm for reconstructing a random picture $P = P_n$ from its k-deck. Roughly, this algorithm proceeds by placing the k-grids in the deck in place, into an expanding *droplet*, one at a time and eventually reconstructs the entire picture P. We begin with our droplet being a single (randomly chosen) k-grid and first extend it to a $3k \times k$ coloured grid. We extend this $3k \times k$ droplet horizontally one column at a time, and extend the subsequent $3k \times n$ droplet vertically one row at a time until we have an $n \times n$ grid.

When putting a deck element in place to extend our droplet, we shall often check if it can be extended further; by 'looking ahead' in this manner, we decrease the probability

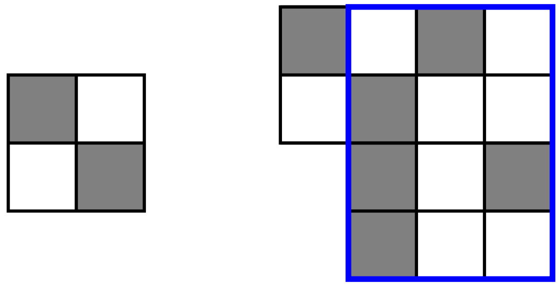


FIGURE 3. On the left is a grid S, and on the right is a $(4,3)$-extension of S to the right using the blue grid.

that the algorithm places an incorrect k-grid into the droplet. To formalize this notion, we introduce three types of extensions. A 'naive extension' simply places the next deck element that fits at a given location in our droplet. An 'internal extension' checks if a row or column of k consecutive k-grids can be placed before placing the first of those k-grids. Finally, a 'corner extension' first checks if a subgrid of size $(2k-1)\times(2k-1)$ can be placed in the corner of the current droplet before fixing one of those k-grids in place.

The majority of the algorithm consists of applying the three types of extension steps to grow the current droplet S using a randomly ordered deck $\mathcal{D}$. We now formally define the extensions alluded to above.

- **Naive extension** (to the right). Find the first element T of $\mathcal{D}$ that extends S to the right. Replace S with the extension of S by T and delete T from $\mathcal{D}$.
- **Internal extension** (to the right). Given i, consider rows i through $i+k-1$ of S and look for a $(2k-1,k)$-extension to the right using elements of $\mathcal{D}$. Once one is found, place the extension's uppermost k-grid T in the droplet and delete T from $\mathcal{D}$. See Figure 4d.
- **Double corner extension** (to the right). For the topmost k rows of S, look for a $(2k-1,2k-1)$-extension to the right using elements of $\mathcal{D}$. When one is found, place the extension's upper-left corner T in the droplet, and delete T from $\mathcal{D}$. Do the same for the bottom-most k rows of S but with the $(2k-1,2k-1)$-extension reflected vertically, and place the extension's lower-left corner in the droplet. See Figures 4b and 4c.

Extensions in other directions are defined analogously, and we refer to the illustrations in Figure 5. We now give a full description of our reconstruction algorithm. Throughout, S will denote the droplet at the current step.

(1) Choose a uniformly random ordering of $\mathcal{D}_k(P)$; let S be the first element, and let $\mathcal{D}=\mathcal{D}_k(P)\setminus S$.

(2) Naively extend S until it has $3k$ rows by first extending downward, then upward if necessary (Figure 4a).

(3) Repeat the following until an extension fails.

> *Single column extension*: Extend S one column to the right by first applying a double-corner extension, then applying $k+1$ internal extensions going from top to bottom (Figures 4b–4d).

The last internal extension adds a k-grid adjacent to the bottom corner k-grid. For the remaining k-grids extending rows $k+2$ through $3k-1$, delete these k-grids from $\mathcal{D}$ (these appear in the reconstructed grid but have not been explicitly placed by the algorithm—see Figure 4e). If these are not all found in $\mathcal{D}$, abort.

(4) When an extension fails, we assume that we are close to the boundary and have our algorithm do the following.

If an internal extension fails, we identify the previous column as the boundary of the grid and undo the column where the failure occurred, returning the deck elements from the failed column back to $\mathcal{D}$.

If a corner extension fails, we are within $k-1$ columns from the boundary. We repeat the following boundary step until it fails, at most $k-1$ times. We identify the column before the failure as the boundary.

> *Boundary step*: Extend S one column to the right by first applying naive extensions at the top and bottom corners and then applying $k+1$ internal extensions going from top to bottom (Figures 4f–4g).

(5) Repeat Steps 3 and 4 extending to the left. If the resulting droplet does not have n columns, abort.

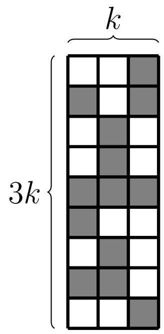
(6) Once our droplet has dimensions $3k \times n$, we repeat the following until an extension fails:

> *Single row extension*: Do the same as in Step 3 but extending S upward one row at a time by first applying a double-corner extension to the upper-left and upper-right corners (where the $(2k-1, 2k-1)$-extensions are upward; see Figure 5a), then applying $n-2k+1$ internal extensions from left to right (Figure 5b).

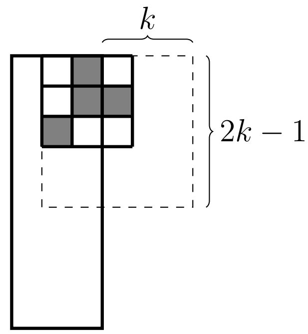
As before, we check if the k-grids that intersect both the last internal extension and the upper-right corner are elements of $\mathcal{D}$; if so, delete them from $\mathcal{D}$ and if not, abort.

(7) When an extension fails, again apply Step 4, but now for extending upward.

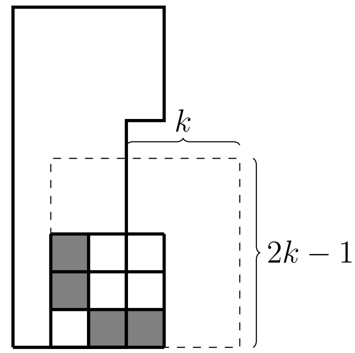
(8) Repeat Steps 6 and 7 but extending downward. If the resulting droplet does not have n rows, abort. Otherwise, output the resulting picture of size n.



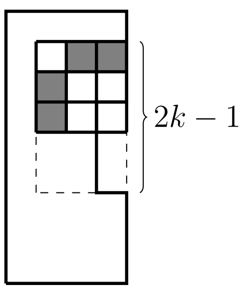
(A) First step: naive extensions.



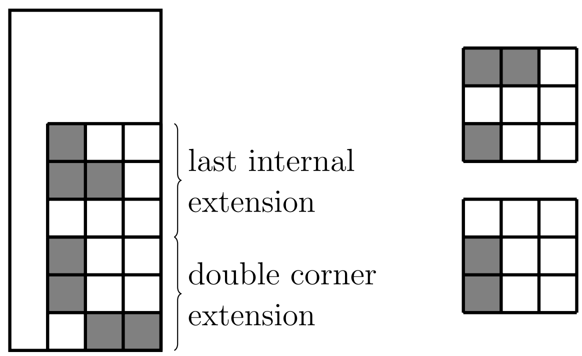
(B) First part of double corner extension.



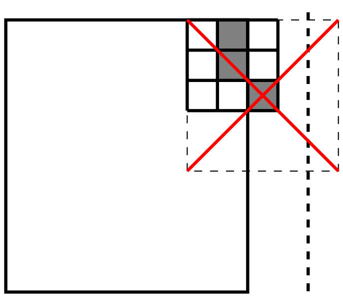
(C) Second part of double corner extension.



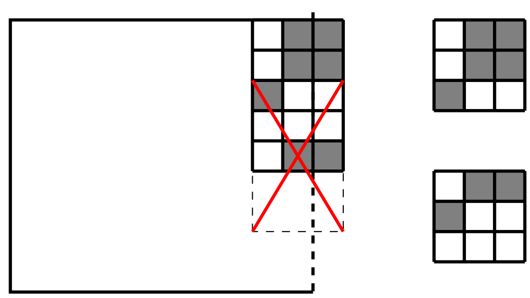
(D) Internal extension.



(E) On the left is the completed column. On the right are the subgrids intersecting the internal and corner extension that must be removed from $\mathcal{D}$.

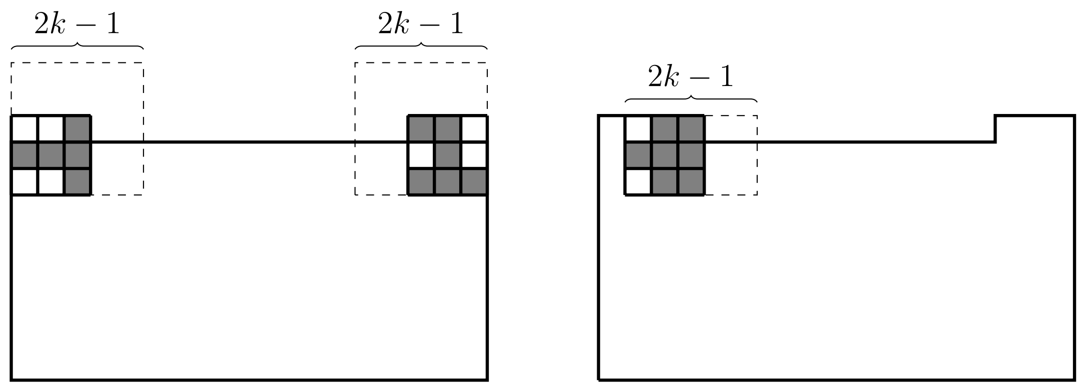


(F) If a corner extension fails, we are close to the boundary.



(G) The boundary finished with naive extensions at the corners. We undo the column that failed and return these subgrids, shown on the right, to $\mathcal{D}$.

Figure 4. Steps 1–4 of the reconstruction algorithm.



(A) Extending upwards at the corners.

(B) Extending upwards internally.

FIGURE 5. Step 6 of the algorithm.

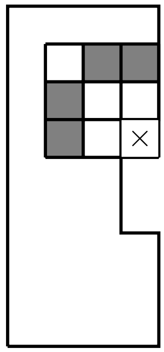
Figures 4 and 5 illustrate several steps of the algorithm. At each step, the colouring of the previously placed grids is not shown to emphasize which grid is being added in the current step.

Let (i_0, j_0) denote the location in P of the (top-right corner of the) subgrid constituting the initial droplet in Step 1; as will be clear, this subgrid, and consequently its location in P, is unique (and hence well-defined) with high probability. We will think of the first subgrid in our droplet as being located at (i_0, j_0), and will think of the positions of all subsequently placed subgrids in our droplet relative to (i_0, j_0). It is then meaningful to talk about the reconstruction algorithm making a *mistake at a location* (i, j) in P; this happens if the subgrid in our droplet at location (i, j) is not identical to the subgrid at the same location in P, and we include the reconstruction algorithm placing a subgrid at (i, j) that extends past the boundary of P in this definition. It is clear that our reconstruction algorithm successfully terminates by outputting P if it never makes a mistake. Going forward, we say a k-grid placed (potentially only temporarily) in the droplet is *bad* if it contains an entry that is incorrect with respect to the original picture P.

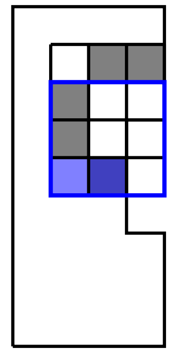
3.2. **Naive Extensions.** We first write down a simple bound for the probability that a naive extension made by the reconstruction algorithm results in its first mistake.

Lemma 3.1. *If* $n^2k2^{-k^2+k} \to 0$, *then for any fixed location in* P, *the probability that the first mistake made by the reconstruction algorithm is in a naive extension at that location is* $o(1/k)$.

Proof. Consider any instance where our reconstruction algorithm extends the current (mistake-free) droplet S to the right, say, using a naive extension. Observe that if T extends S to the right, then there are $k^2 - k$ cells in $S \cap T$. The probability for any



(A) If the coloured grid shown is the first mistakenly placed grid, the incorrect entry must be the one marked with an $\times$.



(B) Condition (II) holds if the shaded blue cells match S.

FIGURE 6

given element of the deck $\mathcal{D}$, other than the correct one (assuming one exists, which may not be the case near the boundary of P), to extend S in this manner is 2^{-k^2+k}.

Thus, the probability of a naive extension at any fixed location resulting in the first mistake is then, by a union bound over all possible choices for the incorrect extension from the deck, at most $n^2 2^{-k^2+k}$, which is $o(1/k)$ by the hypothesis that $n^2 k 2^{-k^2+k} \to 0$. $\square$

3.3. **Internal Extensions.** Next, we consider the event that the *first* mistake made by the reconstruction algorithm is in an internal extension.

Lemma 3.2. *If $n^2 k 2^{-k^2+k} \to 0$, then for any fixed location in P, the probability that the first mistake made by the reconstruction algorithm is in an internal extension at that location is $o(1/(k^2 n^2))$.*

Proof. Fix any valid droplet S and consider the event that the next internal extension made by the reconstruction algorithm to extend S results in a mistake. Suppose that T is the k-grid placed by this internal extension.

By the assumption that placing T is the first mistake made by the reconstructiong algorithm, we know that the entries in S and thus $S \cap T$ are all correct, so there is exactly one incorrect entry in T, as shown in Figure 6a. Moreover, every k-grid in the $(2k-1, k)$-extension that we check before placing T must also contain this incorrect entry.

Let E_i be the event that there exists a (bad) subgrid T_i which can be placed at the ith step of the $(2k-1, k)$-extension. Then,

$$\mathbb{P}[T \text{ is placed}] \leq \mathbb{P}[E_1] \times \mathbb{P}[E_2 \mid E_1] \times \cdots \times \mathbb{P}[E_k \mid E_{k-1}, \dots, E_1].$$

As in the proof for Lemma 3.1, we know that

$$\mathbb{P}[E_1] \leq n^2 2^{-k^2+1},$$

since the probability for any given element of the deck $\mathcal{D}$, other than the correct one, to extend S in this manner is 2^{-k^2+1}.

Claim 3.3. *For $1 < i \leq k$, we have*

$$\mathbb{P}[E_i \mid E_{i-1}, \dots, E_1] \leq 2^{-k+1}\left(1 + o\left(\frac{2}{k}\right)\right).$$

Proof. We claim that

$$\mathbb{P}[E_i \mid E_{i-1}, \dots, E_1] \leq n^2 2^{-k^2+1} + 2^{-k+1} + i4k^2(2^{-2k+2}). \tag{2}$$

Without loss of generality, suppose the direction of the internal extension is to the right.

For a given deck element D, let $D(a,b) \in \{0,1\}$ denote the entry in row a and column b of D. Suppose T_i extends the $(\ell - k + 1)$th through ℓth rows of S, and for notational ease, assume that S has k columns (since we may safely ignore any columns of S beyond the rightmost k for the purposes of understanding the events in question).

Observe that the probability that D can be placed as T_i, conditional on the events that $T_1, \dots, T_{i-1}$ were already placed, is exactly the probability that

(I) $D(a,b) = T_{i-1}(a+1,b)$ for all $1 \leq a \leq k-1$ and $1 \leq b \leq k$, and
(II) $D(k,b) = S(\ell, b+1)$ for $1 \leq b \leq k-1$.

(I) says that D must extend T_{i-1} downward, which means the top $k-1$ rows of D must match the bottom $k-1$ rows of T_{i-1}, revealing nothing about the bottom row of D. (II) says that D must also extend S to the right in the entries not already accounted for in T_{i-1}, which are the leftmost $k-1$ entries in the bottom row of D (i.e., the shaded blue cells in Figure 6b). Observe that the entries of D involved in (I) are independent of those in (II).

To prove (2), we shall consider the grids that intersect with something earlier in the $(2k-1,k)$-extension separately from those that do not. To that end, let $\mathcal{D}_j \subset \mathcal{D}$ be the multiset of deck elements that intersect with T_j in the original picture P. Let R be the k-grid that extends T_{i-1} downward in P.

If $D \notin \cup_{j<i}\mathcal{D}_j$, then the probability D can be placed as T_i is 2^{-k^2+1}, since conditioning on $E_1, \dots, E_{i-1}$ does not expose any information about the content of the cells of such an element D; we crudely bound the number of such elements by n^2, which explains the first term in (2).

Next, we consider the probability of R appearing as T_i. As illustrated in Figure 6b, conditional on $E_1, \dots, E_{i-1}$, the probability of R being placed as T_i is 2^{-k+1}; this is the probability that (II) holds. Indeed, for $D = R$, we know that (I) holds with probability 1. On the other hand, for (II), observe that conditioning on $E_1, \ldots, E_{i-1}$ does not expose any information about the event in question. (Even in the case that $R \in \mathcal{D}_j$ for some $j \le i-1$, knowing that $D(k,b) = S(x,y)$ for some $x < \ell$ does not affect the probability that $D(k,b) = S(\ell, b+1)$ since the entries $S(x,y)$ and $S(k+1,b+1)$ are independent.) From this, we get the second term of (2).

Finally, similar reasoning holds for the case that $D \in \cup_{j<i}\mathcal{D}_j \setminus \{R\}$. Here, we break up (I) into two parts, looking at the first $k-2$ rows and the $(k-1)$th row separately.

(Ia) $D(a,b) = T_{i-1}(a+1,b)$ for all $1 \le a \le k-2$ and $1 \le b \le k$, and
(Ib) $D(k-1,b) = T_{i-1}(k,b)$ for all $1 \le b \le k$.

Again, observe that the entries of D involved in (Ia) are independent of those in (Ib). We bound the probability that (Ia) occurs by 1. (We cannot do better than this—for example, we could have that D is the second extension downward of T_{i-2} in P, in which case (Ia) holds with probability 1.) For (Ib) and (II), it is easy to again see that conditioning on $E_1, \ldots, E_{i-1}$ reveals nothing about these events; as before, any equalities in P revealed by the grids $T_1, \ldots, T_{i-1}$ are independent of the equalities involving the leftmost $k-1$ entries in the bottom two rows of D needed for (Ib) and (II). Thus, the probability of D being placed as T_i in this case is at most 2^{-2k+2}, and the number of choices for such a D is at most $|\cup_{j<i}\mathcal{D}_j| \le i(4k^2) \le 4k^3$; this explains the third term of (2).

Putting these bounds together, we get

$$\begin{aligned}\mathbb{P}[E_i \,|\, E_{i-1}, \ldots, E_1] &\le 2^{-k+1}(n^2 2^{-k^2+k} + 4k^3 2^{-k+1} + 1)\\ &\le 2^{-k+1}\left(1 + o\left(\frac{2}{k}\right)\right). \qquad \square\end{aligned}$$

Using the bound in the claim above, we get

$$\begin{aligned}\mathbb{P}[T \text{ is placed}] &\le n^2 2^{-k^2} \cdot 2^{-(k-1)^2}\left(1 + o\left(\frac{2}{k}\right)\right)^{k-1}\\ &< n^2 2^{-2k^2+2k-1}(e^2)\\ &= o\left(\frac{1}{k^2 n^2}\right),\end{aligned}$$

as claimed. $\square$

3.4. **Corner Extensions.** The bulk of our argument deals with the probability of making a mistake in a corner extension. To estimate this, we first require some notation.

The *grid graph* $G(P)$ is the natural graph associated with P, where each of the $(n+1)^2$ intersection points of gridlines is a vertex and each gridline segment connecting a pair of vertices is an edge. We say an edge of the grid graph $G(P)$ is *incident* to

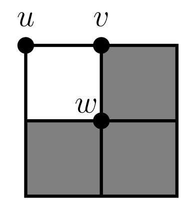


FIGURE 7. The grid graph of a picture S of size 2. The edge uv is incident to one white cell of S whereas vw is incident to one white and one black cell.

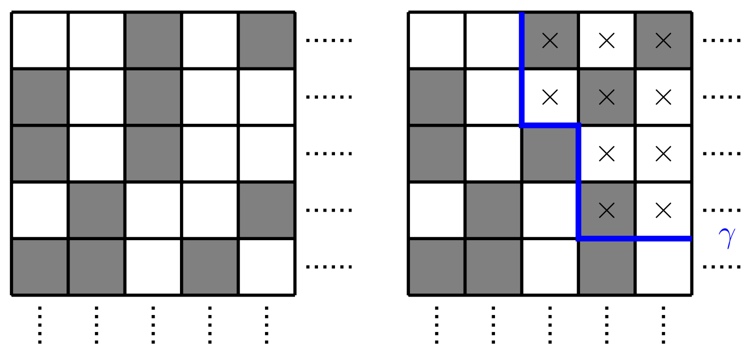


FIGURE 8. On the left is a picture P. On the right is a reconstruction Q, for $k = 2$. Each marked cell is the upper-right corner of a k-grid containing an incorrect entry. An interface path γ is highlighted in blue.

the cells of P bordering it, of which there are either one or two (see Figure 7), and we extend this to say that an edge of the grid graph is *incident* to a k-grid if it is incident to the cell in the top right corner of the k-grid.

For each bad k-grid in a reconstructed picture Q, we mark the cell in its upper-right corner. An *interface path* in the grid graph $G(Q)$ is a path separating marked cells from unmarked cells with a direction imposed, i.e., all the cells incident to the path on one of its sides are marked, and those on the other side are unmarked. See Figure 8 for an example.

We will view γ as a directed path by specifying one of the two endpoints as the initial point of γ, with the other endpoint being the terminal point. There is then a unique assignment of directions to edges of γ that is consistent with our choice of initial and terminal points. We call these directed edges right-steps, left-steps, up-steps, and down-steps. We use $\ell(\gamma)$ to denote the number of steps in γ.

We now analyse the probability that the algorithm makes its first mistake in a double-corner extension.

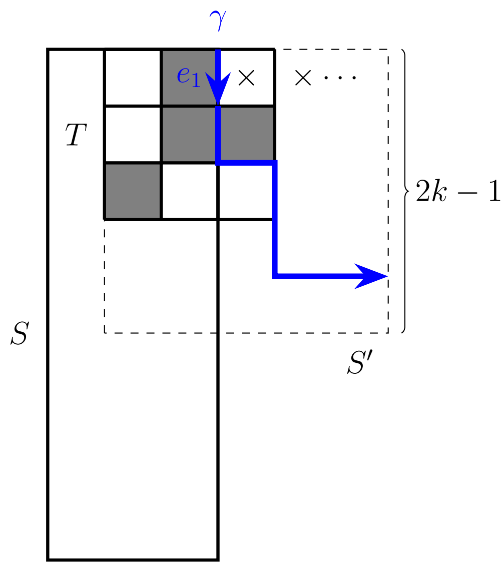


FIGURE 9. Interface path at a corner extension

Lemma 3.4. *If $n^2k2^{-k^2+k} \to 0$, then for any fixed location in P, the probability that the first mistake made by the reconstruction algorithm is in a corner extension at that location is $o(1/n)$.*

Proof. We consider the event that our reconstruction algorithm places a bad k-grid T when extending the upper-right corner of some fixed valid droplet S to the right. Recall that in order to place T, we first check that a $(2k-1, 2k-1)$-extension can be made with T as its upper-left corner. Call this $(2k-1) \times (2k-1)$ coloured grid S'. Observe that if T is bad, then there must exist an interface path γ in $G(S')$ such that an endpoint of γ is at the upper-left corner of S'. Note that every cell in the top row of S' from the top-right corner of T onwards must be marked, since the corresponding k-grids must all contain the mistake in T, and all the cells along the right edge of S are unmarked, since S is a valid mistake-free droplet by assumption, whence there is an interface path separating these. Since we are considering a corner extension to the right, γ must begin with a down-step, and furthermore, given the boundary conditions along the edges, we see that there must exist such an interface path whose terminal vertex lies on the right or bottom edge of $S' \setminus S$. We say that γ is a *rightward path* in the former case and a *downward path* in the latter. See Figure 9 for an illustration of a rightward interface path.

For any fixed γ with edges $\{e_1, \ldots, e_{\ell(\gamma)}\}$, we shall bound the probability that there exist bad k-grids that fit in place along the 'bad side' of γ. Let E_γ be the event that there is a $(2k-1, 2k-1)$-extension S' allowing us to place T with interface path γ, and let $E_i = E_{i,\gamma}$ be the event that we are able to place a bad k-grid incident to e_i in

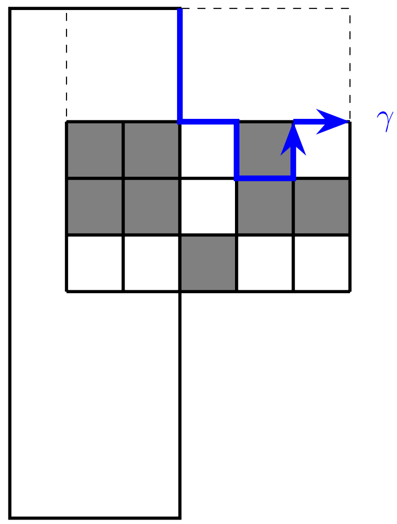


FIGURE 10. Observe that there cannot be an up-step in γ. The leftmost and rightmost 3-grids shown must have correct entries, but the middle 3-grid must contain a mistake, which is impossible.

this $(2k-1, 2k-1)$-extension, whence $E_\gamma \subset \cap E_{i,\gamma}$. We shall estimate the probability of E_γ using

$$\mathbb{P}[E_\gamma] \leq \mathbb{P}[E_1] \times \mathbb{P}[E_2|E_1] \times \cdots \times \mathbb{P}[E_{\ell(\gamma)}|E_{\ell(\gamma)-1}, \ldots, E_1],$$

noting that if T is placed by mistake, then some E_γ must hold.

For $i = 1$, observe that if T extends S to the right, then there are $k^2 - k$ cells in $S \cap T$. As in previous arguments, we have

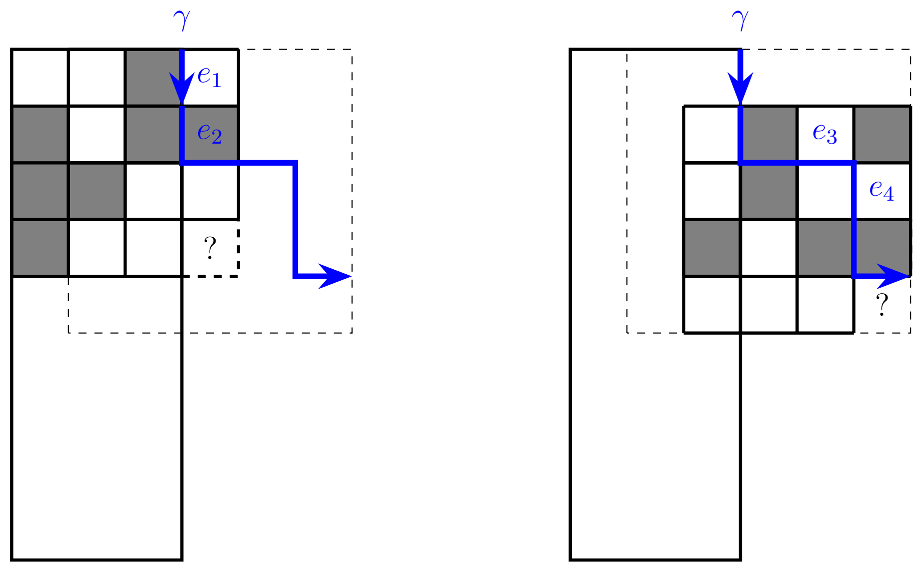
$$\mathbb{P}[E_1] \leq n^2 2^{-k^2+k}.$$

To bound the remaining conditional probabilities, we observe some properties of any valid interface path γ.

(a) There are no up-steps in γ. If e is an up-step in row i, then e is incident to a bad k-grid on its left and a good k-grid on its right. The union of this good subgrid with S contains all of the entries in the bad subgrid, contradicting the fact that the bad subgrid contains a mistake, as exemplified in Figure 10.

(b) If a left-step occurs in γ and a right-step occurs later in the same column, the steps must be separated by at least k cells. This follows by the same argument as in (a).

Say e_i is a *contributing edge* if it is: a down-step preceded by a right-step or down-step, a right-step preceded by a right-step, or a left-step preceded by a down-step or left-step. We claim the following.



(A) An example of two down-steps in a row; all but one entry of T_2 is determined by T_1 and S.

(B) An example of a down-step preceded by a right-step; all but one entry of T_4 is determined by previously placed k-grids.

Figure 11

Claim 3.5. *If $i > 1$, then for every i such that e_i is a contributing edge, we have*

$$\mathbb{P}[E_i \mid E_{i-1}, \ldots, E_1] \leq n^2 2^{-k^2+1} + i(4k^2)(2^{-k+1}) \tag{3}$$

Note the necessity of specifying the type of step before e_i. If the steps form an inner corner, i.e. if e_{i-1} is a left-step and e_i is a down-step, or if e_{i-1} is a down-step and e_i is a right-step, then the bad k-grid incident to e_{i-1} is the same as that incident to e_i, so in such cases, we trivially have $\mathbb{P}[E_i \mid E_{i-1}] = 1$.

Proof. We use a similar strategy to the proof of Lemma 3.2. For each $j < i$, let T_j be the bad k-grid that is placed earlier incident to e_j, and let $\mathcal{D}_j \subset \mathcal{D}$ be the multiset of deck elements that intersect with T_j in the original picture P.

First, if a k-grid is not in $\cup_{j<i}\mathcal{D}_j$, then the probability that it is placed incident to e_i as T_i is at most 2^{-k^2+1}. Indeed, suppose e_i is a down-step. If e_{i-1} is a down-step, say, then T_i extends T_{i-1} downward, and $T_{i-1} \cap T_i$ contains $k^2 - k$ cells. There is also a good k-grid incident to e_i which T_i extends to the right and which intersects T_i in at least $k-1$ cells outside of T_{i-1}. Thus, the probability for a given element of $\mathcal{D} \setminus \cup_{j<i}\mathcal{D}_j$ to be placed as T_i is 2^{-k^2+1}. The case where e_{i-1} is a right-step is analogous, as are the cases where e_i is a contributing right-step or left-step. See Figure 11 for an illustration.

Next, we claim that for any fixed element R of $\cup_{j<i}\mathcal{D}_j$, the probability that R can be placed incident to e_i as T_i is at most 2^{-k+1}. Indeed, recalling Conditions (I) and (II) and

using (a) and (b), we may verify geometrically that if e_i is a contributing down-step, then conditioning on $E_1, E_2, \ldots, E_{i-1}$ does not expose any information about whether (II) holds, i.e., that the $k-1$ cells in the bottom row of R agree with those of the good k-grid on the other side of e_i. Thus, the probability R that can be placed incident to e_i as T_i conditional on $E_1, E_2, \ldots, E_{i-1}$ is at most 2^{-k+1}. We may argue analogously in the cases where e_i is a contributing right-step or left-step (using the right or the left edge of R in each case) to get the same bound. The claim follows from noting that $|\cup_{j<i} \mathcal{D}_j| \le i(4k^2)$. $\square$

From (a), any valid interface path γ has at most $2k-1$ down-steps, and from (b), we see that any such γ has at most $2(k-1)\cdot(2k-2/k+1) < 4(k-1)$ left-steps and right-steps, so the length of γ satisfies $\ell(\gamma) \le 6k$. Thus, we can bound the right-hand side of (3) by

$$p_{n,k} = 2\max\{n^2 2^{-k^2+1}, 3k^3 2^{-k+3})\}.$$

By our assumption $k > k_c(n)$, we know that $n^2 2^{-k^2+k} = o(1/k)$, so it is easy to see that $p_{n,k} \le 100k^3/2^k$.

To finish, we observe that regardless of whether γ is a rightward path or downward path, γ must have at least k contributing edges.

If γ is a rightward path, then γ has at least k right-steps, each of which is either a contributing edge or which is preceded by a contributing down-step.

If γ is a downward path, then γ has exactly $2k-2$ down-steps after the first step. If at least k of these are preceded by down-steps or right-steps, then these are our k contributing edges. Otherwise, at least $k-1$ of these are preceded by left-steps. Given a down-step e_i and left-step e_{i-1}, observe that e_{i-2} is either a down-step or left-step. In either case, e_{i-1} is a contributing edge. Thus, we have at least $k-1$ contributing left-steps. In the case that equality holds, we must have at least one more contributing down-step (in fact, $k-1$ more) since the total number of down-steps is $2k-2$.

To finish, we bound the probability of placing a bad k-grid T when extending the upper-right corner of some fixed valid droplet S to the right by a union bound over all valid interface paths γ. The number of interface paths of length m is crudely at most 3^m, and the total number of valid interface paths (which we know have length at most $6k$) is thus at most 3^{6k+1}. Since any such path has at least k contributing edges, we see from the bounds above that

$$\mathbb{P}[T \text{ is placed}] \le \sum_{\gamma} \mathbb{P}[E_\gamma] \le 3^{6k+1} \cdot n^2 2^{-k^2+k} \cdot p_{n,k}^k \le \frac{3^{6k+1}(100k^3)^k}{k2^{k^2}} = o(1/n),$$

the final estimate holding with room to spare. $\square$

We are ready to apply the previous estimates to our reconstruction algorithm and prove Theorem 1.1.

Proof of the 1*-statement in Theorem* 1.1. We claim that for any fixed location (i, j) in the random picture P, the probability that our reconstruction algorithm makes a mistake when starting with the subgrid at (i, j) in P is $o(1)$; the result follows since the reconstruction algorithm selects its initial subgrid uniformly at random.

First, by an argument identical to that proving Lemma 3.1, we may see that for any fixed location (i, j) in P, the probability of the subgrid at that location occurring more than once in the deck $\mathcal{D}$ is $o(1/k)$. Assuming the starting location is well-defined, it is clear that there is a unique order in which a mistake-free execution of our reconstruction algorithm reconstructs P. In particular, such a mistake-free execution of our algorithm makes $O(k)$ naive extensions (we perform $3k$ such extensions in Step 2 and $O(k)$ in the boundary steps near the four corners of P), $O(n)$ corner extensions, and $O(n^2)$ internal extensions, and the locations of each of these extensions is determined uniquely by the starting location (i, j); let L denote this collection of extensions (along with their locations).

If our reconstruction algorithm fails to reconstruct P, then it makes its first mistake in one of the extensions listed in L. The probability of this first mistake happening at any fixed naive extension in L, of which there are $O(k)$, is $o(1/k)$ by Lemma 3.1, the probability of this happening at any fixed corner extension in L, of which there are $O(n)$, is $o(1/n)$ by Lemma 3.4, and the probability of this happening at any fixed internal extension in L, of which there are $O(n^2)$, is $o(1/n^2)$ by Lemma 3.2. Thus, by a union bound, the probability of our reconstruction algorithm ever making a mistake is $o(1)$ for any fixed starting location (i, j), proving the result. $\square$

4. Conclusion

Determining the behaviour of the reconstruction problem for pictures of size n at the critical threshold $k_c(n)$ is an interesting problem that we have not been able to resolve, though we suspect that the answer is governed purely by entropic considerations.

It is also natural to consider the reconstruction problem for higher-dimensional pictures; here, it seems likely that a similar argument to ours will give reasonable bounds, though to get sharp bounds, it seems necessary to find the appropriate higher-dimensional generalisation of the interface paths in our arguments.

Acknowledgments

This work was supported in part by NSF grants CCF-1814409 and DMS-1800521.

References

1. N. Alon, Y. Caro, I. Krasikov, and Y. Roditty, *Combinatorial reconstruction problems*, J. Combin. Theory Ser. B **47** (1989), 153–161. 1

2. R. Arratia, D. Martin, G. Reinert, and M. S. Waterman, *Poisson process approximation for sequence repeats, and sequencing by hybridization*, J. Comp. Bio **3** (1996), 425–463. 1
3. P. Balister, B. Bollobás, and B. Narayanan, *Reconstructing random jigsaws*, Multiplex and Multilevel Networks, Oxford University Press, 2018, pp. 31–50. 3
4. B. Bollobás, *Almost every graph has reconstruction number three*, J. Graph Theory **14** (1990), 1–4. 1
5. C. Bordenave, U. Feige, and E. Mossel, *Shotgun assembly of random jigsaw puzzles*, Random Structures Algorithms **56** (2020), 998–1015. 3
6. J. Gaudio and E. Mossel, *Shotgun assembly of Erdős-Rényi random graphs*, Electronic Communications in Probability **27** (2022), 1–14. 2
7. F. Harary, *On the reconstruction of a graph from a collection of subgraphs*, Theory of Graphs and its Applications (Proc. Sympos. Smolenice, 1963), Publ. House Czechoslovak Acad. Sci., Prague, 1964, pp. 47–52. 1
8. P. J. Kelly, *A congruence theorem for trees*, Pacific J. Math. **7** (1957), 961–968. 1
9. A. Martinsson, *A linear threshold for uniqueness of solutions to random jigsaw puzzles*, Combin. Probab. Comput. **28** (2019), 287–302. 3
10. E. Mossel and N. Ross, *Shotgun assembly of labeled graphs*, IEEE Trans. Network Sci. Eng. **6** (2019), 145–157. 2, 3
11. E. Mossel and N. Sun, *Shotgun assembly of random regular graphs*, Preprint, arXiv:1512.08473. 2
12. A. S. Motahari, G. Bresler, and D. N. C. Tse, *Information theory of DNA shotgun sequencing*, IEEE Transactions on Information Theory **59** (2013), 6273–6289. 1
13. R. Nenadov, P. Pfister, and A. Steger, *Unique reconstruction threshold for random jigsaw puzzles*, Chic. J. Theoret. Comput. Sci. (2017), Art. 2, 16. 3
14. L. Pebody, *The reconstructibility of finite abelian groups*, Combin. Probab. Comput. **13** (2004), 867–892. 1
15. ———, *Reconstructing odd necklaces*, Combin. Probab. Comput. **16** (2007), 503–514. 1
16. L. Pebody, A. J. Radcliffe, and A. D. Scott, *Finite subsets of the plane are* 18-*reconstructible*, SIAM J. Discrete Math. **16** (2003), 262–275. 1
17. R. Peierls, *On Ising's model of ferromagnetism*, Math. Proc. Cambridge Philos. Soc. **32** (1936), 477–481. 4
18. M. Przykucki, A. Roberts, and A. Scott, *Shotgun reconstruction in the hypercube*, Random Structures Algorithms **60** (2022), 117–150. 2
19. A. J. Radcliffe and A. D. Scott, *Reconstructing subsets of Z_n*, J. Combin. Theory Ser. A **83** (1998), 169–187. 1
20. S. M. Ulam, *A collection of mathematical problems*, Interscience Tracts in Pure and Applied Mathematics, Interscience Publishers, New York-London, 1960. 1

Department of Mathematics, Rutgers University, Piscataway, NJ 08854, USA
Email address: narayanan@math.rutgers.edu

Department of Mathematics, Rutgers University, Piscataway, NJ 08854, USA
Email address: corrine.yap@rutgers.edu