



RESEARCH ARTICLE

Topology of random 2-dimensional cubical complexes

Matthew Kahle¹, Elliot Paquette² and Érika Roldán^{3,4}

¹The Ohio State University, 100 Math Tower, 231 W. 18th Ave., Columbus, OH 43210, USA; E-mail: kahle.70@osu.edu.

²McGill University, 805 Rue Sherbrooke Ouest, Montréal, Québec H3A 0B9, Canada; E-mail: elliot.paquette@mcgill.ca.

³Zentrum Mathematik, Technische Universität München, Boltzmannstraße 3, 85748 Garching bei München.

⁴EPFL SV BMI UPHESS, Station 8 CH-1015 Lausanne, Switzerland; E-mail: erika.roldan@ma.tum.de.

Received: 17 September 2020; **Revised:** 18 August 2021; **Accepted:** 14 September 2021

Keywords: stochastic topology, cubical complexes, random groups

2020 Mathematics Subject Classification: *Primary* – 05C80; *Secondary* – 62R99, 68Q87

Abstract

We study a natural model of a random 2-dimensional cubical complex which is a subcomplex of an n -dimensional cube, and where every possible square 2-face is included independently with probability p . Our main result exhibits a sharp threshold $p = 1/2$ for homology vanishing as $n \rightarrow \infty$. This is a 2-dimensional analogue of the Burtin and Erdős–Spencer theorems characterising the connectivity threshold for random graphs on the 1-skeleton of the n -dimensional cube.

Our main result can also be seen as a cubical counterpart to the Linial–Meshulam theorem for random 2-dimensional simplicial complexes. However, the models exhibit strikingly different behaviours. We show that if $p > 1 - \sqrt{1/2} \approx 0.2929$, then with high probability the fundamental group is a free group with one generator for every maximal 1-dimensional face. As a corollary, homology vanishing and simple connectivity have the same threshold, even in the strong ‘hitting time’ sense. This is in contrast with the simplicial case, where the thresholds are far apart. The proof depends on an iterative algorithm for contracting cycles – we show that with high probability, the algorithm rapidly and dramatically simplifies the fundamental group, converging after only a few steps.

Contents

1	Introduction	2
1.1	Discussion	5
1.2	Overview and organisation	6
2	Preliminaries	6
2.1	The edge-path group	6
2.2	Star notation and the parallel relation	7
2.3	The sizes of the components of $Q(n, p)$	8
2.4	The threshold for maximal 1-faces	10
3	Parallel homotopy algorithm	11
3.1	Stage 1: Explosive growth	11
3.2	Stage 2: Only local defects remain	13
3.3	Stage 3: The final squeeze	15
4	Structure theorem for general p	17
5	Below the threshold for maximal edges	21

1. Introduction

Denote the n -dimensional cube by $Q^n = [0, 1]^n$ and the set of vertices of the n -dimensional cube by Q_0^n . In other words, $Q_0^n = \{0, 1\}^n$, which is the set of all n -tuples with binary entries. More generally, let Q_k^n denote the k -dimensional skeleton of Q^n . For example, Q_1^n is the graph with vertex set Q_0^n and an edge (i.e., a 1-dimensional face) between two vertices if and only if they differ by exactly one coordinate. Define the random 2-dimensional cubical complex $Q_2(n, p)$ as having 1-skeleton equal to Q_1^n , and where each 2-dimensional face of Q^n is included with probability p , independently.

The random complex $Q_2(n, p)$ is a cubical analogue of the random simplicial complex $Y_2(n, p)$ introduced by Linial and Meshulam in [22], whose theory is well developed. The random complex $Y_2(n, p)$ is defined by taking the complete 1-skeleton of the n -dimensional simplex Δ^n and including each 2-face independently and with probability p . In this way, $Q_2(n, p)$ is constructed in exactly the same way as $Y_2(n, p)$, except that the underlying polytope Δ^n is replaced by Q^n .

The random complex $Q_2(n, p)$ is also the 2-dimensional analogue of the random cubical graph; see [21] for a 1992 survey. To make the connection precise and review some of the history, let $Q(n, p)$ denote the random subgraph of the n -cube defined by including all vertices of Q^n , and including each edge in Q_1^n independently with probability p .

Burtin [6] and later Erdoős and Spencer [14] studied the threshold for connectivity in the random cubical graph:

Theorem 1.1 (Burtin [6], Erdoős and Spencer [14]). *For $Q \sim Q(n, p)$ and for any fixed $p \in [0, 1]$,*

$$\lim_{n \rightarrow \infty} \mathbb{P}[Q \text{ is connected}] = \begin{cases} 0 & \text{if } p < 1/2, \\ e^{-1} & \text{if } p = 1/2, \\ 1 & \text{if } p > 1/2. \end{cases} \tag{1}$$

Let $f_n(p) = \mathbb{P}[Q \sim Q(n, p) \text{ is connected}]$. It is easy to see that if $p < 1/2$, then $\lim_{n \rightarrow \infty} f_n(p) = 0$. Burtin showed that this is sharp; if $p > 1/2$, then $\lim_{n \rightarrow \infty} f_n(p) = 1$.

Erdoős and Spencer refined this argument to show what happens at $p = 1/2$ exactly. They also showed that near $p = 1/2$, the only connected components of $Q(n, p)$ are isolated vertices and a giant component. Then it is straightforward to show that the number of isolated vertices has a limiting Poisson distribution with mean 1, and as a consequence

$$\mathbb{P}[\beta_0(Q(n, 1/2)) = k + 1] \rightarrow e^{-1}/k!$$

for every integer $k \geq 0$, where β_0 denotes the number of connected components.

This picture strongly parallels what is seen in Erdoős–Rényi graphs $G \sim G(n, p)$. Let $p = (\log n + c)/n$ with $c \in \mathbb{R}$ fixed. Then

$$\lim_{n \rightarrow \infty} \mathbb{P}[G \text{ is connected}] = e^{-e^{-c}}$$

(see [13] or [5]). Letting $c \rightarrow \pm\infty$, the probability tends to 0 or 1.

The proofs share a strong similarity, in that the method is to enumerate potential cut sets and show that they are rare by making a first moment estimate of the number of cut sets.

It is therefore perhaps reasonable to speculate that the topological phenomenology of the higher-dimensional process $Q_2(n, p)$ mirrors that of $Y_2(n, p)$ after appropriately adjusting how p is chosen as a function of n . We shall show, however, that there are major differences between $Q_2(n, p)$ and $Y_2(n, p)$.

Before discussing our results and illustrating these differences, we introduce some standard terminology. Our focus is on typical behaviour of random objects for large values of n . So we will say that a sequence of statements \mathcal{P}_n holds *with high probability* (abbreviated ‘w.h.p.’) if

$$\lim_{n \rightarrow \infty} \mathbb{P}[\mathcal{P}_n] = 1.$$

We will make use of the Landau notations $O, o, \omega, \Omega, \Theta$ in the asymptotic sense, so that $f = O(g)$ means f/g is eventually bounded above as $n \rightarrow \infty$ and $f = o(g)$ means $f/g \rightarrow 0$ as $n \rightarrow \infty$. Also, $f = \omega(g)$ means $g = o(f)$ and $f = \Omega(g)$ means $g = O(f)$. Finally, we will use $f = \Theta(g)$ to mean $f = O(g)$ and $f = \Omega(g)$. We occasionally display parameters like $O_{a,b,c}(\cdot)$, emphasising that the implied constants depend on a, b, c .

We will also make use of the notion of thresholds. A function $f = f(n)$ is said to be a *threshold* for a property P of a sequence of random objects $G = G_{n,p}$ if $p = \omega(f)$ implies $G \in P$ w.h.p. and $p = o(f)$ implies $G \notin P$ w.h.p. Such a threshold is defined only up to n -independent scalar multiples. If there is a function $g = o(f)$ such that $p \geq f + g$ implies $G \in P$ w.h.p. and $p \leq f - g$ implies $G \notin P$ w.h.p., the threshold is said to be *sharp*.

In this paper we study the *fundamental group* $\pi_1(Q)$ for $Q \sim Q_2(n, p)$. The fundamental group can be given a purely combinatorial representation for a space such as $Q_2(n, p)$, which we discuss in Section 2.1. Our first result establishes the threshold for $\pi_1(Q) = 0$ – that is, for $Q \sim Q_2(n, p)$ to be simply connected. Here we formulate the theorem for p fixed independently of n .

The 2-dimensional analogues of isolated vertices are maximal 1-faces. It is not hard to see that the threshold for maximal 1-faces is $p = 1/2$, and that for any fixed $p \leq 1/2$, there are maximal 1-dimensional faces w.h.p. If there are any maximal 1-faces, then any loop passing through such an edge is not contractible, so $\pi_1(Q) \neq 0$. Most of our work is to prove the following:

Theorem 1.2. *If $Q \sim Q_2(n, p)$ and $p > 1/2$, then w.h.p., $\pi_1(Q) = 0$.*

This theorem marks a substantial difference between the behaviours of the random cubical complex $Q \sim Q_2(n, p)$ and the random simplicial complex $Y \sim Y_2(n, p)$. The threshold for $\pi_1(Y) = 0$ is roughly $p = n^{-1/2}$ [4]. The threshold is sharpened by Luria and Peled in [25]. There is a sharp threshold for $H_1(Y) = 0$ vanishing at $p = 2 \log n/n$. This is due to Linial and Meshulam [22], for the statement with $\mathbb{Z}/2\mathbb{Z}$ coefficients. This was extended by Meshulam and Wallach [26] to all finite coefficient rings, and finally to \mathbb{Z} coefficients by Łuczak and Peled [24]; see also [28] for the extension of the \mathbb{Z} -coefficients result to higher dimensions. In summary, there is a big gap between the thresholds for $\pi_1(Y) = 0$ and $H_1(Y) = 0$, but Theorem 1.2 implies that the thresholds for $\pi_1(Q) = 0$ and $H_1(Q) = 0$ (with any coefficients) coincide.

For $p > 1 - (1/2)^{1/2}$, we characterise the structure of the fundamental group $\pi_1(Q)$ for $Q \sim Q_2(n, p)$. For the following statement, recall that in a simplicial complex, a *maximal 1-face* is a 1-face which is not contained in any 2-face.

Theorem 1.3. *For $p > 1 - \left(\frac{1}{2}\right)^{1/2}$, with high probability, for $Q \sim Q_2(n, p)$,*

$$\pi_1(Q) \cong F_N,$$

where F_N denotes a free group on N generators and N denotes the number of maximal 1-dimensional faces in Q .

Hence, just below the $p = 1/2$ threshold, only maximal 1-faces contribute to the fundamental group. Again, this is in contrast with the corresponding story for $\pi_1(Y)$. The fundamental group of the random 2-dimensional simplicial complex $\pi_1(Y)$ is known to have property (T) with high probability once $p \geq (2 + \epsilon) \log n/n$ [19], and this precludes it being a nontrivial free group or even having any nontrivial free quotients.

Theorem 1.3 has the following corollary, a limiting Poisson distribution for the Betti number β_1 :

Corollary 1.4. *If $c \in \mathbb{R}$ is fixed and*

$$p = \frac{1}{2} \left(1 + \frac{\log n + c}{n} \right),$$

then

$$\mathbb{P}[\beta_1(Q_2(n, 1/2)) = k] \rightarrow \frac{e^{ck} e^{-e^{-c}}}{k!}.$$

Here β_1 represents the dimension of homology $H_1(Q, \mathbb{R})$. It is also the number of generators of the free group $\pi_1(Q)$. This all follows from Theorem 1.3, together with showing that the number of maximal 1-faces is Poisson distributed with mean e^{-c} .

It is also possible to formulate a *process version* of this statement. Here one couples all $Q_2(n, p)$ together for all $p \in [0, 1]$ in a monotone fashion, so the 2-faces of $Q_2(n, p_1)$ are a subset of $Q_2(n, p_2)$ whenever $p_1 \leq p_2$. Let $(Q_p : p \in [0, 1])$ have this distribution, and let N_p be the number of maximal 1-faces in Q_p . Then we can formulate the *stopping times* T_{sc} and T_{2d} as

$$T_{sc} = \inf \{ p : \pi_1(Q_p) = 0 \}, \quad T_{2d} = \inf \{ p : N_p = 0 \}.$$

We have the following:

Corollary 1.5. $T_{sc} = T_{2d}$ w.h.p.

Theorem 1.3 characterises the structure of $\pi_1(Q)$ for $Q \sim Q_2(n, p)$ and

$$1 - \left(\frac{1}{2} \right)^{1/2} < p < \frac{1}{2}.$$

For smaller p , we are not able to completely describe the structure of the fundamental group, but we give a partial characterisation in terms of its free factorisation.

For a finitely presented group H , we say that H is *indecomposable* if whenever H is written as a free product of two groups $H \cong H_1 * H_2$, either H_1 or H_2 is a trivial group. It is well known that a finitely presented group G may be written as a free product

$$G = G_1 * G_2 * \dots * G_\ell,$$

where every G_i is indecomposable. Moreover, this free-product factorisation is unique, up to isomorphism types of the factors and reordering.

Definition 1.6. For a cubical subcomplex T of Q^n , let $e(T)$ denote the number of edges in T . Let \mathcal{T}_p be the set of pure 2-dimensional strongly connected cubical complexes T that are subcomplexes of (the 2-dimensional skeleton of) Q^n for some n and such that $1 - (1/2)^{1/e(T)} > p$.

While we do not characterise all the free factors, we are able to characterise some of them. Essentially, we show that everything that can happen will happen, somewhere, with high probability.

Theorem 1.7. *For $p \in (0, 1)$ fixed and $Q \sim Q_2(n, p)$, let the free-product factorisation of $G = \pi_1(Q)$ be given by*

$$G \cong G_1 * G_2 * \dots * G_\ell.$$

With high probability, for every $T \in \mathcal{T}_p$ with nontrivial fundamental group, $\pi_1(T)$ appears as a free factor – that is, we have $\pi_1(T) \cong G_j$ for some $1 \leq j \leq \ell$.

Indeed, every finitely presented group G can appear as a free factor (see Section 5 for details). One can also ask the extremal topological combinatorial question: For given homotopy classes of 2-dimensional complexes T , what is the smallest $e(T)$ attainable? Furthermore, one can ask what is the threshold for specific groups to appear as a free factor in G . In Section 5, we give some examples of specific complexes which we believe to be minimal – the torus, the projective plane and the Klein bottle.

Using the case of the projective plane, when $0 < p < 1 - \left(\frac{1}{2}\right)^{1/40} \approx 0.017179$ and $Q \sim Q_2(n, p)$, we show that $\pi_1(Q)$ has a $\mathbb{Z}/2\mathbb{Z}$ free factor with high probability (see Corollary 5.4). This shows in particular that $H_1(Q; \mathbb{Z})$ has torsion elements for small enough fixed $p > 0$. The question of whether torsion ever appears in $H_1(Y; \mathbb{Z})$ is an open one, although it appears that there is almost always a short burst of enormous torsion [20].

Conjecture 1.8. For $p > 1 - \left(\frac{1}{2}\right)^{1/40}$ and $Q \sim Q_2(n, p)$, $\pi_1(Q)$ is torsion free with high probability.

We further believe it is possible that for p above this threshold, $\pi_1(Q)$ is free with high probability. For the case of $\pi_1(Y)$ with $Y \sim Y_2(n, p)$, the sharp threshold is found by Newman in [27], improving on previous work of [7].

1.1. Discussion

We have not addressed many results about the fundamental group of $\pi_1(Y)$ for $Y \sim Y_2(n, p)$, which may have interesting analogues in $Q_2(n, p)$ that could further elucidate what appear to be deep differences between simplicial and random cubical complexes. As the body of literature on $Y_2(n, p)$, is substantial, we discuss possible directions of interest for questions about $Q_2(n, p)$.

Many interesting topological phases of $\pi_1(Q)$ are likely to exist when $p \rightarrow 0$ at an appropriate rate as $n \rightarrow \infty$. For $Y_2(n, p)$, a particularly rich regime of p is when the mean degree of an edge np tends to a constant. We would expect this regime to be similarly rich for $Q_2(n, p)$ and to name a few transitions that should appear in this regime: the collapsibility threshold [2], the threshold for a giant shadow [23] and the threshold for $\pi_1(Q)$ to have an irreducible factor in its free-product factorisation with growing rank.

A natural direction is to consider higher-dimensional complexes $Q_d(n, p)$ built in an analogous way to $Y_d(n, p)$. For $Q_d(n, p)$, it may be possible to analyse the higher homotopy groups in a similar fashion to what is done here.

In a different direction, we mention that all of the results we present are about the $n \rightarrow \infty$ limit, but also have some content for some large n . These could provide useful results for understanding 2-dimensional percolation on a sufficiently high-dimensional lattice \mathbb{Z}^n . There are some recent related results for such higher-dimensional cubical percolation [12, 18, 17].

1.1.1. Multiparameter generalisations

In the literature on random cubical graphs, there is a 2-parameter model $Q(n; p_0, p_1)$ (see [21] for a survey of some results). First we take a random induced subgraph of the n -cube, where every vertex has probability p_0 independently. Then we include each of the remaining edges with probability p_1 independently. Bond percolation on the hypercube is the random cubical graph where $p_0 = 1$ and p_1 varies, and site percolation is where $p_1 = 1$ and p_0 varies.

It seems natural to form a higher-dimensional generalisation of this model $Q_2(n; p_0, p_1, p_2)$. Indeed, Costa and Farber [8, 9, 10] have made a detailed study of the analogous model $Y_2(n; p_0, p_1, p_2)$, including many interesting results on the fundamental group. (See also [15], wherein new questions about the fundamental group are discussed for this multiparameter model.) Our discussion has been about the special case where $p_0 = p_1 = 1$ and p_2 varies.

For $Q_2(n; p_0, p_1, p_2)$, it is natural to ask if there is a critical surface for homology H_1 vanishing in the unit cube. The case of setting $p_1 = p_2 = 1$ and letting p_0 vary looks particularly interesting, analogous to the site-percolation model. Higher homology is no longer monotone, as in, for example,

a random clique complex or a Vietoris–Rips complex. Are there separate thresholds for H_0 vanishing, H_1 appearing, H_1 vanishing and H_2 appearing?

1.2. Overview and organisation

We begin with Section 2, where we define some key notions for working with $Q_2(n, p)$, and make some elementary estimates about it. In Section 2.1 we give a combinatorial definition of π_1 . In Section 2.2 we introduce notation to work with subcomplexes of Q^n and we introduce the notions of the random cubical complex $Q_2(n, p)$ and the random graph $Q(n, p^4)$. In Section 2.3 we summarise some estimates from [14] that we need about the random graph $Q(n, p)$. In Section 2.4 we make estimates for the existence of maximal 1-faces and use these to deduce Theorem 1.2 from Theorem 1.3.

In Section 3 we introduce an algorithm for identifying contractible 4-cycles in π_1 . This algorithm reduces the analysis of π_1 to determining the topology of small subcomplexes. We also finish the proof of Theorem 1.2 and give a proof of Theorem 1.3.

In Section 4 we show a general structure theorem that describes the free-product factorisation of $\pi_1(Q_2(n, p))$, and we then prove Theorem 1.7. In Section 5 we construct specific complexes which show that certain interesting free factors appear.

2. Preliminaries

2.1. The edge-path group

For subcomplexes Q of Q^n , the fundamental group $\pi_1(Q)$ has a nice combinatorial definition as the *edge-path group*, which we now define.

Say that two edges (1-faces) of Q^n are adjacent if they intersect at a vertex. An *edge path* in Q^n is defined to be a sequence of edges in which every consecutive pair are adjacent.

In Q^n , any 2-face has a 4-cycle as its boundary. Conversely, every 4-cycle in Q^n is the boundary of a 2-face. Hence, any two adjacent edges are contained in a unique 2-face and therefore in a unique 4-cycle in Q^n .

Two edge paths in Q are said to be *edge-equivalent* if one can be obtained from the other by successively doing one of the following moves:

- 1) Replacing two consecutive adjacent edges by the two opposite edges of the 4-cycle x that contains them, *if* the 2-face that bounds x is in Q
- 2) Replacing one edge contained in a 4-cycle x with the other three consecutive edges in x , *if* the 2-face that bounds x is in Q
- 3) Replacing three consecutive edges in a 4-cycle x with the complementary edge in x , *if* the 2-face that bounds x is in Q
- 4) Removing an edge that appears twice consecutively or adding an edge that appears twice consecutively

Define $\bar{0}$ to be the vector with only zero entries in Q_0^n . An edge loop at $\bar{0}$ is an edge path starting and ending at $\bar{0}$. The edge-path group $\pi_1(Q)$ of a complex $Q \subset Q^n$ (for any $n \geq 1$) is defined as the set of edge-equivalence classes of edge loops at $\bar{0}$ (with product and inverse defined by concatenation and reversal of edge loops).

For random cubical complexes $Q \sim Q_2(n, p)$, we explore first the extremal cases. If $p = 0$, then any $Q \sim Q_2(n, p)$ is equal to Q_1^n – that is, with a probability of 1, the complex Q has not a single 2-face included. Observing that in any graph G the number of independent generators in $\pi_1(G)$ is equal to $E(G) - V(G) + 1$, in the case of an element $Q \sim Q_2(n, p)$ we get

$$E(C) = 2^{n-1}n, \quad V(G) = 2^n, \quad (2)$$

which implies that the number of independent generators in $\pi_1(Q)$ will be at most $2^{n-1}(n-2)+1$. Thus, when $p=0$ we have that $\pi_1(Q)$ is a free group with $2^{n-1}(n-2)+1$ independent generators, and this is the maximum number of independent generators that the edge-path group of a random 2-cubical complex can attain. This number of independent generators is less than the total number of 4-cycles in Q^n , which is $2^{n-3}n(n-1)$. If $p=1$, then any $Q \sim Q_2(n, p)$ will have all the 2-faces included, which implies that $\pi_1(Q) = 0$ with a probability of 1.

2.2. Star notation and the parallel relation

In Q^n , the four vertices belonging to a 4-cycle have $n-2$ equal entries and two coordinate entries that are not equal in all of them. Denote these nonequal coordinate entries as i and j ; then we can uniquely represent a 4-cycle using an n -tuple with $n-2$ fixed binary values and two $*$ s. One $*$ will be located on coordinate i and the other will be located on coordinate j . As an example, the 4-cycles of Q^3 are $\{(0, *, *), (*, 1, *), (*, *, 1), (*, 0, *)\}$, with, for instance, $(*, *, 0) = \{(1, 1, 0), (1, 0, 0), (0, 1, 0), (0, 0, 0)\}$.

For dice, physical realisations of the cube Q^3 , we have a physical intuitive notion of parallel faces; there are three pairs of parallel faces, and if the die is fair, each pair should add up to 7. Using the $*$ notation of 4-cycles just introduced, we extend this notion of parallel faces to Q^n :

Definition 2.1. Two 4-cycles in Q^n are *parallel* if they have the two $*$ s in the same entries, and if their Hamming distance is 1.

Thus, in Q^3 (the cube), there are three pairs of parallel 4-cycles: $(0, *, *)$ and $(1, *, *)$, the 4-cycles $(0, *, *)$ and $(1, *, *)$, and the 4-cycles $(0, *, *)$ and $(1, *, *)$. With this parallel notion, we are able to define a binary relation in the set of 4-cycles of a random 2-cubical complex.

We represent a 3-dimensional cube in Q^n with a vector with n entries, three of which have a fixed $*$ and the rest of which are binary numbers. If we have two parallel faces x and y that have $*$ in the i and j coordinates and differ (only) in the binary value of the k coordinate, then the cube that contains them is represented by a vector with the three fixed $*$ s in entries i, j, k and with the rest of the $n-3$ entries equal to the entries of x (which are also equal to the entries of y).

Observe that the other four 4-cycles of the cube will have all the entries that are not i, j , or k equal to the entries of x , two $*$ s in positions either $\{i, k\}$ or $\{j, k\}$ and a binary number in the remaining coordinate. Also, given two parallel 2-faces in Q^n , there is a unique 3-dimensional cube in Q^n that contains them.

Definition 2.2. Given a subcomplex Q of Q^n , two parallel 4-cycles $x, y \in Q$ are related if the (unique) 3-dimensional *cube* that contains them has a 2-face attached to each one of the other four 4-cycles in the cube. We represent this by $x \parallel y$.

Lemma 2.3. *If two 4-cycles x and y are related ($x \parallel y$), then they are edge-equivalent.*

Definition 2.4 (Graph of parallel related 4-cycles). Given a subcomplex Q of Q^n , we define its graph of parallel cycles, which we represent by $G[Q]$, as the graph with set of vertices V^n whose elements are all the the 4-cycles in Q^n (there are $2^{n-3}n(n-1)$ 4-cycles), and an edge between two of them if they are related by \parallel .

Observe that \parallel is reflexive but not transitive. This implies, for instance, that $G[Q_2^n]$ (remember that Q_2^n is the 2-skeleton of Q^n) is not the complete graph. We can completely characterise $G[Q_2^n]$.

Lemma 2.5. *The graph $G[Q_2^n]$ has $\binom{n}{2}$ components, and each one of these components is isomorphic to a Q_1^{n-2} graph.*

Proof. Fix an $n > 0$. By definition of the relation \parallel , a necessary condition for two 4-cycles to be related is to have their two $*$ s in the same position. There are $\binom{n}{2}$ ways of choosing the positions of two $*$ s in a

vector of size n , which implies that there are at most $\binom{n}{2}$ components in $G [Q_2^n]$. This gives us a partition of the set of vertices, which we represent by

$$V^n = \bigcup_{i=1}^{\binom{n}{2}} V_i^n.$$

Set $1 \leq i \leq \binom{n}{2}$; in what follows we prove that the induced subgraph of V_i^n in $G [Q_2^n]$ is isomorphic to Q_1^{n-2} . Any element in V_i^n has the two *s in the same position, and the rest of the $n - 2$ entries have all possible binary entries. This implies that $|V_i^n| = |Q_0^{n-2}|$. Let $\phi : V_i^n \rightarrow Q_0^{n-2}$ be the natural bijection between these two sets of vertices. It is clear from Definition 2.3 that two 4-cycles x and y in V_i^n are connected in $G [Q_2^n]$ if and only if the Hamming distance of $\phi(x)$ and $\phi(y)$ is 1. This implies that $G [V_i^n] \cong Q_1^{n-2}$. \square

In this paper we restrict ourselves to complexes $Q \sim Q_2(n, p)$, and thus the graphs $G [Q]$ that we analyse are always subgraphs of $G [Q_2^n]$. We say that a vertex in V^n is coloured if the 4-cycle that it represents has its 2-face present in Q , and we say that it is not coloured otherwise. We use the previous established partition of V^n in $\binom{n}{2}$ sets, to denote accordingly the induced subgraphs $G_1, \dots, G_{\binom{n}{2}}$ of the graph $G [Q_2^n]$. For a $Q \sim Q_2(n, p)$, this partition defines $\binom{n}{2}$ random subgraphs, which we represent by $G_1 [Q], \dots, G_{\binom{n}{2}} [Q]$. Then the edges that are included in $G_i [Q]$, for $1 \leq i \leq \binom{n}{2}$, depend on the 2-faces included in Q . The next lemma characterises the probability distribution of each $G_i [Q]$:

Lemma 2.6. *Set $p \in (0, 1)$ and let $Q \sim Q_2(n, p)$. Then for $1 \leq i \leq \binom{n}{2}$, each random graph $G_i [Q]$ is a random graph on Q_1^{n-2} , with each edge included independently with probability p^4 and each vertex coloured independently with probability p . Moreover, the vertex colourings are independent of the edge set. Using the notation established in the Introduction, the uncoloured graph $G_i [Q]$ has the same distribution as $Q (n - 2, p^4)$ for all $1 \leq i \leq \binom{n}{2}$.*

Proof. Set $Q \sim Q_2(n, p)$. From Lemma 2.5, we know that each $G_i [Q]$ is a random graph on Q_1^{n-2} . Let x and y be two vertices in $G_i [Q]$ that are connected in $G [Q_2^n]$, and represent this edge by \overline{xy} . This implies in particular that x and y are 4-cycles that have, with the previously defined star notation, the * in the same entries and Hamming distance equal to 1. Let $c_{\overline{xy}} \in Q^n$ be the unique 3-dimensional cube that contains x and y .

The probability of \overline{xy} being an edge in $G_i [Q]$ is equal to the probability of the other 4-cycles in $c_{\overline{xy}}$ being covered by 2-faces in Q . This event happens with probability p^4 because in $Q_2(n, p)$ each 2-face is added independently with probability p .

Moreover, observe that any of these 4-cycles in $c_{\overline{xy}}$ are not vertices in $G_i [Q]$, because they do not have the two *s in the same location as x (or y). This implies the independence between the colouring of the vertices and the inclusion of the edges in $G_i [Q]$.

Let \mathcal{C} be the set of all 3-dimensional cubes $c_{\overline{xy}}$ with x and y varying among all unordered pairs of vertices in $G_i [Q]$ that are connected in $G [Q_2^n]$. Then, by the uniqueness of the cube $c_{\overline{xy}} \in Q^n$, we have that $|\mathcal{C}|$ is equal to the number of edges in $G_i [Q]$. Finally, edges in $G_i [Q]$ are added independently with probability p^4 because each 4-face in a cube in \mathcal{C} appears in only one 3-dimensional cube in \mathcal{C} . \square

Remark. Even if the edges within each subgraph $G_i [Q]$ of the graph $G [Q]$ are included independently with probability p^4 , this independence does not extend to events related to edges selected from two or more of the subgraphs $G_i [Q]$. An illustrative example of this dependence can be easily seen in the 3-dimensional cube where there is only one cube containing all the 4-cycles.

2.3. The sizes of the components of $Q(n, p)$

For a given $p \in (0, 1)$, we want to understand the sizes of the components of $Q(n, p)$. We will require an argument from [14] that rules out components of small sizes. Because we slightly adapt those lemmas, we include proofs.

Denote by \mathcal{Q}_s the set of all subsets of vertices in Q^n which are connected and have cardinality s . Given a subset S of vertices in Q^n that are connected, define

$$b(S) = |\{(u, v) \in Q^n \mid (u, v) \text{ is an edge in } Q^n, u \in S, \text{ and } v \notin S\}|. \tag{3}$$

Let

$$g(s) = \sum_{S \in \mathcal{Q}_s} (1 - p)^{b(S)}. \tag{4}$$

Then from a union bound, $g(s)$ is an upper bound for the probability of the existence of a connected component on s vertices appearing in $Q(n, p)$.

Lemma 2.7. *We have*

$$g(s) \leq 2^n (ns)^s (1 - p)^{s(n - \lfloor \log_2(s) \rfloor)}. \tag{5}$$

Proof. Set $s \geq 1$; then for any $S \in \mathcal{Q}_s$ we have (from [16])

$$b(S) \geq s(n - \lfloor \log_2(s) \rfloor). \tag{6}$$

Also, using the fact that the degree of each vertex of Q^n is at most n ,

$$|\mathcal{Q}_s| \leq 2^n (n)(2n)(3n) \cdots ((s - 1)n) \leq 2^n (ns)^s. \tag{7}$$

Hence,

$$g(s) \leq \sum_{S \in \mathcal{Q}_s} (1 - p)^{s(n - \lfloor \log_2(s) \rfloor)} \leq 2^n (ns)^s (1 - p)^{s(n - \lfloor \log_2(s) \rfloor)}. \tag{8}$$

□

Lemma 2.8. *For any $p \in (0, 1)$, there is a number $T_p \in \mathbb{N}$ and there exist $\delta, \epsilon > 0$ such that*

$$\sum_s g(s) < 2^{-\delta n} \tag{9}$$

with the sum over all s such that $T_p \leq s \leq 2^{\epsilon n}$.

Proof. Let T_p be defined by

$$T_p = \inf_{T \in \mathbb{N}} \{2 \cdot (1 - p)^T\} < 1.$$

Then for $s \leq 2^{\epsilon n}$, by Lemma 2.7,

$$g(s) \leq 2^n (ns)^s (1 - p)^{s(n - \lfloor \log_2(s) \rfloor)} \leq 2^n (1 - p)^{-2\epsilon n} (n2^{\epsilon n} (1 - p)^n)^s$$

for all n sufficiently large. Then

$$\begin{aligned} \sum_{s=T_p}^{\lfloor 2^{\epsilon n} \rfloor} g(s) &\leq 2^n (1 - p)^{-2\epsilon n} \sum_{s=T_p}^{\infty} (n2^{\epsilon n} (1 - p)^n)^s \\ &\leq 2^n (1 - p)^{-2\epsilon n} (n2^{\epsilon n} (1 - p)^n)^{T_p} (1 + o(1)), \end{aligned}$$

provided ϵ is chosen so that $2^\epsilon (1 - p) < 1$ and n is taken large. Taking ϵ sufficiently small,

$$\alpha = 2^{1+\epsilon T_p} (1 - p)^{T_p - 2\epsilon} < 1.$$

Hence, in terms of α ,

$$\sum_{s=T_p}^{\lfloor 2^{\epsilon d} \rfloor} g(s) \leq \alpha^n n^{T_p} (1 + o(1)) \leq 2^{-\delta n}$$

for some $\delta > 0$ sufficiently small and all n sufficiently large. □

2.4. The threshold for maximal 1-faces

Any element $Q \sim Q_2(n, p)$ has $2^{n-1}n$ edges, which we represent by

$$e_1, e_2, \dots, e_{2^{n-1}n},$$

with each one of these edges being in $n - 1$ different 4-cycles. We represent by I_i the indicator function of the event that the edge e_i is maximal – that is, that none of the $(n - 1)$ 4-cycles that contain e_i has an attached 2-face. Then

$$\mathbb{E}[I_i] = (1 - p)^{n-1}.$$

Let $\mathcal{I}(Q)$ be the random variable that counts the number of maximal edges in Q :

$$\mathcal{I} = \sum_{i=1}^{2^{n-1}n} I_i.$$

Then

$$\mathbb{E}[\mathcal{I}] = 2^{n-1}n(1 - p)^{n-1}. \tag{10}$$

Observe that if $p = 1/2$, then $\mathbb{E}[\mathcal{I}] = n$.

We now prove that Theorem 1.2 follows from Theorem 1.3:

Proof of Theorem 1.2. We first establish that for $p > \frac{1}{2}$, $\mathcal{I} = 0$ w.h.p., and for $p \leq \frac{1}{2}$, $\mathcal{I} \geq 2$ w.h.p. For the first claim, the expectation (10) tends to 0. For the second, again from equation (10), if $(1 - p) \geq 1/2$ for a random 2-cubical complex, $Q \sim Q_2(n, p)$, then

$$\mathbb{E}[\mathcal{I}] = 2^{n-1}d(1 - p)^{n-1} \geq n.$$

Thus $\mathbb{E}[\mathcal{I}] \rightarrow \infty$ as $n \rightarrow \infty$. Now, we use a second moment argument (see [1, Corollary 4.3.5]) to prove that $\mathbb{P}[\mathcal{I} \geq 2] \rightarrow 1$ as $n \rightarrow \infty$.

Fix an edge e_i . Any other edge e_j such that I_j is not independent from I_i – we represent this nonindependent relation between edges e_i and e_j by $j \sim i$ – will be an edge of one and only one of the $(n - 1)$ 4-cycles that contain e_i . There are $3(n - 1)$ such edges and $\mathbb{P}[I_j | I_i] = (1 - p)^{n-2}$. If we define $\Delta_i^* = \sum_{j \sim i} \mathbb{P}[I_j | I_i]$, then

$$\Delta_i^* = \sum_{j \sim i} \mathbb{P}[I_j | I_i] = 3(n - 1)(1 - p)^{n-2}.$$

Thus $\Delta_i^* = o(\mathbb{E}[I_n])$, which implies that $\mathbb{P}[\mathcal{I} > n/2] \rightarrow 1$ as $n \rightarrow \infty$.

Hence from Theorem 1.3, we have that for $p > \frac{1}{2}$, $\pi_1(Q_2(n, p)) = 0$ with high probability, and for any $p < \frac{1}{2}$, $\pi_1(Q_2(n, p)) = G * \mathbb{Z} * \mathbb{Z}$ for some group G with high probability. The event that $Q_2(n, p)$ has such a free factorisation is a *decreasing* event, in that for any complex Q that satisfies $\pi_1(Q) = G * \mathbb{Z} * \mathbb{Z}$

for some group G , removing any 2-face (i.e., removing relations from $\pi_1(Q)$) yields a complex Q' so that $\pi_1(Q') = G' * \mathbb{Z} * \mathbb{Z}$ for some other group G' . It follows that for any $p \leq \frac{1}{2}$,

$$\mathbb{P}[\exists G : \pi_1(Q_2(n, p)) = G * \mathbb{Z} * \mathbb{Z}] \geq \mathbb{P}\left[\exists G : \pi_1\left(Q_2\left(n, \frac{1}{2}\right)\right) = G * \mathbb{Z} * \mathbb{Z}\right] \rightarrow 1$$

as $n \rightarrow \infty$, which completes the proof. □

3. Parallel homotopy algorithm

In this section, we introduce a simple iterative algorithm for finding contractible 4-cycles. For $Q_2(n, p)$ with $p > 0$, this algorithm rapidly and dramatically simplifies the fundamental group to its nontrivial parts.

We begin by introducing the algorithm. We have defined V^n as the set of all 4-cycles in Q_1^n . For any subset $V \subset V^n$, we define the *graph of parallel related 4-cycles* denoted by $G(V)$ in a similar fashion to Definition 2.4: the vertex set of $G(V)$ is given by the V^n , and two 4-cycles x and y are connected if they have $*$ s in the same positions and are contained in a 3-cube c and all other 4-cycles in c are in V . That is, two 4-cycles x and y are connected in $G(V)$ if they are connected in $G[Q]$, with Q being the cubical complex $Q \subset Q_2^n$ such that the 2-faces included in Q are exactly those corresponding to the 4-cycles that are in V .

Given a $Q \sim Q_2(n, p)$, we denote by V_1^n the subset of V^n which contains the boundaries of 2-faces in Q . We then iteratively run the following procedure, with $t \in \mathbb{N}$:

Stage t: Build the graph of parallel related 4-cycles $G(V_t^n)$. Define the set of 4-cycles V_{t+1}^n as the set of 4-cycles that are in a component in $G(V_t^n)$ that contains a 4-cycle in V_t^n .

The algorithm stops at the first t for which $V_{t+1}^n = V_t^n$. This will happen, in particular, if there exists a t such that $V_t^n = V^n$, in which case all 4-cycles of Q^n are in a component of $G(V_t^n)$ that contains a 4-cycle in V_t^n . Thus, Theorem 1.2 follows from the following result:

Theorem 3.1. For $p > 1/2$,

$$\lim_{n \rightarrow \infty} \mathbb{P}\left[V_3^n = V^n\right] = 1.$$

3.1. Stage 1: Explosive growth

For any set of 4-cycles $V \subset V^n$, say that a set of vertices S in $G(V)$ is a *quasicomponent* if S is connected in $G(V^n)$ and S is disconnected from its complement in $G(V)$.

Theorem 3.2. Set $Q \sim Q_2(n, p)$. For any $p \in (0, 1)$, there exists an integer T_p so that w.h.p. there is no quasicomponent of $G(V_1^n)$ bigger than T_p that is disjoint from V_1^n .

Proof. The proof will follow from two key steps. Let A_s be the event that a quasicomponent of size s in $G(V_1^n)$ exists.

(Step i) For any $p \in (0, 1)$, there exist an integer T_p and $\epsilon, \delta > 0$ such that for all n sufficiently large,

$$\mathbb{P}\left[\bigcup_{s=T_p}^{2^{\epsilon n}} A_s\right] < 2^{-\delta n}. \tag{11}$$

(Step ii) For any $\epsilon > 0$, the probability that there exists a component of $G(V_1^n)$ bigger than $2^{\epsilon n}$ with no vertex in V_1^n tends to zero with n .

We show first how to complete the proof from this point. We now suppose that there is no quasicomponent of $G(V_1^n)$ of size between T_p and $2^{\epsilon n}$ for some integer T_p and some $\epsilon > 0$, and we further suppose that there is no component of $G(V_1^n)$ of size $2^{(\epsilon/2)n}$ with no vertex in V_1^n .

Let X be a minimal quasicomponent of $G(V_1^n)$ which is bigger than T_p and which is disjoint from V_1^n , if it exists. Then the size of X must be at least $2^{\epsilon n}$. As it is so large (in particular larger than $2^{(\epsilon/2)n}$), it cannot be that X is connected. So we must be able to remove a connected component (say by building a spanning tree in $G(V)$ of X by connecting spanning trees of its components and removing a component which contains a leaf) to form a smaller quasicomponent X' . The component we remove must be smaller than $2^{(\epsilon/2)n}$, and so for all n sufficiently large, we conclude that X' is a strictly smaller quasicomponent of $G(V_1^n)$ which is bigger than T_p and which is disjoint from V_1^n – a contradiction.

We turn to proving Step (i). For any s , we have $\mathbb{P}[A_s] \leq g(s)$ (recalling g from equation (4)), as $g(s)$ is the expected number of quasicomponents of size s in $Q_2(n, p)$, and hence from a union bound,

$$\mathbb{P}\left[\bigcup_{s=T_p}^{2^{\epsilon n}} A_s\right] \leq \sum_s g(s),$$

with the sum over all $T_p \leq s \leq 2^{\epsilon n}$. The existence of ϵ, δ, T_p such that formula (11) holds is then a direct consequence of Lemma 2.8.

We turn to showing Step (ii). Let W be the event that there exists a component in $G(V_1^n)$ of size greater than or equal than $2^{\epsilon n}$ that does not intersect V_1^n . We show in what follows that $\mathbb{P}[W] \rightarrow 0$ as $n \rightarrow \infty$. First, we observe that $G(V_1^n) = G[Q]$ and that the vertices in V_1^n are precisely the coloured vertices in $G[Q]$, which by Lemma 2.6 are coloured independently with probability equal to p . For $1 \leq i \leq \binom{n}{2}$, define W_i as the event that there exists in $G_i[Q]$ a component of size greater than or equal to $2^{\epsilon n}$ that has all its vertices uncoloured. Thus, by Lemma 2.6,

$$W = \bigcup_{i=1}^{\binom{n}{2}} W_i. \tag{12}$$

Set $1 \leq i \leq \binom{n}{2}$. Conditioned on knowing $G_i[Q]$ – in particular on knowing that there are exactly l components with uncoloured vertices and with sizes s_1, s_2, \dots, s_l greater than $2^{\epsilon n}$ in $G_i[Q]$ – we get

$$\mathbb{P}[W_i \mid G_i] \leq \sum_{k=1}^l (1-p)^{s_k}. \tag{13}$$

Observing that $s_1 + s_2 + \dots + s_l \leq 2^{n-2}$, it has to be the case that $l \leq 2^{n-2}$, and because $(1-p) < 1$, we have $(1-p)^{s_k} \leq (1-p)^{2^{\epsilon n}}$ for all $1 \leq k \leq l$. Thus from formula (13) we get

$$\mathbb{P}[W_i \mid G_i] \leq \sum_{k=1}^l (1-p)^{2^{\epsilon n}} \leq 2^{n-2} (1-p)^{2^{\epsilon n}}. \tag{14}$$

This implies that $\mathbb{E}[\mathbb{P}[W_i \mid G_i]] \leq 2^{n-2} (1-p)^{2^{\epsilon n}}$, and thus

$$\mathbb{P}[W_i] \leq 2^{n-2} (1-p)^{2^{\epsilon n}} \tag{15}$$

for all $1 \leq i \leq \binom{n}{2}$. Finally, by a union bound argument on equation (12) and inequality (15), we have that

$$\mathbb{P}[W] \leq \binom{n}{2} 2^{n-2} (1-p)^{2^{\epsilon n}}, \tag{16}$$

with

$$\lim_{n \rightarrow \infty} \binom{n}{2} 2^{n-2} (1-p)^{2^{\epsilon n}} = 0. \tag{17} \quad \square$$

3.2. Stage 2: Only local defects remain

Let \mathcal{F} be the event that there is no quasicomponent of $G(V_1^n)$ bigger than T_p that is disjoint from V_1^n . This event was shown to hold w.h.p. by Theorem 3.2.

Lemma 3.3. *In the event \mathcal{F} , any 4-cycle ν with at least T_p neighbors in $G(V_2^n)$ is in V_3^n . Likewise, any 4-cycle ν with at least T_p neighbors in $G(V_1^n)$ is in V_2^n .*

Proof. Suppose \mathcal{F} holds, and let ν be any 4-cycle. Suppose that ν has at least T_p neighbors in $G(V_1^n)$. Then the connected component of ν in $G(V_1^n)$ has at least T_p neighbors, and therefore this connected component intersects V_1^n . It follows by the definition of V_2^n that $\nu \in V_2^n$.

Suppose now that ν has at least T_p neighbors in $G(V_2^n)$. We would like to show that at least one of these neighbors is also in V_2^n , for then in the next step of the parallel transport algorithm, we would have $\nu \in V_3^n$. We may also suppose that ν is not in a component of $G(V_1^n)$ that intersects V_1^n , for if it is, then $\nu \in V_2^n$ and we are done.

Suppose by way of a contradiction that none of the neighbors of ν is in V_2^n , which implies that each is in a component of $G(V_1^n)$ disjoint from V_1^n . Hence the union of these components and the component of $G(V_1^n)$ containing ν is a quasicomponent of $G(V_1^n)$ that is disjoint from V_1^n . Moreover, it is a quasicomponent which is larger than T_p , which is disjoint from V_1^n . This does not exist in \mathcal{F} , and therefore ν has a neighbor in V_2^n . Hence $\nu \in V_3^n$. □

We will show that as a consequence of Lemma 3.3, in Stage 2 all those 4-cycles whose every constituent edge has a high enough degree will be collapsed. For any p , define

$$M_p = \inf_{M>0} \mathbb{P} \left(\text{Binomial} \left(\lfloor M/4 \rfloor, p^3 \right) < T_p \right) < \left(\frac{1}{2} \right)^{1/4}. \tag{18}$$

For any 1-face f in Q^n , define $\text{deg}(f)$ as the number of 2-faces in Q containing f . Call a 1-face of $Q \sim Q_2(n, p)$ *light* if its degree is less than or equal to M_p ; otherwise, call it *heavy*. We show that 4-cycles made from heavy edges are contracted in the second stage of the algorithm:

Lemma 3.4. *For any $p \in (0, 1)$, with probability tending to 1 as $n \rightarrow \infty$, every 4-cycle whose every 1-face is heavy is contained in V_3^n .*

We will introduce some additional notation for working with faces of Q . For two disjoint sets $U, W \subset [n]$, let (U^*, W^1) denote the $|U|$ -dimensional face of Q with *s in the positions given by U and 1s exactly in the positions given by W .

Using symmetry, it will be enough to analyse the 4-cycle $(\{1, 2\}^*, \emptyset^1)$. With M_p from equation (18), define \mathcal{E} as the event that all the 1-faces in the 4-cycle $(\{1, 2\}^*, \emptyset^1)$ are heavy – that is,

$$\mathcal{E} = \left\{ \begin{aligned} &\text{deg} \left(\left(\{1\}^*, \emptyset^1 \right) \right) > M_p, \text{deg} \left(\left(\{1\}^*, \{2\}^1 \right) \right) > M_p, \\ &\text{deg} \left(\left(\{2\}^*, \emptyset^1 \right) \right) > M_p, \text{deg} \left(\left(\{2\}^*, \{1\}^1 \right) \right) > M_p \end{aligned} \right\}.$$

To prove Lemma 3.4, it suffices to show the following:

Lemma 3.5. *For any $p \in (0, 1)$, there is an $\epsilon > 0$ such that*

$$\mathbb{P} \left[\mathcal{E} \cap \mathcal{F} \cap \left\{ \text{the degree of } \left(\{1, 2\}^*, \emptyset^1 \right) \text{ in } G(V_2^n) \text{ is less than } T_p \right\} \right] \leq n^{O(1)} 2^{-(1+\epsilon)n}.$$

Proof. The possible neighbors of $(\{1, 2\}^*, \emptyset^1)$ in $G(V^n)$ all have the form $(\{1, 2\}^*, \{j\}^1)$ for some $3 \leq j \leq n$. To have an edge between these 4-cycles in $G(V_2^n)$, we must have

$$\begin{aligned} (\{1, j\}^*, \emptyset^1) &\in V_2^n, & (\{1, j\}^*, \{2\}^1) &\in V_2^n, \\ (\{2, j\}^*, \emptyset^1) &\in V_2^n, & (\{2, j\}^*, \{1\}^1) &\in V_2^n. \end{aligned}$$

In the event \mathcal{F} , we must only lower-bound the degree of these 4-cycles in $G(V_1^n)$ to ensure that they are in V_2^n . Hence, define

$$\begin{aligned} Y_{1j} &= \mathbf{1} \left\{ \deg \left((\{1, j\}^*, \emptyset^1) \right) \geq T_p \right\}, & Y_{2j} &= \mathbf{1} \left\{ \deg \left((\{1, j\}^*, \{2\}^1) \right) \geq T_p \right\}, \\ Y_{3j} &= \mathbf{1} \left\{ \deg \left((\{2, j\}^*, \emptyset^1) \right) \geq T_p \right\}, & Y_{4j} &= \mathbf{1} \left\{ \deg \left((\{2, j\}^*, \{1\}^1) \right) \geq T_p \right\}. \end{aligned} \tag{19}$$

Here, ‘deg’ refers to the degree of the 4-cycle in $G(V_1^n)$. We would like to show that there are at least T_p choices j for which all $Y_{\ell j}$, $\ell \in \{1, 2, 3, 4\}$, are 1.

In the event \mathcal{E} , there are four disjoint sets $R_\ell \subset \{3, 4, \dots, d\}$ for $\ell \in \{1, 2, 3, 4\}$ of size $\lfloor M_p/4 \rfloor$ such that

$$\begin{aligned} (\{1, k\}^*, \emptyset^1) &\in V_1^n \text{ for } j \in R_1, & (\{1, k\}^*, \{2\}^1) &\in V_1^n \text{ for } j \in R_2, \\ (\{2, k\}^*, \emptyset^1) &\in V_1^n \text{ for } j \in R_3, & (\{2, k\}^*, \{1\}^1) &\in V_1^n \text{ for } j \in R_4. \end{aligned}$$

Observe that the possible neighbors of $(\{1, j\}^*, \emptyset^1)$, $j \in \{3, 4, \dots, n\}$, are given by $(\{1, j\}^*, \{k\}^1)$ for $k \notin \{1, j\}$. For simplicity, we will also discard the case $k = 2$. To have this edge in $G(V_1^n)$, we would need

$$\begin{aligned} (\{1, k\}^*, \emptyset^1) &\in V_1^n, & (\{1, k\}^*, \{j\}^1) &\in V_1^n, \\ (\{j, k\}^*, \emptyset^1) &\in V_1^n, & (\{j, k\}^*, \{1\}^1) &\in V_1^n. \end{aligned}$$

In particular, for $k \in R_1$ the first of these requirements is guaranteed. Hence we can define

$$Z_{1jk} = \mathbf{1} \left\{ (\{1, k\}^*, \{j\}^1) \in V_1^n, (\{j, k\}^*, \emptyset^1) \in V_1^n, (\{j, k\}^*, \{1\}^1) \in V_1^n \right\},$$

and define

$$Z_{1j} = \sum_{k \in R_1} Z_{1jk}.$$

Then Z_{1j} is a lower bound for $\deg((\{1, j\}^*, \emptyset^1))$, and so if Z_{1j} is at least T_p , then $Y_{1j} = 1$.

We do a similar construction for $\ell \in \{2, 3, 4\}$, making appropriate modifications. We list them for clarity:

$$\begin{aligned} Z_{2jk} &= \mathbf{1} \left\{ (\{1, k\}^*, \{2, j\}^1) \in V_1^n, (\{j, k\}^*, \{2\}^1) \in V_1^n, (\{j, k\}^*, \{1, 2\}^1) \in V_1^n \right\}, \\ Z_{3jk} &= \mathbf{1} \left\{ (\{2, k\}^*, \{j\}^1) \in V_1^n, (\{j, k\}^*, \emptyset^1) \in V_1^n, (\{j, k\}^*, \{1\}^1) \in V_1^n \right\}, \\ Z_{4jk} &= \mathbf{1} \left\{ (\{2, k\}^*, \{1, j\}^1) \in V_1^n, (\{j, k\}^*, \{1\}^1) \in V_1^n, (\{j, k\}^*, \{1, 2\}^1) \in V_1^n \right\}. \end{aligned}$$

In terms of these, we set $Z_{\ell j} = \sum_{k \in R_\ell} Z_{\ell jk}$.

Let $J = \{3, 4, \dots, d\} \setminus (\cup_1^4 R_\ell)$. Then the family

$$\{Z_{\ell jk} : \ell \in \{1, 2, 3, 4\}, j \in J, k \in R_\ell\}$$

is independent random variables. Moreover, for any $\ell \in \{1, 2, 3, 4\}$ and $j \in J$, from equation (18) we have

$$\mathbb{P}(Z_{\ell j} < T_p) < \mathbb{P}(\text{Binomial}(\lfloor M_p/4 \rfloor, p^3) < T_p) \leq \left(\frac{1}{2}\right)^{1/4}.$$

It follows that with

$$Z = \sum_{j \in J} \prod_{\ell=1}^4 \mathbf{1}\{Z_{\ell j} \geq T_p\}$$

and with $q = \mathbb{P}(Z_{\ell j} \geq T_p)^4 > \frac{1}{2}$, we have

$$\mathbb{P}(Z < T_p) \leq \mathbb{P}(\text{Binomial}(n - 3 - M_p, q) < T_p) = n^{O(1)}(1 - q)^n,$$

which completes the proof. □

3.3. Stage 3: The final squeeze

In this section, we draw conclusions on what remains noncontracted in the complex in the third stage.

3.3.1. The simply connected regime, $p > \frac{1}{2}$

We begin by showing that for $p > 1/2$, there are simply no light 1-faces. Hence in fact for $p > \frac{1}{2}$, $V_3^n = V_n$ with high probability (proving Theorem 3.1).

Lemma 3.6. *For any $p > 1/2$, there is an $M_p > 0$ such that with probability tending to 1 with n , for every 1-face f of $Q \sim Q_2(n, p)$, $\deg(f) > M_p$.*

Proof. The degree of a 1-face is distributed as Binomial($n - 2, p$). For $p > \frac{1}{2}$, the probability that this is less than any fixed constant M is $n^{O(1)}(1 - p)^n$. Hence by a union bound, the lemma follows. □

3.3.2. Completely shielded 1-faces

Call a 1-face $f \in Q \sim Q_2(n, p)$ *completely shielded* if every 3-face $c \in Q^n$ that contains f contains only heavy 1-faces of Q , besides possibly f . Completely shielded 1-faces modify the fundamental group of Q in a simple way, contributing a free factor of \mathbb{Z} if f is maximal.

To see this we begin with the following definition:

Definition 3.7. Let f be any 1-face of Q^n . Define the *n -bubble around f* to be the subcubical complex of Q^n given by the union of the complete 1-skeletons of all 3-faces containing f and every 2-face on this skeleton which does *not* contain f .

An n -bubble has fundamental group \mathbb{Z} .

Lemma 3.8. *For any $n \geq 3$ and any n -bubble X around f ,*

$$\pi_1(X) \cong \mathbb{Z}.$$

Furthermore, the complex $X \setminus \{f\}$ and the complex $X \cup \{e\}$, where e is any 2-face containing f , are simply connected.

Proof. Without loss of generality, suppose that f is the face $(\{1\}^*, \emptyset^1)$. The 3-faces containing f all have the form $(\{1, i, j\}^*, \emptyset^1)$, and so the 1-skeleton of X is

$$\left\{ (\{i\}^*, A^1) : A \subset \{1, 2, \dots, n\}, i \notin A, |A \cup \{i\}| \leq 3 \right\}.$$

We claim that all the 4-cycles containing f are homotopic. As all other 4-cycles are contractible from the definition of X , the statements in the lemma follow.

The 4-cycles that contain f are boundaries of the 2-faces of Q^n of the form

$$\left\{ (\{1, i\}^*, \emptyset^1) : 2 \leq i \leq n \right\}.$$

For any $2 \leq i < j \leq n$, the 3-face $c = (\{1, i, j\}^*, \emptyset)$ intersected with X contains four 2-faces. Moreover, the 2-faces $(\{1, i\}^*, \emptyset)$ and $(\{1, j\}^*, \emptyset)$ are adjacent in this cube. Hence, these cycles can be deformed through c to one another. As this holds for any such i and j , the proof follows. \square

Lemma 3.9. *For any $p \in (0, 1)$, let \hat{Q} be the cubical complex that results from deleting from $Q \sim Q_2(n, p)$ every completely shielded 1-face f and any 2-face of Q containing f . Then with high probability,*

$$\pi_1(Q) \cong \pi_1(\hat{Q}) * \underbrace{(\mathbb{Z} * \mathbb{Z} * \dots * \mathbb{Z})}_N,$$

where N denotes the number of completely shielded 1-faces in Q that are isolated.

Proof. From Lemma 3.4, all 4-cycles whose every 1-face is heavy are contractible. In particular, we do not modify the fundamental group of Q if we include all those 2-faces into Q whose boundary is in V_3^n . Let \hat{Q} be this cubical complex.

We now remove completely shielded 1-faces from \hat{Q} one at a time, tracking the changes to the fundamental group. We will show what happens after removing the first. It will be clear that by using induction, a similar analysis would give the claim in the lemma.

Let f be a completely shielded 1-face of \hat{Q} . Let Q_1 be the complex that results after removing f from \hat{Q} and any 2-face containing f , and let Q_2 be the union of all the complete 2-skeletons of all 3-faces that contain f . Then Q_2 contains an n -bubble, and it is exactly an n -bubble if f is isolated.

As $Q_2 \cup Q_1 = \hat{Q}$ and $Q_1 \cap Q_2$ is open and path-connected (compare Lemma 3.8, as this complex is an n -bubble with its central 1-face deleted). Moreover, every 4-cycle in $Q_1 \cap Q_2$ is contractible, and so $\pi_1(Q_1 \cap Q_2)$ is trivial. From the Siefert–van Kampen theorem, we therefore have

$$\pi_1(\hat{Q}) \cong \pi_1(Q_1) * \pi_1(Q_2).$$

If f is maximal, then from Lemma 3.8 the fundamental group $\pi_1(Q_2)$ is isomorphic \mathbb{Z} . \square

3.3.3. The velvety bubble phase

For $p > 1 - \left(\frac{1}{2}\right)^{1/2} \approx 0.292893$, we further show that the fundamental group completely reduces to its maximal 1-faces. In this phase, while light 1-faces may exist in $Q_2(n, p)$ (for $p \leq \frac{1}{2}$), they are well separated.

Lemma 3.10. *For $p > 1 - \left(\frac{1}{2}\right)^{1/2}$, with high probability, there are no 3-faces $c \in Q^n$ that contain more than one light 1-face of $Q \sim Q_2(n, p)$.*

Proof. For a fixed c and a fixed choice of two 1-faces f_1, f_2 , the degrees of f_1 and f_2 are both light with probability at most $n^{O(1)}(1-p)^{2n-1}$. Hence for any p as in the statement of the lemma, there is an $\epsilon > 0$

such that the probability that this occurs is $2^{-(1+\epsilon)n+O(\log n)}$. As there are $2^n n^{O(1)}$ many ways to pick a 3-face with two designated edges, the lemma follows from a first moment estimate. \square

We now give the proof of Theorem 1.3, which we recall for convenience:

Theorem 3.11. For $p > 1 - \left(\frac{1}{2}\right)^{1/2}$, with high probability, for $Q \sim Q_2(n, p)$,

$$\pi_1(Q) \cong \underbrace{(\mathbb{Z} * \mathbb{Z} * \cdots * \mathbb{Z})}_N,$$

where N denotes the number of maximal 1-faces in Q .

Proof. From Lemma 3.4, with high probability every 4-cycle containing only heavy 1-faces is in V_3^n . From Lemma 3.10, with high probability no 3-faces $c \in Q^n$ contain more than one light face. Hence taking \tilde{Q} as Q together with all 2-faces bounded by some element of V_3^n (so that $\pi_1(\tilde{Q}) = \pi_1(Q)$), every light 1-face f of \tilde{Q} is completely shielded in \tilde{Q} . Moreover, every 4-cycle of \tilde{Q} either intersects a light 1-face or is the boundary of a 2-face. Hence in the notation of Lemma 3.9, $\pi_1(\hat{Q}) = 0$. It follows that from Lemma 3.9 we have a free group on N' generators, with N' the number of completely shielded maximal 1-faces. As every light 1-face is completely shielded w.h.p., it follows that $N' = N$ with high probability. \square

4. Structure theorem for general p

In this section, we prove Theorem 1.7. For convenience, we recall some definitions from the introduction. Recall Definition 1.6:

Definition 4.1. For a cubical subcomplex T of any cube Q^n , let $e(T)$ denote the number of edges in T . Let \mathcal{T}_p be the set of pure 2-dimensional strongly connected cubical complexes T that are subcomplexes of Q_2^n for some n and such that $(1 - \left(\frac{1}{2}\right)^{1/e(T)}) < p$.

We will prove Theorem 1.7, which we recall:

Theorem 4.2. For any $p \in (0, 1)$ and for $Q \sim Q_2(n, p)$, let the free-product factorisation of $\pi_1(Q)$ be given by

$$\pi_1(Q) \cong F * \pi_1(X_1) * \pi_1(X_2) * \cdots * \pi_1(X_\ell),$$

with F a free group. With high probability, any $T \in \mathcal{T}_p$ appears as a factor $\pi_1(X_j)$ for some $1 \leq j \leq \ell$.

Our main technical tool will be the following:

Definition 4.3. For a cubical subcomplex T of a cubical complex $W \subset Q^n$, denote by $h(T)$ the minimal cubical subcomplex of W such that

1. the 1-skeleton of $h(T)$ is the 1-skeleton of a k -dimensional hypercube,
2. every 2-face of W that is incident to T is contained in $h(T)$ and
3. every 2-face of Q_2^n with 1-skeleton in $h(T)$ which is *not* incident to T is in $h(T)$.

Also denote by $H(T) \subset W$ a complete 2-skeleton of a k -dimensional hypercube which is parallel to $h(T)$, so that any 2-face that has an edge in $h(T) \setminus T$ and another edge in $H(T)$ is contained in W .

We emphasise that T need not be connected in any sense and that $H(T)$ is not unique; we just need to choose one.

Lemma 4.4. For any $p \in (0, 1)$, with $Q \sim Q_2(n, p)$, there exists a number k_p such that w.h.p., every $((M_p + 2) \times k_p)^2$ -dimensional cube in Q^n contains fewer than k_p light edges of Q .

Proof. We argue by a first moment estimate. For any ℓ , the number of ℓ -dimensional cubes in Q^n is given by $2^{n-\ell} \binom{n}{\ell}$. The probability that any such cube contains k light edges is $O_{k,\ell,p}((1-p)^{nk})$. Hence taking $\ell = ((M_p + 2)k)^2$, if we pick k sufficiently large that $(1-p)^k < \frac{1}{2}$, then the expected number of ℓ -dimensional cubes containing more than k light edges tends to 0 exponentially in n . \square

Lemma 4.5. *Set $p \in (0, 1)$ and let $\ell \in \mathbb{N}$ be fixed; then w.h.p., for $Q \sim Q_2(n, p)$, every ℓ -dimensional cube X has a parallel cube Y that has no light edges in Q and for which there are no light edges in Q between X and Y .*

Proof. This is similar to Lemma 4.4. We argue by a first moment estimate. For any ℓ , the number of ℓ -dimensional cubes in Q^n is given by $2^{n-\ell} \binom{n}{\ell}$. For a fixed ℓ -dimensional cube $X \subset Q^n$, the probability that every parallel ℓ -dimensional cube Y contains either

- (i) at least one light edge or
- (ii) the endpoint of a light edge between X and Y

is at most

$$\left((\ell + 2)2^{\ell-1} \right)^{n-\ell} (1-p)^{n(n-\ell)} = o\left(n^\ell 2^{-n} \right).$$

Hence from a first moment estimate, for any fixed ℓ and any $p \in (0, 1)$, w.h.p. every ℓ -dimensional cube X has a parallel cube Y that contains no light edges of Q and shares no endpoint of a light edge between X and Y . \square

Theorem 4.6. *Set $p \in (0, 1)$ and k_p as in Lemma 4.4, and $Q \sim Q_2(n, p)$. Let \bar{Q} be Q with all the 2-faces bounded by 4-cycles having no light edges. With high probability, there are disjoint cubical complexes $\{\tau_1, \tau_2, \dots, \tau_\ell\}$ in Q such that*

- (i) *the union of 1-faces over all $\{\tau_j : 1 \leq j \leq \ell\}$ is the set of all light 1-faces;*
- (ii) *for each $1 \leq j \leq \ell$, both $h(\tau_j)$ and $H(\tau_j)$ exist in \bar{Q} ; and*
- (iii) *for each $1 \leq i \neq j \leq \ell$, the Hamming distance between the 0-skeletons of $h(\tau_j)$ and $h(\tau_i)$ is at least 2.*

We need the next definition for proving Theorem 4.6:

Definition 4.7. Let X and Y be two subcomplexes of Q^n . Define $X \square Y$ to be the face of smallest dimension Q^m such that $Q^m \subset Q^n$ and $X \cup Y \subset Q^m$. Observe that $m \leq n$. More generally, let X_1, X_2, \dots, X_l be any finite collection of subcomplexes of Q^n and $I = \{1, 2, \dots, l\}$. We define

$$\square_{i \in I} X_i$$

as the face of smallest dimension Q^m such that $Q^m \subset Q^n$ and such that

$$[X_1 \cup X_2 \cdots \cup X_l] \subset Q^m.$$

In this case, $m \leq n$ as well.

Proof of Theorem 4.6. We first show that for every light 1-face e there is a cubical complex σ_e containing e and having all its 1-faces light such that $h(\sigma_e)$ exists in \bar{Q} . We will then merge these $h(\sigma_e)$ to form the partition claimed to exist in the theorem.

Let $e_1 := e$ be any light edge of Q . Let T_1 be the cubical complex which is the down closure of e_1 . Let X_1 be the smallest induced complex in \bar{Q} which contains T_1 , which contains all 2-faces of \bar{Q} incident to e_1 and whose 1-skeleton is a hypercube. If $X_1 = h(T_1)$, we are finished. Otherwise, by definition, there must be a 2-face f of \bar{Q}_2^n with 1-skeleton in X_1 but not itself in X_1 . Then there must be at least one light edge $e_2 \in X_1 \setminus T_1$.

We then define T_2 as the induced subcomplex of Q on edges e_1, e_2 . Let X_2 be the smallest induced complex in Q which contains T_2 , which contains all 2-faces of Q incident to T_2 and whose 1-skeleton is a hypercube. Once more, if $X_2 = h(T_2)$, we are done. Otherwise, we proceed inductively by the same argument.

This produces a nested sequence of complexes $\{T_k\}$ each having k edges. It also produces a sequence of complexes $\{X_k\}$ such that each $X_k \supset T_k$ and each X_k contains at least k light edges and such that X_k has the 1-skeleton of a hypercube of dimension at most $k \times M_p$. By Lemma 4.4, with high probability, this sequence must terminate at some $k^* \leq k_p$. The complex $X_{k^*} = h(T_{k^*})$ by definition, and we define $\sigma_e = T_{k^*}$.

We define a graph G with vertex set given by the collection of σ_e . Two vertices $\sigma_{e_1}, \sigma_{e_2}$ in this graph are connected if the Hamming distance between $h(\sigma_{e_1})$ and $h(\sigma_{e_2})$ is less than 2. Let $\{\tau_1, \tau_2, \dots, \tau_\ell\}$ be the unions of the connected components in G . Then for each $1 \leq j \leq \ell$, we construct the hypercube

$$\Sigma_j = \square_{e \in \tau_j} h(\sigma_e).$$

It is easy to see that $\Sigma_j = h(\tau_j)$, which implies that $h(\tau_j)$ exists and is exactly Σ_j .

The dimension of Σ_j is at most

$$\sum_{\sigma_e} [\dim(h(\sigma_e)) + 2], \tag{20}$$

where the sum is over all σ_e contained in τ_j . Therefore by Lemma 4.4, each τ_j has at most k_p edges. Hence by Lemma 4.5, each $H(\tau_j)$ exists as well. □

Lemma 4.8. *Let W be a subcomplex of Q_2^n and T a subcomplex of W . Suppose that $h(T)$ and $H(T)$ exist in W . Let \hat{W} be the complex formed by adding to W the complete 1-skeleton of $h(T) \square H(T)$ and any 2-face of Q_2^n with 1-skeleton in $h(T) \square H(T)$. Then*

$$\pi_1(W) \cong \pi_1(\hat{W}) * \pi_1(W \cap (h(T) \square H(T))).$$

Recall that for a disconnected cubical complex X , we define $\pi(X)$ as the free product of its connected components.

Proof. Let \hat{T} be all the 1-faces in T and any 2-face of W incident to T . Let \tilde{T} be the down closure of \hat{T} . Let $S = W \cap (h(T) \square H(T))$.

Let $X = (W \setminus \hat{T}) \cap S$. We claim that $\pi_1(X) \cong 1$. The 1-faces of X that are in $h(T)$ are not in T . Therefore, by Definition 4.3, for every edge $e \in X \cap h(T)$, the unique 4-cycle connecting e to $H(T)$ is the boundary of a 2-face in X . Thus, every closed curve in X is homotopic to a curve in $H(T)$. Since $\pi_1(H(T)) \cong 1$, it follows that $\pi_1(X) \cong 1$. Therefore, the Siefert–van Kampen theorem states that

$$\pi_1(W) \cong \pi_1(W \setminus \hat{T}) * \pi_1(S).$$

We now show that $\pi_1(W \setminus \hat{T}) \cong \pi_1(\hat{W})$. Define a complex S^* as the down closure of all 2-faces in Q_2^n incident to T with 1-skeleton in $h(T) \square H(T)$, union with $H(T)$. Any 1-face e of $S^* \cap (W \setminus \hat{T})$ that is in $h(T)$ must be in $h(T) \setminus T$. In particular, there is a 2-face f containing e which has a 1-face in $H(T)$. Hence, any closed curve in $S^* \cap (W \setminus \hat{T})$ is homotopic to one in $H(T)$, which is simply connected. Therefore by the Siefert–van Kampen theorem,

$$\pi_1(\hat{W}) = \pi_1(S^* \cup (W \setminus \hat{T})) \cong \pi_1(W \setminus \hat{T}) * \pi_1(S^*).$$

It remains to evaluate the fundamental group of S^* . Any edge in $S^* \cap h(T)$ has a 4-cycle that has an edge in $H(T)$. By construction, we know that this 4-cycle has a 2-face added. Therefore, any closed curve in S^* is homotopic to a closed curve in $H(T)$. Thus S^* is simply connected, because $H(T)$ is by definition. □

Lemma 4.9. *Let W be a subcomplex of Q_2^n and T a subcomplex of W . Suppose that $h(T)$ and $H(T)$ exist in W . Let P_1, \dots, P_m be all the pure 2-dimensional strongly connected components completely contained in T , such that any 2-face adjacent to the 1-skeleton of any P_i is also contained in P_i . Suppose that $T = \cup_{i=1}^m P_i$. Then there is a free group F such that*

$$\pi_1(W \cap (h(T) \square H(T))) \cong \pi_1(P_1) * \pi_1(P_2) * \dots * \pi_1(P_m) * F.$$

Proof. Let $S = W \cap (h(T) \square H(T))$. Suppose we fill $H(T)$ by taking the flag cubical complex of $H(T)$. The fundamental group of S is unchanged, and we can contract $H(T)$ to a point x . We denote this complex by \hat{S} . If e is an edge in T , then e forms an unfilled triangle with x in \hat{S} . Let $T_x \subset \hat{S}$ be the union of T, x and all the edges between T and x . Any edge $f \in h(T)$ which is not contained in T is the base of a filled triangle with x in \hat{S} , and so any closed curve in \hat{S} is homotopic to a closed curve in T_x . Hence

$$\pi_1(S) = \pi_1(\hat{S}) = \pi_1(T_x) = \pi_1(P_1) * \pi_1(P_2) * \dots * \pi_1(P_m) * F,$$

where F is a free group. □

Theorem 4.10. *Fix $p \in (0, 1)$. For $Q \sim Q_2(n, p)$, w.h.p., if $\tau_1, \tau_2, \dots, \tau_\ell$ are as constructed in Theorem 4.6, then with $S_j = C \cap (h(\tau_j) \square H(\tau_j))$ for all $1 \leq j \leq \ell$,*

$$\pi_1(Q) \cong \pi_1(S_1) * \pi_1(S_2) * \dots * \pi_1(S_\ell).$$

Proof. Let \bar{Q} be Q with all the 2-faces bounded by 4-cycles having no light edges. By Lemma 3.4, all 4-cycles with no light edges are in V_3^n w.h.p., and so $\pi_1(\bar{Q}) = \pi_1(Q)$. We apply Lemma 4.8 inductively to each of the complexes τ_j . As a result, we have

$$\pi_1(\bar{Q}) = \pi_1(J) * \pi_1(S_1) * \pi_1(S_2) * \dots * \pi_1(S_\ell),$$

where J is the complex \bar{Q} together with all 2-faces in Q_2^n having 1-skeleton contained in some $h(\tau_j) \square H(\tau_j)$ for some $1 \leq j \leq \ell$.

It just remains to prove that $\pi_1(J) \cong 1$. The 1-skeleton of J is Q_1^n , and so it suffices to show that every 4-cycle in J is contractible. The only 4-cycles x in J that do not bound a 2-face are those that contain a 1-face e of some τ_j for $1 \leq j \leq \ell$ but which are not contained in $h(\tau_j) \square H(\tau_j)$. However, as e is in $h(\tau_j)$, it has a parallel 1-face f in $H(\tau_j)$. The unique cube $c = x \square e$ that contains x and e has all 2-faces except for the face bounded by x , which implies that x is contractible. □

Proof of Theorem 1.7. Let $T \in \mathcal{T}_p$ be fixed. By assumption, there is a k -dimensional cube X such that T is a subcomplex of X , and we may choose k minimal. We do not take the full 2-skeleton for X , but instead we choose exactly those 2-faces which either are in T or share no edge with T . Note that this makes $X = h(T)$. Let ϕ be a cubical embedding of the 2-skeleton of X into Q_2^n . Define the event \mathcal{E}_ϕ , for $Q \sim Q_2(n, p)$:

1. The 2-faces of Q that are contained in the 1-skeleton of $\phi(X)$ are exactly the 2-faces of $\phi(X)$.
2. No other 2-face in Q contains a 1-face of $\phi(T)$.
3. There are no light 1-faces in $\phi(X \setminus T)$ and no light 1-faces within Hamming distance $2k_p + 2$ of $\phi(T)$, except possibly those in $\phi(T)$. Here, k_p is defined as in Lemma 4.4.

We now estimate the probability of \mathcal{E}_ϕ under the law of $Q_2(n, p)$. Note that this probability does not depend on ϕ , and so these estimates will be uniform in ϕ . First, observe that each edge of $\phi(T)$ has degree bounded independently of n on this event, so there are $(e(T) * d) - O(1)$ 2-faces which must be absent for \mathcal{E}_ϕ to hold. There are $O(1)$ 2-faces that must be present for \mathcal{E}_ϕ to hold, also. There are also

$O(d^{2k_p+2})$ 1-faces which are contained in the Hamming distance $(2k_p + 2)$ -neighbourhood of $\phi(X)$, which we require to be not light. As the probability that a 1-face is light is $O((1-p)^n)$, we conclude that

$$\mathbb{P}(\mathcal{E}_\phi) = \Theta\left((1-p)^{e(T)d}\right) = \Omega\left(2^{-(1-\epsilon)d}\right),$$

for some $\epsilon > 0$, where the second equality follows from Definition 1.6. So the expected number of occurrences of \mathcal{E}_ϕ goes to infinity exponentially fast as $n \rightarrow \infty$.

We can now show that some \mathcal{E}_ϕ now occurs with high probability by using a second moment computation (see [1, Corollary 4.3.5]). Observe that if the Hamming distance of $\phi(X)$ to $\psi(X)$ is greater than 4, then the events \mathcal{E}_ϕ and \mathcal{E}_ψ are independent. Let $\psi \sim \phi$ if \mathcal{E}_ϕ and \mathcal{E}_ψ are not independent. Then

$$\Delta_\phi^* = \sum_{\psi \sim \phi} \mathbb{P}[\mathcal{E}_\psi \mid \mathcal{E}_\phi] \leq \sum_{\psi \sim \phi} 1 = O(d^{O(1)}),$$

which is much smaller than the expected number of \mathcal{E}_ϕ that occur (which grows exponentially in n).

Hence, with the factorisation given by Theorem 4.10,

$$\pi_1(Q) \cong \pi_1(S_1) * \pi_1(S_2) * \dots * \pi(S_\ell),$$

where $S_j = Q \cap (h(\tau_j) \square H(\tau_j))$ and where τ_j are the complexes from Theorem 4.6. For any embedding ϕ , if \mathcal{E}_ϕ occurs, then $\phi(X) \in Q$ is such that $\phi(X) = h(\phi(T))$. Moreover, $\phi(T) = \tau_j$ for some j with $1 \leq j \leq \ell$, as the Hamming distance of $\phi(T)$ to any other light 1-face is at least $2k_p + 2$. By Lemma 4.9, $\pi_1(S_j) \cong F * \pi_1(T)$ for some free group F . □

5. Below the threshold for maximal edges

In this section, we discuss in slightly more detail the idea that everything that can happen will happen.

First we show that in terms of finitely presented groups, everything can happen. The following is well known (see, e.g., [11, Appendix A] or [3, Section 2]):

Theorem 5.1. *Let S be a finite simplicial complex with k vertices. Then S is homeomorphic to a cubical complex. Indeed, S is homeomorphic to a subcomplex of the k -dimensional cube.*

Theorem 5.1 has the following immediate corollary:

Corollary 5.2. *Let G be any finitely presented group. Then there exists a number $P_G > 0$ such that whenever $0 < p < P_G$ and $Q \sim Q_2(n, p)$, we have that G exists as a free factor in $\pi_1(Q)$ with high probability.*

We give concrete bounds on P_G for a few groups G in the following. The constructions used in Theorems 5.4 and 5.5 and Figure 1 are from joint work of Dejan Govc and the third author of this paper.

Theorem 5.3. *Let T^2 be the 2-dimensional torus. For $0 < p < \left(1 - (1/2)^{1/2^5}\right) \approx 0.021428$, $\pi_1(T_2) \cong \mathbb{Z} \times \mathbb{Z}$ is a free factor of $\pi_1(Q)$ for $Q \sim Q_2(n, p)$ with high probability.*

Proof. We will create a cubical complex which is a subcomplex of Q_2^4 and is homeomorphic to T^2 . Let T_1^\square be the minimal cubical subcomplex of Q_2^4 with 2-faces given by

$$\begin{aligned} &\{(*, 0, 0, *), (0, *, 0, *), (*, 1, 0, *), (1, *, 0, *), \\ &(*, 0, *, 1), (0, *, *, 1), (*, 1, *, 1), (1, *, *, 1), \\ &(*, 0, 1, *), (0, *, 1, *), (*, 1, 1, *), (1, *, 1, *), \\ &(*, 0, *, 0), (0, *, *, 0), (*, 1, *, 0), (1, *, *, 0)\}. \end{aligned}$$

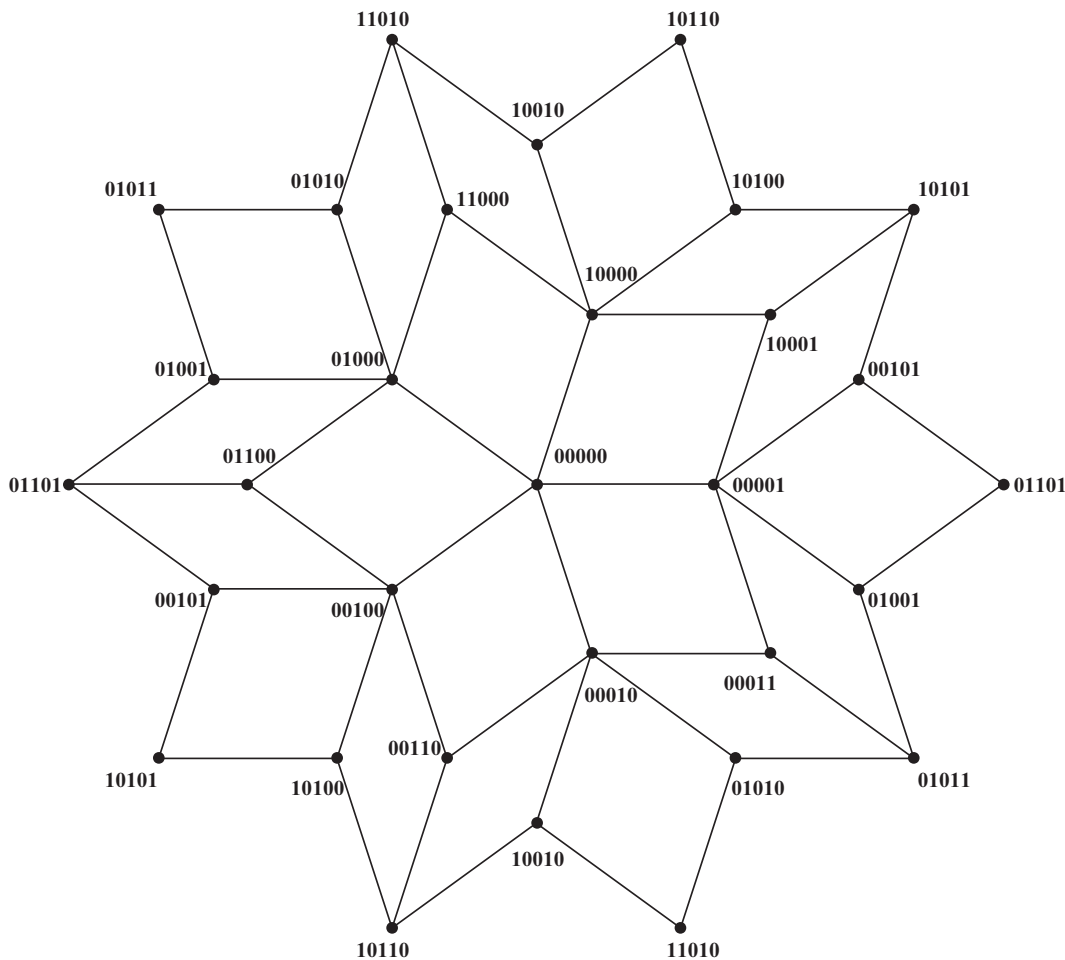


Figure 1. Diagram of T_2^\square with the vertices labeled.

Observe that T_1^\square has $e(T_1^\square) = 32$. Hence, from Theorem 1.7, for $0 < p \neq 0$, if $p < (1 - (1/2)^{1/32})$, the fundamental group of $Q \sim Q_2(n, p)$ has a copy of $\mathbb{Z} \times \mathbb{Z}$ in its free-product factorisation with high probability. \square

Theorem 5.4. Let T_2 be the projective plane. For $p \neq 0$, if $p < (1 - (1/2)^{1/40}) \approx 0.017179$, $\pi_1(T_2) \cong \mathbb{Z}/2\mathbb{Z}$ is a free factor of $\pi_1(Q)$ for $Q \sim Q_2(n, p)$ with high probability.

Proof. Let T_2^\square be the minimal cubical subcomplex of Q_2^5 with 2-faces given by

$$\begin{aligned} &\{(0, 0, 0, *, *), (0, 0, 1, *, *), (0, 0, *, 1, *), (0, 0, *, *, 1), \\ &(0, 1, *, *, 0), (0, *, 0, 0, *), (0, *, 1, *, 0), (0, *, *, 0, 0), \\ &(0, *, *, 1, 0), (1, 0, *, 0, *), (1, *, 0, 0, *), (1, *, 0, *, 0), \\ &(*, 0, 0, *, 0), (*, 0, 1, 0, *), (*, 0, *, 0, 0), (*, 0, *, 0, 1), \\ &(*, 1, 0, 0, *), (*, *, 0, 0, 1), (*, 1, 0, *, 0), (*, *, 0, 1, 0)\}, \end{aligned}$$

so that $e(T_2^\square) = 40$. Hence for $p \neq 0$, if $p < (1 - (1/2)^{1/40}) \approx 0.017179$, the fundamental group of $Q \sim Q_2(n, p)$ has a torsion group in its free-product factorisation with high probability. \square

Theorem 5.5. Let $G = \langle x, y \mid xyx^{-1}y \rangle$ be the fundamental group of the Klein bottle. If $0 < p < (1 - (1/2)^{1/56}) \approx 0.01230134$ and $Q \sim Q_2(n, p)$, then with high probability, $\pi_1(Q)$ has G as a free factor.

Proof. Let T_3^\square be the minimal cubical subcomplex of Q_2^5 with 2-faces given by

$$\begin{aligned} & \{ (*, *, 1, 0, 0), (*, *, 0, 0, 0), (0, *, *, 0, 0), (1, *, *, 0, 0), \\ & (0, 1, *, *, 0), (1, 1, *, *, 0), (*, 1, 0, *, 0), (*, 1, 1, *, 0), \\ & (0, *, *, 1, 0), (1, *, *, 1, 0), (*, *, 1, 1, 0), (*, 0, *, 1, 0), \\ & (0, *, 0, 1, *), (1, *, 0, 1, *), (*, 0, 0, 1, *), (*, 1, 0, 1, *), \\ & (*, 1, *, 1, 1), (0, *, *, 1, 1), (1, *, *, 1, 1), (*, *, 1, 1, 1), \\ & (0, 0, *, *, 1), (1, 0, *, *, 1), (*, 0, 0, *, 1), (*, 0, 1, *, 1), \\ & (0, 0, *, 0, *), (1, 0, *, 0, *), (*, 0, 0, 0, *), (*, 0, 1, 0, *) \}. \end{aligned}$$

Then $e(T_3^\square) = 56$. □

Acknowledgments. We would like to thank Yuval Peled for interesting conversations that helped launch this work. We thank Dejan Govc for the constructions used in Theorems 5.4 and 5.5, and for Figure 1.

Funding statement. The first author was supported by NSF DMS #1547357, DMS #2005630 and CCF #1740761. He is also grateful to the Simons Foundation for a Simons Fellowship, and to the Deutsche Forschungsgemeinschaft (DFG) for a Mercator Fellowship. The second author is supported by an NSERC Discovery grant. The third author was supported in part by NSF-DMS #1352386 and NSF-DMS #1812028. She has also received funding from the European Union's Horizon 2020 research and innovation program under Marie Skłodowska-Curie grant agreement No. 754462.

Conflicts of interest: None.

References

- [1] N. Alon and J. H. Spencer, *The Probabilistic Method*, fourth edn, Wiley Series in Discrete Mathematics and Optimization (John Wiley & Sons, Inc., Hoboken, NJ, 2016).
- [2] L. Aronshtam, N. Linial, T. Łuczak and R. Meshulam, 'Collapsibility and vanishing of top homology in random simplicial complexes', *Discrete Comput. Geom.* **49**(2) (2013), 317–334.
- [3] E. K. Babson, L. J. Billera and C. S. Chan, 'Neighborly cubical spheres and a cubical lower bound conjecture', *Israel J. Math.* **102** (1997), 297–315.
- [4] E. Babson, C. Hoffman and M. Kahle, 'The fundamental group of random 2-complexes', *J. Amer. Math. Soc.* **24**(1) (2011), 1–28.
- [5] B. Bollobás, *Random Graphs*, second edn, Cambridge Studies in Advanced Mathematics (Cambridge University Press, Cambridge, UK, 2001).
- [6] J. D. Burtin, 'The probability of connectedness of a random subgraph of an n -dimensional cube', *Problemy Peredachi Informatsii* **13**(2) (1977), 90–95 (in Russian).
- [7] D. Cohen, A. Costa, M. Farber and T. Kappeler, 'Topology of random 2-complexes', *Discrete Comput. Geom.* **47**(1) (2012), 117–149.
- [8] A. Costa and M. Farber, 'Large random simplicial complexes, I', *J. Topol. Anal.* **8**(3) (2016), 399–429.
- [9] A. Costa and M. Farber, 'Large random simplicial complexes, II: The fundamental group', *J. Topol. Anal.* **9**(3) (2017), 441–483.
- [10] A. Costa and M. Farber, 'Large random simplicial complexes, III: The critical dimension', *J. Knot Theory Ramifications* **26**(2) (2017), 1740010.
- [11] M. W. Davis, *The Geometry and Topology of Coxeter Groups*, London Mathematical Society Monographs Series vol. **32** (Princeton University Press, Princeton, NJ, 2008).
- [12] K. Dowling and E. Lundberg, 'Homotopy types of random cubical complexes', Preprint, 2019, [arXiv:1910.12803](https://arxiv.org/abs/1910.12803).
- [13] P. Erdős and A. Rényi, 'On the evolution of random graphs', *Magyar Tud. Akad. Mat. Kutató Int. Közl.* **5** (1960), 17–61.
- [14] P. Erdős and J. Spencer, 'Evolution of the n -cube', *Comput. Math. Appl.* **5**(1) (1979), 33–39.
- [15] M. Farber and T. Nowik, 'Topological embeddings into random 2-complexes', Preprint, 2019, [arXiv:1912.03939](https://arxiv.org/abs/1912.03939).
- [16] S. Hart, 'A note on the edges of the n -cube', *Discrete Math.* **14**(2) (1976), 157–163.
- [17] Y. Hiraoka and T. Shirai, 'Tutte polynomials and random-cluster models in Bernoulli cell complexes', Preprint, 2016, [arXiv:1602.04561](https://arxiv.org/abs/1602.04561).

- [18] Y. Hiraoka and K. Tsunoda, ‘Limit theorems for random cubical homology’, *Discrete Comput. Geom.* **60**(3) (2018), 665–687.
- [19] C. Hoffman, M. Kahle and E. Paquette, ‘Spectral gaps of random graphs and applications’, *Int. Math. Res. Not. IMRN* **11** (2021), 8353–8404.
- [20] M. Kahle, F. H. Lutz, A. Newman and K. Parsons, ‘Cohen–Lenstra heuristics for torsion in homology of random complexes’, *Experimental Mathematics* **29**(3) (2020), 347–359.
- [21] A. V. Kostochka, A. A. Sapozhenko and K. Weber, ‘On random cubical graphs’, in *Fourth Czechoslovakian Symposium on Combinatorics, Graphs and Complexity (Prachatice, 1990)*, Annals of Discrete Mathematics vol. **51** (North-Holland, Amsterdam, 1992), 155–160.
- [22] N. Linial and R. Meshulam, ‘Homological connectivity of random 2-complexes’, *Combinatorica* **26**(4) (2006), 475–487.
- [23] N. Linial and Y. Peled, ‘On the phase transition in random simplicial complexes’, *Ann. of Math. (2)* **184**(3) (2016), 745–773.
- [24] T. Łuczak and Y. Peled, ‘Integral homology of random simplicial complexes’, *Discrete Comput. Geom.* **59**(1) (2018), 131–142.
- [25] Z. Luria and Y. Peled, ‘On simple connectivity of random 2-complexes’, Preprint, 2018, [arXiv:1806.03351](https://arxiv.org/abs/1806.03351).
- [26] R. Meshulam and N. Wallach, ‘Homological connectivity of random k -dimensional complexes’, *Random Structures Algorithms* **34**(3) (2009), 408–417.
- [27] A. Newman, ‘Freeness of the random fundamental group’, *J. Topol. Anal.* **12**(01) (2020), 29–35.
- [28] A. Newman and E. Paquette, ‘The integer homology threshold in $Y_d(n, p)$ ’, Preprint, 2018, [arXiv:1808.10647](https://arxiv.org/abs/1808.10647).