

A power saving for planar unit distances

OpenAI

September 23, 2026

Abstract

We prove that, for some absolute $\beta < 4/3$, every set of n points in the Euclidean plane determines $O(n^\beta)$ unordered pairs at unit distance. If configurations approached exponent $4/3$, we would obtain a dense bipartite graph whose edges carry ratios of coordinate differences along unit-distance edges. Each ratio has large logarithmic height, but multiplying the ratios on two opposite edges of a four-cycle and dividing by the other two gives negligible height on the same scale. Prediction from earlier test responses and the product formula give these height bounds. Passing to an ultraproduct turns the negligible-height quotients into constants for a derivation. The differentiated distance equations involve square roots whose radicands lie in a function field. We prove that no nonempty product of these radicands is a square there. Kummer theory then lets us change one root's sign while fixing the others, contradicting those equations. The saving is fixed but not explicit.

1 Introduction

For a finite set $X \subset \mathbb{R}^2$, write

$$u(X) = \#\{\{x, y\} \subset X : \|x - y\|_2 = 1\}, \quad u(n) = \max_{\substack{X \subset \mathbb{R}^2 \\ |X|=n}} u(X).$$

Thus pairs are unordered and their endpoints are distinct. We prove a fixed power saving over the exponent $4/3$.

Theorem 1.1. *There are absolute constants $0 < C < \infty$ and $1 \leq \beta < 4/3$ such that $u(n) \leq Cn^\beta$ for every nonnegative integer n .*

Equivalently, $u(n) = O(n^{4/3-\delta})$ for an absolute $\delta > 0$, uniformly over all configurations.

1.1 Earlier bounds and structure

Erdős introduced the repeated-distance problem in 1946 [8]. Rescaling turns any specified positive distance into a unit distance, so the problem concerns the largest possible multiplicity of one distance. His lattice construction gives $u(n) \geq n^{1+c/\log \log n}$ for a positive constant c and sufficiently large n [8, Theorem 2]. Spencer, Szemerédi and Trotter established $u(n) = O(n^{4/3})$ [21]. Székely's crossing-number proof places this bound within a general point–curve incidence theorem [22, Theorems 4 and 8]: two unit circles meet in at most two points, and two distinct points belong to at most two unit circles. The crossing inequality is due independently to Ajtai, Chvátal, Newborn and Szemerédi and to Leighton [3, 13]. Later work improved the leading constant [1].

Katz and Silier describe near-extremal configurations for both point–line and point–unit-circle incidences. For N points and N unit circles, they obtain a cell decomposition from two families

of circles through two points, and a construction using $O(N^{1/3})$ parameters that captures a near-extremal subconfiguration [10, Theorems 1.4, 5.26 and 5.28]. This describes the structure possible at the $4/3$ scale; the remaining question is whether unit-distance configurations can attain that scale. Pach, Raz and Solymosi obtain a logarithmic improvement of the upper bound conditional on their rigidity conjecture [18, Conjecture 7 and Theorem 8]. Our proof rules out sequences with $n^{4/3-o(1)}$ unit distances by combining incidence density with arithmetic constraints on the unit-distance equations.

The historical conjecture of Erdős proposed the stronger upper bound $u(n) = n^{1+o(1)}$. An OpenAI construction disproved it by giving $n^{1+\varepsilon}$ unit distances, for a fixed $\varepsilon > 0$, along an unbounded sequence of sizes [17, Theorem 1.1]. Alon et al. gave a simplified account and explained its antecedents in class-group arguments and number-field towers with prescribed splitting [4, Sections 1.2–1.3]. Sawin obtained $u(n) \gg n^{1.014114}$ along an unbounded sequence [19, Theorem 1]. These lower bounds and Theorem 1.1 leave a gap between the known exponents.

1.2 The proof idea

Suppose there were a sequence with $t^{4-o(1)}$ ordered unit pairs, where $t = n^{1/3}$. We replace the coordinates by real algebraic numbers while preserving the unit pairs, then use sampling to retain a graph with fewer points than circle centers and controlled degrees (Section 2). Its edges cannot concentrate too heavily between small subsets of the two vertex sets. Choose a uniform edge and, conditional on its center, independently choose a second neighboring point. The unequal vertex-set sizes and the bound of two common centers make the mutual information between these two points $o(\log t)$.

Tardos and Katz–Tardos use entropy inequalities to lower-bound the number of distinct distances. Katz–Tardos also resamples within classes while preserving the marginal, as we do when coupling points with the same earlier test answers [23, 11]. Ahlswede reduces dependence by conditioning on selected coordinates [2, Section 4, Lemma 3]. Here we must predict a point’s response using only its answers to earlier tests, while evaluating the prediction with a new test drawn independently of those answers and the sampled points. The acceptance criterion may depend on the neighboring center. We subtract from each endpoint’s test feature its mean conditional on that endpoint’s earlier answers. Since the two endpoints are independent given their common center, the expected inner product of these residuals is the expected squared norm of their conditional mean at the center. The information bound controls this quantity and hence the prediction error (Section 3).

Write $z = x_1 + ix_2$ and $w = x_1 - ix_2$ for $x = (x_1, x_2)$. Along a unit edge, the coordinate differences are d and d^{-1} . At an absolute value v and level s , test whether $|d|_v < e^{-s}$ and group nearby coordinate values under a common label. When one outcome is rare, most vertices share a label in the coordinate selected by the more frequent outcome. We compare pairs of points sharing a center with pairs of centers sharing a point, conditioning the latter sampling on earlier test answers. The product formula [14, Theorem 7.15] cancels the weighted logarithmic terms in the upper bounds for both samples’ scores. The contributions from vertices outside the prevalent label class have opposite signs in their lower bounds and cancel when added. Together with prediction, this shows that thresholds where both outcomes are frequent contribute negligible disagreement of the sampled edge bits.

A positive amount of variation remains on the other thresholds, and Section 4 shows that $S \rightarrow \infty$, where S is the weighted integral of the less frequent bit’s probability, defined in Section 4.2. Integrating each point’s signed exception indicators over thresholds gives a profile

indexed by absolute values. For many pairs p, r with a chosen common center q , the logarithms of $\lambda_{pr} = (z_p - z_q)/(z_r - z_q)$ are approximately the profile at r minus the profile at p . Each endpoint profile therefore cancels in $\lambda_{pr}\lambda_{p'r'}/(\lambda_{pr'}\lambda_{p'r})$ whenever all four pairs are retained. This is a rectangle in the bipartite graph, with no geometric rectangle required. Absolute logarithmic height is the weighted sum of positive logarithms of absolute values. [Proposition 7.1](#) gives a scale $J \rightarrow \infty$ for which every retained ratio has height at least $J/3$ and every such quotient has height $o(J)$ uniformly.

To use these height bounds, we pass to an ultraproduct and choose a derivation that vanishes on the field of negligible-height elements. Graph density supplies left and right points with every cross-pair retained. One left point remains free of extra polynomial relations on its plane or curve after the coefficient field is enlarged to contain the right-hand positions, their coordinate derivatives, and the scalars obtained by differentiating the ratios. We work in the function field of that plane or curve over this enlarged coefficient field.

The differentiated distance equations involve square roots of radicands in this function field. For each radicand, we construct a zero of odd order at which all the other radicands have order zero. Thus no nonempty product is a square. The common-unit-circle case would make every radicand a square; this is why [Section 7](#) first removes pairs on unit circles containing many points. Kummer theory now permits one root's sign to change while fixing the other roots and the function field [[15](#), Theorem 5.30 and Remark 5.32]. This change contradicts the differentiated distance equations together with the relations obtained from the four-edge quotients ([Section 8](#)).

The contradiction excludes every sequence approaching exponent $4/3$, and hence gives a fixed positive saving. The proof does not estimate its size. The deduction of the main theorem from the precise reductions appears in [Section 2](#).

Throughout the proof, asymptotic notation refers to the chosen sequence $t \rightarrow \infty$. Constants denoted by c, C may change between occurrences. All logarithms are natural. Every appeal to a slowly varying parameter will be made after the quantities it must dominate have been specified.

2 An incidence piece and its probability laws

The input to this section is a sequence of planar configurations with $t = n^{1/3} \rightarrow \infty$ and at least $t^{4-o(1)}$ ordered unit pairs. We construct an unbalanced incidence graph whose degrees are controlled on both sides. For a uniformly chosen edge, we bound the probability that its endpoints belong to prescribed subsets of the two vertex sets; the bound is uniform even for arbitrarily small subsets. We then prove two consequences of that estimate and show that the endpoints of a random two-step walk have mutual information $o(\log t)$.

Such an input sequence exists if [Theorem 1.1](#) fails. Indeed, for each $j \geq 2$, failure of a bound with exponent $4/3 - 1/j$ allows the constant in that bound to be chosen large enough to force a counterexample with $n \geq j$. Its ordered unit-pair count is at least $n^{4/3-1/j} = t^{4-3/j}$. Passing to an increasing sequence of sizes gives the claimed input. We use separate copies of the configuration as points and as centers of unit circles, so its ordered unit pairs are exactly the incidences between these two copies.

2.1 The reduction and the deduction of the main theorem

The incidence output needed throughout the proof is the following. Its construction is given in [Section 2.4](#); the intervening specialization and sampling arguments supply its two geometric inputs.

Proposition 2.1. *Every sequence of $n = t^3$ real planar points with at least $t^{4-o(1)}$ ordered unit pairs admits, after algebraic specialization and restriction, finite real algebraic point and center sets P, Q and a nonempty unit-incidence graph $E \subseteq P \times Q$ with the following properties. For a parameter $a = a(t) > 0$ with $a \rightarrow 0$ and $a \log t \rightarrow \infty$,*

$$|P| = t^{1+2a+o(a)}, \quad |Q| = t^{2+a+o(a)}, \quad |E| = t^{2+2a+o(a)}. \quad (2.1)$$

Writing $d(p), d(q)$ for graph degrees, uniformly over all vertices,

$$d(p) = t^{1+o(a)}, \quad d(q) = t^{a+o(a)}. \quad (2.2)$$

The uniform edge law $\mathbb{P}_E$, with expectation $\mathbb{E}_E$, has marginals $\mu(p) = d(p)/|E|$ and $\kappa(q) = d(q)/|E|$ satisfying

$$\mu(p) \geq \frac{\gamma}{|P|}, \quad \kappa(q) \geq \frac{\gamma}{|Q|}, \quad \gamma = \frac{27}{50} = 0.54. \quad (2.3)$$

Set $\eta = \gamma - \frac{1}{2} = 1/25$. There is a single sequence $\xi_t \rightarrow 0$ such that, for every $F \subseteq P$ and $G \subseteq Q$, with counting fractions $u = |F|/|P|$ and $v = |G|/|Q|$,

$$\mathbb{P}_E(p \in F, q \in G) \leq \sqrt{uv}((uv)^\eta + \xi_t). \quad (2.4)$$

All these estimates hold along the given sequence; the same ξ_t works for every pair of subsets.

Proof of Theorem 1.1 from the reductions. If the theorem fails, the sequence reduction above gives configurations with $t^{4-o(1)}$ ordered unit pairs. Apply Proposition 2.1 to obtain the algebraic incidence graphs. The height construction in Section 4 applies to every such sequence: Proposition 4.3 excludes bounded variation, and Lemma 4.5 supplies the coordinate tests. With the parameters in Section 5.1, the score comparisons in Propositions 6.4 and 6.5 give the two inputs to Proposition 7.1.

That proposition produces at least $t^{2-o(1)}$ pairs on two copies of an algebraic point set of size at most $t^{1+o(1)}$, together with the exact identity (7.1), the lower height bound (7.2), the uniform rectangle bound (7.3), and the stated exclusion of pairs on unit circles containing at least $t^{9/10}$ points. These are precisely the hypotheses of Proposition 8.1, with $c = 1/3$; its conclusion is a contradiction. The statements of the extraction and obstruction are at the entrances to Sections 7 and 8, and their deductions are in Sections 7.1 and 8.5.

Hence some fixed exponent below $4/3$ admits a uniform bound. Increasing the exponent to at least one preserves this bound and keeps it below $4/3$. Enlarging the positive constant covers the finitely many remaining sizes $n \geq 2$, and passing from ordered to unordered pairs only changes the constant. The cases $n = 0, 1$ are immediate. $\square$

2.2 Preserving incidences under algebraic specialization

Later arguments use number-field absolute values. The following specialization permits this without losing any of the unit pairs.

Lemma 2.2. *Let $x_1, \dots, x_n$ be distinct points of $\mathbb{R}^2$, and let $\mathcal{E}$ be any collection of pairs (i, j) for which $\|x_i - x_j\|_2 = 1$. There are distinct points $x'_1, \dots, x'_n$ with real algebraic coordinates, arbitrarily close to the respective original points, such that $\|x'_i - x'_j\|_2 = 1$ for every $(i, j) \in \mathcal{E}$.*

Proof. Fix a neighborhood of each original point and choose a rational open rectangle containing that point inside its neighborhood. For variables $X_i = (X_{i,1}, X_{i,2})$, impose membership in these rectangles, the equations

$$(X_{i,1} - X_{j,1})^2 + (X_{i,2} - X_{j,2})^2 = 1 \quad ((i, j) \in \mathcal{E}),$$

and the strict inequalities

$$(X_{i,1} - X_{j,1})^2 + (X_{i,2} - X_{j,2})^2 > 0 \quad (i \neq j).$$

This finite system has rational coefficients and a real solution, namely the original configuration. Apply the transfer principle [12, Theorems 3.1 and 4.1] with ordered base field $\mathbb{Q}$ and its real closed extensions $\mathbb{R}$ and $\overline{\mathbb{Q}} \cap \mathbb{R}$. Every rectangle endpoint is rational, so the equations, strict inequalities, and neighborhood restrictions belong to the same system over this common base. Its new solution therefore lies in the field of real algebraic numbers. The equations preserve the required unit pairs, the inequalities preserve distinctness, and the rectangles give the prescribed closeness. $\square$

We henceforth apply this lemma separately to each configuration. Extra unit pairs created by the specialization are harmless; all subsequent graphs may retain only the incidences under consideration.

2.3 Regions from a random sample and the incidence bound

For finite sets $\mathcal{P}$ of points and $\mathcal{Q}$ of distinct unit-circle centers, let $I(\mathcal{P}, \mathcal{Q})$ denote the number of incidences. We identify each center with its unit circle when speaking of a family of circles. Two distinct unit circles meet in at most two points. Also, two distinct points lie on at most two unit circles: their possible centers are intersections of the two unit circles centered at the points. These two facts provide the pair-counting bounds used below. Clarkson bounds the number of input objects meeting a region defined by a bounded number of sampled objects [5, Section 4]. Here this sampling argument partitions the points into regions that each meet few of the original circles. We give the circle boundary conventions explicitly so that every retained point belongs to its assigned open region.

Lemma 2.3. *Let $\mathcal{P}$ be a finite planar point set and let $\mathcal{Q}$ be a nonempty set of l distinct unit circles. For every $R \geq 1$, sample $s = \lceil CR \log(2 + l) \rceil$ circles independently and uniformly with replacement, where C is a sufficiently large absolute constant. With probability at least $9/10$, after deleting the points on sampled circles, all remaining points can be assigned to at most $C'(1 + s^2)$ open regions, where C' is absolute. Each region intersects at most l/R original circles and contains every point assigned to it.*

Proof. Choose an auxiliary horizontal direction before taking the sample. Require that an input point and an intersection point of two original circles have different horizontal coordinates unless they are the same point. Require also that an input point have a horizontal coordinate different from that of a horizontal extremum of an original circle, unless it is that extremum. Only finitely many directions are excluded. For the latter assertion, if c is a circle center and p is a point, an equality with an extremal abscissa has the form $e \cdot (p - c) = \pm 1$ for a unit horizontal direction e , and has at most finitely many solutions in e . This auxiliary direction does not change the coordinates used elsewhere in the proof.

Ignore repetitions among the sampled circles. Express each sampled circle as its upper and lower semicircle graphs on their open horizontal domains. The event points are their horizontal extrema and all their pairwise intersections, including tangencies. Between consecutive event abscissas, the graphs have fixed vertical order; consecutive graphs, with infinite outer boundaries allowed, bound open gaps.

Sweep these gaps from left to right. Continue a gap across an event abscissa precisely when its two boundary graphs and their adjacency continue unchanged and its closed limiting vertical

segment contains no event point. Otherwise start a new piece on the right. Each resulting piece is the open region between two fixed semicircle graphs, or infinite boundaries, over an open interval. It contains no part of a sampled circle. In particular, a continued gap has no start, end, tangency or crossing inside it at the abscissa across which it continues. An input point at an event abscissa must, by the choice of direction, itself be an event point and therefore lies on a sampled circle. Thus every point not deleted is in one of the open pieces.

There are $O(1 + s^2)$ pieces, including in the presence of multiple intersections or several events with the same abscissa. A new right-hand gap can start only if an event point lies in its closed limiting vertical segment. If c_x distinct sampled circles pass through an event point x , only $O(1 + c_x)$ right-hand gaps can have this property, by their vertical ordering. Summing over event points costs $O(s + s^2)$: extrema contribute $O(s)$, and at an intersection with $c_x \geq 2$ one may charge c_x to $2\binom{c_x}{2}$, while each pair of circles has at most two intersections. This also covers extrema that coincide with intersections. There is only a constant number of initial unbounded gaps.

Finally, all possible pieces have a description from the full circle set using a bounded number of circles. Each of their two horizontal endpoints is an extremal or intersection abscissa, or infinity; each boundary graph is specified by a circle and a sign, or is infinite. At most two circles specify the boundary graphs, and at most two specify each finite endpoint abscissa. These six choices, together with finitely many choices of signs and event types, determine the region. Consequently at most $C_0(1 + l)^6$ candidate regions can occur, with an absolute constant C_0 . For any fixed candidate region intersecting more than l/R circles, the probability that no sampled circle intersects it is at most

$$(1 - 1/R)^s \leq e^{-s/R},$$

with the evident interpretation when $R = 1$. An actual open piece avoids every sampled circle. A union bound over the candidate regions, with C sufficiently large, therefore shows, with failure probability at most $1/10$, that no resulting region intersects more than l/R original circles. $\square$

For m distinct points and l distinct unit circles in the plane, let I be their number of incidences. Then

$$I \leq C \log^2(2 + m + l)(m^{2/3}l^{2/3} + m + l). \quad (2.5)$$

Here C is absolute, and the bound includes $m = 0$ or $l = 0$.

Székely's point-curve incidence theorem [22, Theorem 8] gives the stronger bound $I = O(m^{2/3}l^{2/3} + m + l)$ when distinct curves meet in at most two points and distinct points belong to at most two curves. Distinct unit circles satisfy both hypotheses. If simple curves are taken to be arcs, cut each circle at two points outside the finite input point set; the resulting $2l$ arcs satisfy the same pair bounds and preserve every incidence. The logarithmically weakened form (2.5) follows, including the empty-set cases.

When m, l are bounded by a fixed power of t , the logarithmic factor in (2.5) is $(\log t)^{O(1)}$.

2.4 Degree control by selecting vertex subsets

We now use the regions from Lemma 2.3 to obtain the prescribed imbalance, and then choose a subset of points and a subset of centers, retaining all edges between them. We maximize the retained edge fraction divided by a fixed power of the product of the two vertex fractions. Maximality bounds how many edges any further restriction can capture and also gives lower degree bounds.

Proof of Proposition 2.1. First use Lemma 2.2. Choose $a \rightarrow 0$ sufficiently slowly that the initial incidence count is at least $t^{4-o(a)}$ and every fixed power of $\log t$ is $t^{o(a)}$. In particular $a \log t \rightarrow \infty$. Apply Lemma 2.3 with $R = t^{1-a}$ and $s = O(t^{1-a} \log t)$. There are t^3 points and circles. Incidences on the sampled circles number at most st^3 . A deleted point on any other circle lies at an intersection of that circle with a sampled circle. Charging its incidence to one such pair of circles costs at most two incidences per pair, and hence at most $2st^3$ in total. Thus the deletion costs $O(t^{4-a} \log t) = o(t^{4-o(a)})$.

Each remaining region meets at most t^{2+a} circles. Further split its assigned point set into sets of size at most $2t^{1+2a}$. The total number of resulting pieces is at most the original region count plus $O(t^3/t^{1+2a})$, hence at most $t^{2-2a+o(a)}$. Some piece has at least $t^{2+2a-o(a)}$ incidences. Denote its two vertex sets temporarily by P_0, Q_0 , and its edge count by e_0 . At this stage

$$|P_0| \leq 2t^{1+2a}, \quad |Q_0| \leq t^{2+a}, \quad t^{2+2a-o(a)} \leq e_0 \leq t^{2+2a+o(a)},$$

where the upper bound follows from (2.5).

Choose $b = o(a)$ so slowly that $b \log t \rightarrow \infty$ and b dominates all exponent errors in these estimates and the logarithmic factor in (2.5). Delete point vertices of degree greater than t^{1+b} , and center vertices of degree greater than t^{a+b} , with degrees measured before these deletions. The counts of such vertices are at most

$$t^{1+2a-b/2} \quad \text{and} \quad t^{2+a-b/2},$$

respectively, for all sufficiently large t . Apply (2.5) to the excessive vertices on one side and all vertices on the other. In either case the mixed term is at most $t^{2+2a-b/3+o(b)}$; the two linear terms are bounded by $t^{1+2a+o(b)} + t^{2+a+o(b)}$. Their sum is $o(e_0)$. Both deletions therefore retain $t^{2+2a+o(a)}$ edges and give the degree upper bounds $t^{1+o(a)}$, $t^{a+o(a)}$. Dividing the retained edge count by these two upper bounds shows that both retained vertex counts attain their stated upper scales within $t^{o(a)}$. Relabel this trimmed graph as (P_0, Q_0, E_0) .

For subsets of these trimmed vertex sets, let u, v be their respective counting fractions and let w be their captured edge fraction. The incidence bound and the maximum-degree bounds give, with one uniform factor $T_t = t^{o(a)}$,

$$\begin{aligned} w &\leq T_t((uv)^{2/3} + t^{-a}v + t^{-1}u), \\ w &\leq T_t \min(u, v). \end{aligned} \tag{2.6}$$

To use both bounds simultaneously, write $A = (uv)^{2/3}$, $B = t^{-a}v$, $C = t^{-1}u$, and $D_0 = \min(u, v)$. For nonnegative quantities,

$$\min(A + B + C, D_0) \leq A + \min(B, D_0) + \min(C, D_0) \leq A + \min(B, u) + \min(C, v).$$

Using $\min(x, y) \leq \sqrt{xy}$ on the last two terms yields

$$w \leq T_t \sqrt{uv}((uv)^{1/6} + 2t^{-a/2}), \tag{2.7}$$

after increasing T_t if necessary. For example, $\min(t^{-a}v, u) \leq t^{-a/2} \sqrt{uv}$; the other linear term uses $t^{-1/2} \leq t^{-a/2}$ for large t .

Put $\tau = t^{-a/8}$. Among all products of vertex subsets, with the edges they induce in the trimmed graph and with counting fractions $u, v \geq \tau$ and captured edge fraction w , maximize

$$\frac{w}{(uv)^\gamma}, \quad \gamma = \frac{27}{50}.$$

We choose $\gamma > 1/2$ so that a bound by $(uv)^\gamma$ improves on $\sqrt{uv}$ for small sets. The inequality $\gamma < 2/3$, where $2/3$ is the incidence exponent in (2.6), will ensure that the maximizing rectangle retains the scales (2.1). This is a maximum over a nonempty finite family, and its value is at least one because the full rectangle is feasible. Write U, V for the maximizing counting fractions and W for its captured edge fraction. We claim

$$U, V, W \geq t^{-o(a)}. \quad (2.8)$$

Indeed (2.7) gives

$$1 \leq T_t \left((UV)^{2/3-\gamma} + 2t^{-a/2}(UV)^{-\eta} \right).$$

By feasibility, $UV \geq t^{-a/4}$, so the second term on the right is at most $2t^{-49a/100+o(a)}$, which tends to zero. If $UV \leq t^{-ca}$ along a subsequence for any fixed $c > 0$, the first term tends to zero as well, since $2/3 - \gamma = 19/150 > 0$. This is impossible. Thus $UV \geq t^{-o(a)}$, which implies the same for each of U, V , and $W \geq (UV)^\gamma$ gives the claim for W .

Let (P, Q, E) be this maximizing rectangle, retaining precisely its edges from the trimmed graph. The scale estimates (2.1) and the degree upper bounds survive by (2.8). For subsets $F \subseteq P$, $G \subseteq Q$, write $\mathbb{P}_E(F, G) = \mathbb{P}_E(p \in F, q \in G)$. For further subsets with fractions u, v inside P, Q , optimality gives

$$\mathbb{P}_E(F, G) \leq (uv)^\gamma$$

whenever $Uu, Vv \geq \tau$. Otherwise $uv \leq \tau / \min(U, V) = t^{-a/8+o(a)}$. Apply (2.7) with the old fractions Uu, Vv and divide by W . Uniformly in all such infeasible subsets, this gives

$$\mathbb{P}_E(F, G) \leq t^{o(a)} \sqrt{uv} ((uv)^{1/6} + 2t^{-a/2}) \leq \xi_t \sqrt{uv},$$

where one may take $\xi_t = t^{-a/48+o(a)} \rightarrow 0$, with the exponent error depending only on the chosen rectangle and the previous uniform bounds. Combining the feasible and infeasible cases proves (2.4). Empty subsets satisfy it trivially.

Finally (2.8) implies that removing any single vertex of the maximizing rectangle remains feasible for all sufficiently large t . Optimality with a point p removed gives

$$1 - \mu(p) \leq (1 - |P|^{-1})^\gamma \leq 1 - \gamma |P|^{-1},$$

the last inequality following from concavity. The center comparison is identical. This proves (2.3), and multiplying by $|E|$ supplies the lower degree bounds in (2.2). $\square$

In particular, the minimum degree on either side tends to infinity, and $\max_p \mu(p) = t^{-1+o(1)}$. The lower marginal bounds also convert (2.4) to the form needed for probability laws. For $h = \mu(F)$, $k = \mu(G)$, we have $|F|/|P| \leq h/\gamma$, $|G|/|Q| \leq k/\gamma$, and hence

$$\mathbb{P}_E(F, G) \leq C \sqrt{hk} ((hk)^\eta + \xi_t), \quad (2.9)$$

with a fixed absolute C . Here $\mathbb{P}_E(F, G)$ abbreviates $\mathbb{P}_E(p \in F, q \in G)$. The same ξ_t works for all subsets.

2.5 Matching labels and uniformly rare sets

We will repeatedly assign finite labels to the vertices, with no bound on the number of labels. If the endpoint labels agree on almost every edge, the subset bound (2.4) forces one label to occur with probability close to one under both vertex marginals. Applying the same bound to the two complementary label classes makes their total probability proportional to the edge-label disagreement probability, even when that probability is smaller than ξ_t .

Lemma 2.4. *For the sequence of edge laws in Proposition 2.1, there are absolute constants $s_0 > 0, C < \infty$ and an index after which the following holds uniformly for every pair of label maps $\ell_P : P \rightarrow \mathcal{A}, \ell_Q : Q \rightarrow \mathcal{A}$ into a finite set. If*

$$s = \mathbb{P}_E(\ell_P(p) \neq \ell_Q(q)) \leq s_0,$$

some label $a_0 \in \mathcal{A}$ satisfies

$$\mu\{\ell_P \neq a_0\} + \kappa\{\ell_Q \neq a_0\} \leq Cs.$$

Moreover, for independent $p, p' \sim \mu$, at every value of s ,

$$\mathbb{P}(\ell_P(p) \neq \ell_P(p')) \leq Cs. \quad (2.10)$$

The analogous statement holds for independent centers of law κ .

Proof. Let u_i, v_i be the counting fractions of the two classes with label i , including zero fractions. Sum (2.4) over the matched rectangles. Cauchy–Schwarz gives $\sum_i \sqrt{u_i v_i} \leq 1$, and therefore

$$1 - s \leq \max_i (u_i v_i)^\eta + \xi_t.$$

For sufficiently small fixed s_0 and sufficiently large t , one label a_0 thus has both counting fractions arbitrarily close to one. Apply (2.4) to either complementary class against the whole opposite side. The marginal exception masses

$$h = \mu\{\ell_P \neq a_0\}, \quad k = \kappa\{\ell_Q \neq a_0\}$$

are then as small as any prescribed fixed constant, uniformly over the label maps. This first conclusion need not be proportional to s .

To obtain proportionality, put $J = \mathbb{P}_E(\ell_P \neq a_0, \ell_Q \neq a_0)$. By (2.9) and $\sqrt{hk} \leq (h+k)/2$, we may choose the preceding constants so that $J \leq (h+k)/4$. The events on which exactly one endpoint has label a_0 are mismatches, so

$$s \geq h + k - 2J \geq \frac{1}{2}(h + k).$$

This proves the first assertion, including when $s = 0$, with no additive error. Two independent point labels differ with probability at most $2h$, giving (2.10) for small s . For $s > s_0$, the bound follows by increasing C so that $Cs_0 \geq 1$. The center assertion is identical. $\square$

Consider the following law on two-step walks, allowing $p = r$. On triples $(p, q, r) \in P \times Q \times P$ we first choose $q \sim \kappa$ and then independently choose p, r according to the conditional edge law given q . Explicitly,

$$\mathbb{P}(p, q, r) = \begin{cases} (|E|d(q))^{-1}, & (p, q), (r, q) \in E, \\ 0, & \text{otherwise.} \end{cases} \quad (2.11)$$

Each point marginal is μ , and each of the two point–center marginals is the uniform edge law.

Lemma 2.5. *Let $\delta_t \rightarrow 0$ be any nonnegative sequence. The estimates below hold uniformly over all indicated subsets, with error factors depending only on δ_t, ξ_t and the fixed constants in (2.9):*

(i) *If $F \subseteq P, G \subseteq Q$, and $h = \mu(F) \leq \delta_t$, while $k = \kappa(G)$ is arbitrary, then*

$$\mathbb{P}_E(F, G) \leq C(\delta_t^\eta + \xi_t)\sqrt{hk} = o(1)\sqrt{hk}. \quad (2.12)$$

The same assertion holds with the sides interchanged.

(ii) If also $k \leq \delta_t$, then

$$\mathbb{P}_E(F, G) \leq \frac{C}{2}(\delta_t^{2\eta} + \xi_t)(h + k) = o(1)(h + k). \quad (2.13)$$

(iii) If $F \subseteq P$ and $h = \mu(F) \leq \delta_t$, then under the two-step walk law,

$$\mathbb{P}(p \in F, r \in F) = o(1)h. \quad (2.14)$$

In particular, the error factors may be taken outside sums or integrals over any family of sets whose marginal masses obey the same uniform bound δ_t .

Proof. The first two assertions follow directly from (2.9), using $hk \leq \delta_t$ in the first case and $hk \leq \delta_t^2$ and $\sqrt{hk} \leq (h + k)/2$ in the second.

For the third, let $g(q) = \mathbb{P}(p \in F \mid q)$, so that $\mathbb{E}_\kappa g = h$ and the probability in question is $\mathbb{E}_\kappa g^2$. There is nothing to prove when $h = 0$. For any $d \in (0, 1]$, put $G_d = \{q : g(q) > d\}$. Markov's inequality gives $\kappa(G_d) \leq h/d$. Splitting at this threshold,

$$\mathbb{E}_\kappa g^2 \leq dh + \mathbb{E}_\kappa [g \mathbf{1}_{G_d}] = dh + \mathbb{P}_E(F, G_d) \leq h \left[d + Cd^{-1/2}((\delta_t^2/d)^\eta + \xi_t) \right].$$

Choose $d = d_t \rightarrow 0$ sufficiently slowly that $\delta_t^{2\eta} d_t^{-1/2-\eta} \rightarrow 0$ and $\xi_t d_t^{-1/2} \rightarrow 0$. The bracket then tends to zero, independently of F . This proves the uniform assertion and hence its integrated form. $\square$

2.6 The information carried by a two-step walk

The small point side and the bound of two common centers for distinct points together ensure that the two point endpoints of a two-step walk share little information. We record the atom estimates as well, since they later convert a probability of two-step walks into a count of distinct point pairs.

For probability laws a, b on the same finite set, define their relative entropy by

$$\mathrm{D}_{\mathrm{KL}}(a \| b) = \sum_x a(x) \log \frac{a(x)}{b(x)}, \quad (2.15)$$

with a zero summand when $a(x) = 0$ and value $+\infty$ if $a(x) > 0 = b(x)$.

Write $\mathcal{L}(X)$ for the probability law of a random variable X .

Lemma 2.6. *For the two-step walk law (2.11), every distinct pair $p, r \in P$ has probability*

$$\mathbb{P}(p, r) \leq \frac{2}{|E| \min_{q \in Q} d(q)} = t^{-2-3a+o(a)}. \quad (2.16)$$

The diagonal probability is

$$\mathbb{P}(p = r) = \frac{|Q|}{|E|} = t^{-a+o(a)} = o(1). \quad (2.17)$$

The mutual information of the point endpoints satisfies

$$D := \mathrm{D}_{\mathrm{KL}}(\mathcal{L}(p, r) \| \mu \otimes \mu) = o(\log t). \quad (2.18)$$

Proof. For fixed p, r , summing (2.11) gives

$$\mathbb{P}(p, r) = \frac{1}{|E|} \sum_{q: (p, q), (r, q) \in E} \frac{1}{d(q)}.$$

There are at most two summands when $p \neq r$, proving the off-diagonal estimate by (2.1)–(2.2). Summing instead over the diagonal yields

$$\mathbb{P}(p = r) = \frac{1}{|E|} \sum_{q \in Q} \frac{d(q)}{d(q)} = \frac{|Q|}{|E|}.$$

The asymptotics follow from the same estimates and $a \log t \rightarrow \infty$.

For a finite random variable X , write $H(X) = -\sum_x \mathbb{P}(X = x) \log \mathbb{P}(X = x)$, with $0 \log 0 = 0$. Every off-diagonal atom is at most $M_t = 2/(|E| \min_{q \in Q} d(q))$. Dropping the nonnegative diagonal entropy contributions therefore gives

$$H(p, r) \geq (1 - \mathbb{P}(p = r)) \log(1/M_t) \geq (2 - o(1)) \log t.$$

On the other hand, $H(p) = H(r) \leq \log |P| = (1 + o(1)) \log t$. Expanding the definition of relative entropy gives

$$0 \leq D = H(p) + H(r) - H(p, r) = o(\log t),$$

as required. $\square$

We retain this graph and these probability laws for the rest of the proof: the uniform edge law, its point and center marginals, and the two-step walk law. The expansion consequences hold for every finite labeling and every sufficiently small exceptional set, with constants independent of any tests subsequently assigned to the vertices.

3 Prediction from short histories

The incidence reduction supplies a probability law on point–center pairs whose point endpoints in a random two-step walk have mutual information $D = o(\log t)$. A point’s history consists of its answers to earlier tests. Using only those answers, we predict its response to an independently drawn new test, with an evaluation criterion that may depend on the neighboring center. We prove this for an arbitrary finite probability law and then apply it to coordinate bins and to agreement with the center in exactly one of two coordinate labels.

The proof sums the information contributed by successive tests. Since that sum is at most D , a sufficiently long history leaves little new dependence on average. Ahlswede similarly reduces dependence by conditioning on a short list of coordinates [2, Section 4, Lemma 3]. Here the proof must additionally compare each endpoint with its own conditional law and leave the next test independent of the selected history.

3.1 The probability laws and the prediction statement

Let P, Q be finite nonempty sets and let $\omega(p, q)$ be a probability mass function on $P \times Q$, with marginals μ and κ . Delete atoms of zero marginal mass. Write

$$\omega(p | q) = \frac{\omega(p, q)}{\kappa(q)}, \quad \omega(q | p) = \frac{\omega(p, q)}{\mu(p)}.$$

Choose q with law κ , then choose p, r independently given q . The resulting joint law and its point-pair marginal are

$$\Omega(p, q, r) = \kappa(q)\omega(p | q)\omega(r | q), \quad \pi(p, r) = \sum_{q \in Q} \Omega(p, q, r). \quad (3.1)$$

Both marginals of π are μ . Relative entropy is defined in (2.15). For finite probability laws a, b , set

$$\text{TV}(a, b) = \frac{1}{2} \sum_x |a(x) - b(x)|.$$

The information available to the argument is

$$D = \text{D}_{\text{KL}}(\pi \| \mu \otimes \mu). \quad (3.2)$$

It is finite, since $\pi(p, r) > 0$ implies $\mu(p)\mu(r) > 0$. For a map ϕ and a probability measure ν , write $\phi_*\nu$ for the image measure, defined by $(\phi_*\nu)(A) = \nu(\phi^{-1}(A))$.

A test T is a random element of a probability space, independent of (p, q, r) , which specifies a code map $c_T : P \rightarrow [M]$, where $[M] = \{1, \dots, M\}$. The test space may be infinite; all maps and integrands below are assumed measurable. Only finite probability laws of codes are used. All vector norms and inner products in this section are Euclidean. For a fixed history map $H : P \rightarrow \mathcal{H}$, put

$$\mu_h(x) = \frac{\mu(x)\mathbf{1}_{\{H(x)=h\}}}{\mu(H=h)} \quad \text{when } \mu(H=h) > 0. \quad (3.3)$$

Its values on histories of zero mass are immaterial.

Lemma 3.1 (Prediction from a finite history). *Let ω, π, D be as in (3.1)–(3.2), let $K, M \geq 1$ be integers, and let $T_1, \dots, T_K$ be independent copies of T , independent of (p, q, r) . Each test specifies, in addition to c_T , a feature map $f_T : [M] \rightarrow \mathbb{R}^d$ and a coefficient map $g_T : Q \rightarrow \mathbb{R}^d$, for an integer $d \geq 1$ and constants $B, B' \geq 0$, satisfying*

$$\|f_T(c)\| \leq B, \quad \|g_T(q)\| \leq B'$$

for all codes, centers and tests. Choose J uniformly from $[K]$, independently of all tests, and retain only the strict prefix

$$H(p) = (c_{T_1}(p), \dots, c_{T_{J-1}}(p)).$$

Let T be an independent fresh test. For the posterior μ_h in (3.3), define

$$m_{T,h} = \sum_{x \in P} \mu_h(x) f_T(c_T(x)).$$

Then

$$\mathbb{E}_{J, T_1, \dots, T_{J-1}, T} \left| \mathbb{E}_{(p,q) \sim \omega} \langle f_T(c_T(p)) - m_{T,H(p)}, g_T(q) \rangle \right| \leq 4BB' \left(\frac{D}{K} \right)^{1/4}. \quad (3.4)$$

Every realized history has at most M^{K-1} values. In particular, a history can be fixed before drawing the fresh test so that its average absolute discrepancy in (3.4) is at most the right side.

The proof is in Section 3.5. We first draw its two consequences used later: acceptance of coordinate bins at three neighbors, and prediction of equality in exactly one coordinate. Both use the same posterior law (3.3).

3.2 Acceptance tests and the prediction coupling

Corollary 3.2 (Acceptance of predicted codes). *For the finite law ω , test codes and integers K, M in Lemma 3.1, suppose each test specifies a set $\mathcal{A}_T(q) \subseteq [M]$ of accepted codes for every $q \in Q$. Put*

$$s_0 = \mathbb{E}_T \mathbb{P}_{(p,q) \sim \omega} \{c_T(p) \notin \mathcal{A}_T(q)\}.$$

There is a fixed history H , formed from tests preceding a chosen step, with at most M^{K-1} values, such that

$$\mathbb{E}_T \sum_{p,q} \omega(p,q) \sum_x \mu_{H(p)}(x) \mathbf{1}_{\{c_T(x) \notin \mathcal{A}_T(q)\}} \leq s_0 + 4\sqrt{M}(D/K)^{1/4}. \quad (3.5)$$

The average is over a fresh test independent of the history construction.

Proof. Let e_c be the c th standard basis vector of $\mathbb{R}^M$. Use $f_T(c) = e_c \in \mathbb{R}^M$, of norm one, and use the acceptance indicator vector as $g_T(q)$, of norm at most $\sqrt{M}$. Actual and predicted failure probabilities differ by the negative of their acceptance probabilities. Averaging (3.4) and then fixing a strict prefix proves the assertion. $\square$

We record the sampling laws used with this result. For any fixed history H and integer $r \geq 1$, draw one point, one prediction, and r neighbors according to

$$\mathbb{P}\{p, \hat{p}, q_1, \dots, q_r\} = \mu(p) \mu_{H(p)}(\hat{p}) \prod_{j=1}^r \omega(q_j | p), \quad (3.6)$$

and draw the fresh test independently of this entire tuple. The two point marginals are both μ , each (p, q_j) has law ω , and each q_j has marginal κ . In particular each pair $(p, \hat{p})$ is independent of the fresh test, although its two points need not be independent. These assertions follow by summing (3.6), using $\sum_{p: H(p)=h} \mu(p) \mu_h(x) = \mu(x) \mathbf{1}_{\{H(x)=h\}}$. For every fixed history,

$$\mathbb{P}\{p = \hat{p}\} = \sum_{h: \mu(H=h) > 0} \frac{\sum_{p: H(p)=h} \mu(p)^2}{\mu(H=h)} \leq |H(P)| \max_p \mu(p). \quad (3.7)$$

Given $H(p)$, the prediction is independent of all neighbors in (3.6). Applying a union bound to (3.5) therefore bounds the chance that at least one of the r acceptance comparisons rejects this *same* predicted point at the same fresh test by $r(s_0 + 4\sqrt{M}(D/K)^{1/4})$. The union bound needs no independence of the rejection events.

There is also a useful equivalent way to draw the point pair: draw h with probability $\mu(H=h)$, then draw two points independently from μ_h . This again has both marginals μ , and its collision probability satisfies (3.7). Once H is fixed, this pair law can be used unchanged for every fresh test. Sampling a second object from the first object's fiber also appears in the entropy argument of Katz and Tardos [11, Section 2, proof of Lemma 1]; the identities above specify the marginals and independent test needed here.

For the coordinate-quantization application, a test will be an independently chosen field embedding, a code will specify bounded coordinate bins together with an overflow code, and $\mathcal{A}_T(q)$ will consist of codes achieved by actual neighbors of q . The exact hypotheses needed from that construction are a fixed finite alphabet of size M and acceptance of every actual edge, so $s_0 = 0$. If $K = o(\log t)$, $D/K = o(1)$, and $\max_p \mu(p) \leq t^{-1+o(1)}$, then the prediction failure and the collision probability both tend to zero. Formula (3.6) with $r = 3$ is the simultaneous three-neighbor experiment. Geometric estimates valid for all couplings with these marginals remain valid for each fixed history. More generally, additional nonnegative error quantities can be averaged together with the prediction error before choosing a prefix; this always leaves the test fresh.

3.3 Equality of one coordinate

The next application allows an arbitrarily large set of state labels without paying for its size in the feature norm. A state is a pair in a product $\mathcal{A} \times \mathcal{B}$. For a state $y = (a, b)$, define

$$R_y = \{(a', b') : (a' = a \text{ and } b' \neq b) \text{ or } (a' \neq a \text{ and } b' = b)\}. \quad (3.8)$$

Thus exactly one coordinate agrees. A fresh test specifies state maps $y_T(x) = (A_T(x), B_{1,T}(x))$ for all $x \in P \cup Q$. Only the finitely many labels attained on these vertex sets need be retained. The labels and their equality relations are assumed measurable in T .

Corollary 3.3 (Prediction of one-coordinate agreement). *Let ω be the finite point-center law, π its two-step point-pair law, and $D = D_{\text{KL}}(\pi \| \mu \otimes \mu)$. Let $K \geq 1, L \geq 2$ be integers, and let T be an independent random test giving the state maps just described. Set*

$$s_0 = \mathbb{E}_T \mathbb{P}_{(p,q) \sim \omega} \{y_T(p) \notin R_{y_T(q)}\}.$$

There exists a fixed history $H : P \rightarrow \mathcal{H}$, with $|H(P)| \leq L^{2K}$, for which the laws of the true, unhashed states

$$\nu_{h,T} = (y_T)_* \mu_h$$

satisfy

$$\mathbb{E}_T \mathbb{E}_{(p,q) \sim \omega} [1 - \nu_{H(p),T}(R_{y_T(q)})] \leq s_0 + \frac{4}{L} + 18(D/K)^{1/4}. \quad (3.9)$$

The history is formed from earlier tests and does not use T .

Proof. Augment each test by two fresh independent random maps from its attained coordinate labels into $[L]$, assigning independent uniform values to distinct labels. Numbering attained labels by their first occurrence in a fixed vertex order makes this construction measurable. Use the same map on both vertex sets for each coordinate, and use independent maps for the two coordinates and for different tests. The point code is its two hashed coordinates, so the alphabet has size $M = L^2$. Here e_a, e_b denote standard basis vectors in their respective copies of $\mathbb{R}^L$, and $e_{(a,b)}$ denotes the standard basis vector indexed by (a, b) in $\mathbb{R}^{[L]^2}$. For a code $(a, b) \in [L]^2$, use the feature

$$f(a, b) = (e_a, e_b, e_{(a,b)}) \in \mathbb{R}^L \oplus \mathbb{R}^L \oplus \mathbb{R}^{L^2}, \quad \|f(a, b)\| = \sqrt{3}.$$

For the hashed center state (a_q, b_q) , use

$$g(q) = (e_{a_q}, e_{b_q}, -2e_{(a_q, b_q)}), \quad \|g(q)\| = \sqrt{6}.$$

Their inner product is $\mathbf{1}_{\{a=a_q\}} + \mathbf{1}_{\{b=b_q\}} - 2\mathbf{1}_{\{(a,b)=(a_q, b_q)\}}$, exactly the indicator of the hashed relation that exactly one of the two code coordinates agrees. Therefore [Lemma 3.1](#) compares its actual and predicted probabilities with error at most $4\sqrt{18}(D/K)^{1/4}$, averaged over strict prefixes and the fresh augmented test.

For any two fixed true states, hashing preserves an equality, and creates an equality between two unequal coordinate labels with probability $1/L$. Thus the indicator that exactly one true coordinate agrees differs from its hashed counterpart with probability at most $2/L$. This applies both to the actual state pair and to the predicted pair: conditional on the earlier prefix, the fresh true test and the chosen vertices, the current hashes are still independent uniform maps. In particular the predicted point is sampled by (3.3) using no current hash. Comparing actual true, actual hashed, predicted hashed, and predicted true failures gives, in that order, total extra error

at most $4/L + 4\sqrt{18}(D/K)^{1/4}$. Average the current test and current hashes before fixing a prefix, and use $4\sqrt{18} < 18$. The resulting true-state expression no longer involves the current hashes. Its history has at most $(L^2)^{K-1} \leq L^{2K}$ values. $\square$

One form of [Corollary 3.3](#) that will be convenient is an integrated version. Let $\mathcal{T}$ be a measurable family of state tests with finite measure W , and let $m(\tau)$ bound actual failure of one-coordinate agreement at τ . For $W > 0$, sample T from the normalized measure $d\tau/W$. The resulting fixed history satisfies

$$\int_{\mathcal{T}} \mathbb{E}_{(p,q) \sim \omega} [1 - \nu_{H(p),\tau}(R_{y_T(q)})] d\tau \leq \int_{\mathcal{T}} m(\tau) d\tau + W(4/L + 18(D/K)^{1/4}). \quad (3.10)$$

For $W = 0$, the assertion holds with the empty history. In the later level application, $\mathcal{T}$ is the set of levels where both bits have appreciable frequency, with the weighted measure defined in [Section 4.2](#), and m is the incidence inconsistency probability. Consistency implies $y_T(p) \in R_{y_T(q)}$, which is the only geometric hypothesis needed for (3.10). We suppress the test subscript and write ν_h when the level being tested is fixed.

3.4 Choosing the history length

There is no conflict between small prediction error and a subpolynomial number of histories. Write $T_0 = \log t$, and suppose $D = o(T_0)$. For example, take

$$K = \left\lceil \sqrt{(D+1)T_0} \right\rceil.$$

Then $K \rightarrow \infty$, $K = o(T_0)$, and $D/K \rightarrow 0$. For a fixed alphabet this gives $M^K = t^{o(1)}$. For hashing one may additionally take

$$L = \left\lceil \exp \sqrt{T_0/K} \right\rceil.$$

Eventually $L \geq 2$, $L \rightarrow \infty$, and $K \log L = o(T_0)$, so $L^{2K} = t^{o(1)}$. These statements require no quantitative rate for D/T_0 . If a later argument requires an integer $k \rightarrow \infty$ satisfying $kK \log L = o(T_0)$ and $k^{20}\eta_t \rightarrow 0$ for an additional positive sequence $\eta_t \rightarrow 0$, it suffices to choose

$$k = \left\lceil \min \left\{ \left(\frac{T_0}{K \log L} \right)^{1/2}, \eta_t^{-1/40} \right\} \right\rceil.$$

Further finitely many slow-growth requirements can be imposed by taking a smaller such diverging sequence. Thus the prediction lemma supplies both vanishing error and the small history counts required in the subsequent geometric arguments.

3.5 The entropy argument

The sum of the conditional relative entropies of the successive code pairs is at most D . Subtracting each endpoint's mean feature conditional on its own earlier codes gives two residuals. Given the common center, the endpoints are independent, so the expected inner product of the residuals equals the expected squared norm of their mean conditional on that center. Bounding this square controls discrepancy against a center-dependent feature, although each prediction uses only its point's earlier codes.

For joint laws a_{XY}, b_{XY} on the same finite product set, write a_X, b_X for their first marginals and $a_{Y|x}, b_{Y|x}$ for their conditional second-coordinate laws. The chain rule for relative entropy [7, Theorem 2.5.3] splits the information in a transcript into conditional contributions from successive tests:

$$D_{\text{KL}}(a_{XY} \| b_{XY}) = D_{\text{KL}}(a_X \| b_X) + \sum_x a_X(x) D_{\text{KL}}(a_{Y|x} \| b_{Y|x}). \quad (3.11)$$

Nonnegativity and the chain rule imply that applying a deterministic map cannot increase relative entropy: adjoin the map to the outcome and then discard the nonnegative conditional term. Pinsker's inequality [7, Lemma 11.6.1, equation (11.138)] turns the resulting relative-entropy bounds into total-variation bounds, which control the expected test features. If D_2 denotes the source's relative entropy with base-two logarithms, its statement is $D_2(a \| b) \geq \|a - b\|_1^2 / (2 \log 2)$. For every pair of finite laws, our conventions give

$$D_{\text{KL}}(a \| b) = (\log 2) D_2(a \| b), \quad \|a - b\|_1 = 2 \text{TV}(a, b).$$

Substitution gives

$$2 \text{TV}(a, b)^2 \leq D_{\text{KL}}(a \| b). \quad (3.12)$$

For any real-valued function F on the common finite sample space, with $\|F\|_\infty = \max_x |F(x)|$, total variation controls expectations by

$$|\mathbb{E}_a F - \mathbb{E}_b F| \leq 2 \|F\|_\infty \text{TV}(a, b). \quad (3.13)$$

Proof of Lemma 3.1. First freeze all K tests. Apply their complete transcript map to both coordinates of the pair (p, r) . The actual transcript law is the image of π ; the reference law is the image of $\mu \otimes \mu$. Data processing bounds their relative entropy by D . For $j \in [K]$, write H_j for the transcript strictly before step j . Conditional on prefixes (h, h') , the reference next-code law is exactly

$$b_{j,h,h'} = (c_{T_j})_* \mu_h \otimes (c_{T_j})_* \mu_{h'}.$$

Here μ_h is formed with H_j . Let $a_{j,h,h'}$ denote the actual conditional law of the next two codes. All prefix pairs of positive actual probability have positive reference probability. Iterating (3.11) therefore gives

$$\sum_{j=1}^K \mathbb{E}_{(p,r) \sim \pi} D_{\text{KL}}(a_{j,H_j(p),H_j(r)} \| b_{j,H_j(p),H_j(r)}) \leq D. \quad (3.14)$$

Consequently, by (3.12) and Cauchy–Schwarz,

$$\frac{1}{K} \sum_{j=1}^K \mathbb{E}_{(p,r) \sim \pi} \text{TV}(a_{j,H_j(p),H_j(r)}, b_{j,H_j(p),H_j(r)}) \leq \sqrt{\frac{D}{2K}}. \quad (3.15)$$

For this fixed test sequence and step j , subtract the prediction belonging to each point's own prefix:

$$U_j(p) = f_{T_j}(c_{T_j}(p)) - \sum_x \mu_{H_j(p)}(x) f_{T_j}(c_{T_j}(x)).$$

The norm is at most $2B$. Under each reference conditional product, both residuals are centered and independent, so the conditional expectation of $\langle U_j(p), U_j(r) \rangle$ is zero. This inner product has absolute value at most $4B^2$. Thus (3.13) and (3.15) imply

$$\frac{1}{K} \sum_{j=1}^K \mathbb{E}_{(p,r) \sim \pi} \langle U_j(p), U_j(r) \rangle \leq 8B^2 \sqrt{\frac{D}{2K}}. \quad (3.16)$$

Each summand is nonnegative, for conditional independence in the two-step walk law gives the exact identity

$$\mathbb{E}_{(p,r) \sim \pi} \langle U_j(p), U_j(r) \rangle = \mathbb{E}_{q \sim \kappa} \left\| \mathbb{E}_{p \sim \omega(\cdot|q)} U_j(p) \right\|^2. \quad (3.17)$$

The discrepancy against a center-dependent feature is controlled by this same conditional mean: for each fixed test sequence and step,

$$\left| \mathbb{E}_\omega \langle U_j(p), g_{T_j}(q) \rangle \right| \leq B' \left(\mathbb{E}_\kappa \left\| \mathbb{E}_{p|q} U_j(p) \right\|^2 \right)^{1/2}.$$

Thus Cauchy–Schwarz over the tests and step, together with (3.16)–(3.17), bounds the average absolute discrepancy by

$$B' \left(8B^2 \sqrt{\frac{D}{2K}} \right)^{1/2} \leq 4BB'(D/K)^{1/4}.$$

Everything proved so far holds after averaging the initially frozen tests. The residual at step j uses only $T_1, \dots, T_j$. Since the tests are independent and identically distributed, its joint law with the strict prefix is unchanged when T_j is replaced by a fresh test after that prefix. This proves (3.4). Finally average over the strict prefixes and choose one attaining at most the average. There are at most $M^{j-1} \leq M^{K-1}$ possible histories for a prefix of length $j-1$. $\square$

The important comparison in this proof is conditional on two possibly different histories. The reference law uses the first endpoint’s prediction at the first coordinate and the second endpoint’s prediction at the second. No independence of the actual endpoints after their histories have been specified is asserted. Their independence is used only conditional on the center, in (3.17).

4 Heights and coordinate tests

We apply the prediction lemma to the incidence pieces of [Proposition 2.1](#). Their coordinates are real algebraic numbers. We compare the logarithms of the coordinate differences along edges at all absolute values of a number field. Their integrated threshold variation, defined below as S , cannot remain bounded. We then group nearby coordinate values into labels so that agreement of the endpoint labels tells us which of an edge’s two reciprocal coordinate differences is below its threshold, except for an error with bounded integral.

Write

$$z_x = x_1 + ix_2, \quad w_x = x_1 - ix_2$$

for a point or center $x = (x_1, x_2)$. For every edge $e = (p, q)$, put

$$d_e = z_p - z_q, \quad w_p - w_q = d_e^{-1}. \quad (4.1)$$

The second equality follows from the unit-distance equation, and in particular $d_e \neq 0$. Distinct real points have distinct z -coordinates and distinct w -coordinates. These nonzero differences remain nonzero under every field embedding. At an embedding other than the original one, we use the two embedded coordinates separately; they need not be complex conjugates.

4.1 A weighted product formula

For each configuration choose a finite normal number field $\mathcal{N}$ containing all its coordinates and i . We use repeated absolute values to express the product formula with equal weights.

Lemma 4.1 (Weighted absolute values). *There is a countable indexed list of genuine absolute values $|\cdot|_v$ on $\mathcal{N}$, allowing repetitions, with weight $1/[\mathcal{N} : \mathbb{Q}]$ at each index. We call each indexed occurrence an entry. The archimedean entries are $|u|_v = |\sigma(u)|$ for the $[\mathcal{N} : \mathbb{Q}]$ embeddings $\sigma : \mathcal{N} \rightarrow \mathbb{C}$, so their total weight is one. All other entries are nonarchimedean. For each nonzero $u \in \mathcal{N}$, only finitely many nonarchimedean entries have $|u|_v \neq 1$, and*

$$\sum_v^* \log |u|_v = 0. \quad (4.2)$$

Here and below a star denotes the sum with the stated weights.

Proof. The classification of number-field places supplies the archimedean and nonarchimedean absolute values, and the product formula makes their normalized logarithms sum to zero [14, Theorems 7.14–7.15]. We express that sum using repeated genuine absolute values with equal weights. At a finite place $\mathfrak{p}$ above a rational prime ℓ , normalize the genuine absolute value by $|\ell|_{\mathfrak{p}} = \ell^{-1}$, and put $n_{\mathfrak{p}} = [\mathcal{N}_{\mathfrak{p}} : \mathbb{Q}_{\ell}]$, where $\mathcal{N}_{\mathfrak{p}}$ is the completion. Use $n_{\mathfrak{p}}$ copies of this absolute value, each of weight $1/[\mathcal{N} : \mathbb{Q}]$. Write $\|\cdot\|_{\mathfrak{p}}$ for the finite-place value in Milne’s product formula. His local normalization formula [14, Lemma 8.6] gives, for every $u \in \mathcal{N}^{\times}$,

$$\log \|u\|_{\mathfrak{p}} = n_{\mathfrak{p}} \log |u|_{\mathfrak{p}}.$$

At a complex place his value is $|\sigma(u)|^2$; its logarithm is $\log |\sigma(u)| + \log |\bar{\sigma}(u)|$. Thus the repeated genuine absolute values have exactly the same total logarithm as the source’s normalized values. Taking logarithms of its product formula and dividing by $[\mathcal{N} : \mathbb{Q}]$ proves (4.2). Ordinary modulus also supplies the triangle inequality needed below; squared modulus itself does not satisfy it. There are exactly $[\mathcal{N} : \mathbb{Q}]$ embeddings, giving archimedean weight one. Finally, the principal fractional ideal of any nonzero u has finitely many prime factors [14, Theorem 3.20]; at every other finite place $|u|_v = 1$. $\square$

Using the weighted sum from Lemma 4.1, define the absolute logarithmic height [20, Section 4.1] by $h(0) = 0$ and

$$h(u) = \sum_v^* \log^+ |u|_v \quad (u \neq 0).$$

Here $\log^+ r = \max(0, \log r)$ for $r > 0$, and $\log^+ 0 = 0$.

Lemma 4.2 (Height inequalities). *The value $h(u)$ is independent of the chosen finite normal field and of the extensions of the prime absolute values in Lemma 4.1. For $u \in \mathcal{N}^{\times}$,*

$$h(u^{-1}) = h(u), \quad \sum_v^* |\log |u|_v| = 2h(u). \quad (4.3)$$

For arbitrary $u, v \in \mathcal{N}$,

$$h(uv) \leq h(u) + h(v), \quad h(u+v) \leq h(u) + h(v) + \log 2. \quad (4.4)$$

If a random $X \in \mathcal{N}$ is independent of a uniformly sampled embedding $\sigma : \mathcal{N} \rightarrow \mathbb{C}$, then, for $B > 1$ and $0 < r < 1$,

$$\mathbb{P}\{|\sigma(X)| > B\} \leq \frac{\mathbb{E}h(X)}{\log B}, \quad (4.5)$$

$$\mathbb{P}\{X \neq 0, |\sigma(X)| < r\} \leq \frac{\mathbb{E}h(X)}{\log(1/r)}. \quad (4.6)$$

Proof. For a finite extension $L/\mathcal{N}$, the decomposition into completions [14, Proposition 8.2] gives $\sum_{\mathfrak{q}|\mathfrak{p}} [L_{\mathfrak{q}} : \mathcal{N}_{\mathfrak{p}}] = [L : \mathcal{N}]$. Hence

$$\sum_{\mathfrak{q}|\mathfrak{p}} \frac{[L_{\mathfrak{q}} : \mathbb{Q}_{\ell}]}{[L : \mathbb{Q}]} = \frac{[\mathcal{N}_{\mathfrak{p}} : \mathbb{Q}_{\ell}]}{[\mathcal{N} : \mathbb{Q}]}.$$

The extended absolute values restrict to the chosen normalization on $\mathcal{N}$, so the finite-place contribution of any $u \in \mathcal{N}$ is unchanged. Each embedding of $\mathcal{N}$ has $[L : \mathcal{N}]$ extensions, giving the same conclusion at infinity. A common finite normal overfield compares any two field choices. The list of absolute-value entries includes all normalized extensions with their local-degree multiplicities, so their enumeration involves no further choice. This proves height independence; $h(0) = 0$ is fixed.

The product formula equates the sums of positive and negative logarithms, giving (4.3). Multiplication gives the first height inequality termwise. The triangle inequality contributes at most $\log 2$ at an archimedean entry and no constant at a nonarchimedean entry; the total archimedean weight is one. This proves the second inequality, including zero summands by the stated convention. For the tail estimates, the average positive logarithm over embeddings is at most $h(X)$. For $X \neq 0$, its average negative logarithm is at most $h(X^{-1}) = h(X)$. Markov's inequality proves both assertions. $\square$

4.2 Thresholds and variation of logarithmic profiles

For every entry v define

$$n_e(v) = -\log |d_e|_v.$$

Choose a finite closed interval $I_v = [\ell_v, u_v]$ containing 0 and all $n_e(v)$, enlarged so that for distinct $p, p' \in P$,

$$\ell_v \leq -\log |z_p - z_{p'}|_v, \quad u_v \geq \log |w_p - w_{p'}|_v. \quad (4.7)$$

All the differences here are nonzero. By Lemma 4.1, we can take $I_v = [0, 0]$ at all but finitely many entries. Define the level space $\mathcal{L}$ to be the disjoint union of the $\{v\} \times I_v$, with weighted Lebesgue measure. For an integrable function $g : \mathcal{L} \rightarrow \mathbb{R}$, write

$$\int_{\mathcal{L}} g := \sum_v^* \int_{\ell_v}^{u_v} g(v, s) \, ds.$$

We usually suppress the domain $\mathcal{L}$ in an integral. At a level (v, s) put

$$\begin{aligned} Y_e(v, s) &= \mathbf{1}_{\{n_e(v) > s\}}, & \pi(v, s) &= \mathbb{P}_E\{Y_e(v, s) = 1\}, \\ f(v, s) &= \min(\pi(v, s), 1 - \pi(v, s)), & S &= \int f. \end{aligned} \quad (4.8)$$

Enlarging any I_v further does not change S , because all edge bits agree outside the range of the $n_e(v)$.

The choice of threshold turns distance between logarithmic profiles into bit disagreement. For independent uniform edges e, e' , Tonelli's theorem and (4.3) give

$$2\mathbb{E}_{e,e'} h(d_e/d_{e'}) = \mathbb{E}_{e,e'} \sum_v^* |n_e(v) - n_{e'}(v)| = 2 \int \pi(1 - \pi) \leq 2S. \quad (4.9)$$

4.3 Excluding bounded profile variation

Proposition 4.3. *For every sequence of incidence pieces supplied by Proposition 2.1, the profile variation S defined in (4.8) tends to infinity.*

If the variation were bounded, the height-transfer estimate below would give bounded mean heights after rescaling and translating the coordinates. Corollary 3.2 would then produce a point distinct from the original point whose predicted coordinate bins are accepted by three neighbors. At a suitably separated choice of these neighbors, this point approximately satisfies all three unit equations. The determinant in (4.15) excludes that outcome. We keep this geometric contradiction here, since it establishes the divergent scale required by all subsequent level estimates.

We first use matching labels to transfer height control from edges to points. The argument is useful because it bounds heights of differences before any translation has been chosen.

Lemma 4.4 (Transferring edge heights). *Let ω be a probability law on finite vertex sets $P \times Q$, with point marginal μ . Suppose that, for every finite set of labels and every assignment of those labels to both vertex sets, the disagreement probability of two independent μ -points is at most C_0 times the disagreement probability under ω . Then every coordinate map $U : P \cup Q \rightarrow \mathcal{N}$ satisfies*

$$\mathbb{E}_{p,p' \sim \mu} h(U_p - U_{p'}) \leq C_1 (\mathbb{E}_{(p,q) \sim \omega} h(U_p - U_q) + 1),$$

where C_1 depends only on C_0 .

Proof. Fix an entry v and $s \geq 0$. At a nonarchimedean entry give two vertices the same label if their U -coordinates differ by at most e^s in absolute value. The strong triangle inequality makes this an equivalence relation. At an archimedean entry, use the cell of a square grid in $\mathbb{C}$ of side e^s , shifted by a uniformly random fractional vector. Only the labels occurring on the finite vertex set matter.

For complex numbers whose distance is a , the probability of different grid cells is at most

$$\min(1, \sqrt{2} a e^{-s}).$$

Indeed a mismatch in either real coordinate has probability at most its coordinate distance divided by the side length, and the two coordinate distances sum to at most $\sqrt{2}a$. Its integral over $s \geq 0$ is at most $C(1 + \log^+ a)$. At a nonarchimedean entry the integral of mismatch is exactly $\log^+ |U_p - U_q|_v$, with value zero for a zero difference. Thus the integrated edge mismatch, averaged over shifts at infinity, is at most $C(\mathbb{E}_{(p,q) \sim \omega} h(U_p - U_q) + 1)$.

Apply the assumed label inequality at each fixed level and shift. For a point pair at a nonarchimedean entry, mismatch is exactly the event $|U_p - U_{p'}|_v > e^s$. At infinity, distance greater than $\sqrt{2}e^s$ forces mismatch for every shift. Integration of these tail events bounds the point-pair height, with an additional $\log \sqrt{2}$ from the total archimedean weight. This proves the result. $\square$

Proof of Proposition 4.3. Suppose $S \leq S_0$ on a subsequence. All constants in this proof may depend on the fixed S_0 . By (4.9), there is an edge value d_0 such that

$$\mathbb{E}_E h(d_e/d_0) \leq S_0.$$

Temporarily replace the two coordinates by z/d_0 and $d_0 w$. Their edge jumps are $D_{pq} = D_e = d_e/d_0$ for $e = (p, q)$ and D_e^{-1} , whose mean heights are at most S_0 . [Lemma 2.4](#) supplies the label hypothesis of [Lemma 4.4](#). Applying it to both coordinates bounds the sum of the two mean point-pair difference heights. Averaging over the second point then chooses $p_0 \in P$ such that the translated coordinates

$$Z_x = (z_x - z_{p_0})/d_0, \quad W_x = d_0(w_x - w_{p_0})$$

satisfy a fixed bound on $\mathbb{E}_\mu(h(Z_p) + h(W_p))$. Since $Z_q = Z_p - D_{pq}$ and $W_q = W_p - D_{pq}^{-1}$ on an edge, [Lemma 4.2](#) also bounds the center means. Thus there is a fixed $C_h < \infty$ with

$$\mathbb{E}_\mu(h(Z_p) + h(W_p)) + \mathbb{E}_\kappa(h(Z_q) + h(W_q)) + \mathbb{E}_E(h(D_e) + h(D_e^{-1})) \leq C_h. \quad (4.10)$$

The reciprocal rescalings preserve the edge identity for this same choice of d_0 and p_0 : on every edge,

$$(Z_p - Z_q)(W_p - W_q) = \frac{z_p - z_q}{d_0} d_0(w_p - w_q) = 1.$$

This identity remains exact after every embedding, although the two embedded coordinates need not be conjugate.

We next give the selection argument, keeping the embedding independent of the points it tests. For *any* fixed finite history map H on P , use the joint law

$$\mu(p) \mu(\hat{p} \mid H(\hat{p}) = H(p)) \prod_{j=1}^3 \mathbb{P}_E(q_j \mid p), \quad (4.11)$$

and independently choose a uniformly random embedding σ . Both p and $\hat{p}$ have marginal μ ; each q_j has marginal κ ; and each (p, q_j) has the original edge law. In particular, the three jumps $D_j = D_{pq_j}$ have the correct individual edge marginals. No assertion of mutual independence of these jumps is needed. The height inequality bounds $\mathbb{E}h(Z_{\hat{p}} - Z_p)$ and each $\mathbb{E}h(D_j - D_l)$ using only (4.10). These bounds are uniform in H .

Choose a fixed $B \geq 2$ so large that, by (4.5), the probability that any of

$$Z_p, W_p, Z_{\hat{p}}, W_{\hat{p}}, (Z_{q_j}, W_{q_j}, D_j, D_j^{-1})_{j=1}^3$$

has embedded absolute value greater than B is less than $1/100$. Then choose a fixed $0 < r < 1$ so small that (4.6) makes the probability of any of

$$\begin{aligned} p \neq \hat{p}, \quad |\sigma(Z_{\hat{p}} - Z_p)| < r, \\ D_j \neq D_l, \quad |\sigma(D_j - D_l)| < r \quad (1 \leq j < l \leq 3) \end{aligned}$$

less than $1/100$ in total. The equalities have deliberately been excluded from these small-difference events. These estimates hold for every H because they were integrated over a fresh independent embedding using only the fixed marginal bounds.

Choose now

$$0 < \tau < \min \left(1, \frac{r^4}{128B^5} \right). \quad (4.12)$$

At an embedding, encode each point by quantizing the real and imaginary parts of both Z_p, W_p in $[-B-1, B+1]$, in half-open intervals of length at most τ . Add an overflow code for points outside this box, using an endpoint convention to make a partition. This is a finite alphabet of size

$M = M(B, \tau)$ independent of t and of the field degree. An acceptance test at q asks whether the point code is attained by some actual neighbor of q . The actual point on an edge always passes.

By Lemma 2.6, $D = o(\log t)$. Choose $K \rightarrow \infty$ with $K = o(\log t)$ and $D/K \rightarrow 0$. Apply Corollary 3.2 to independent uniform embeddings as tests. It gives a fixed history map formed from the tests preceding a chosen step with at most $M^{K-1} = t^{o(1)}$ values for which the predicted acceptance failure at a fresh embedding is $o(1)$. The choice of history is made after averaging the prediction failure over that fresh embedding; the embedding itself is not fixed. The geometric error bounds just proved remain valid for this selected history.

Use this H in (4.11), with one common prediction $\hat{p}$ and three independently sampled neighbors of p . Each pair $(\hat{p}, q_j)$ has the prediction law in the acceptance estimate, so all three tests pass outside probability $o(1)$ by a union bound. Moreover

$$\mathbb{P}(\hat{p} = p) = \sum_{h: \mu(H=h) > 0} \frac{\sum_{p: H(p)=h} \mu(p)^2}{\mu(H=h)} \leq |H(P)| \max_p \mu(p) = t^{-1+o(1)}. \quad (4.13)$$

For $j \neq l$, the conditional uniform neighbor law gives $\mathbb{P}(q_j = q_l) \leq 1/\min_p d(p) = o(1)$. Distinct centers have distinct Z -coordinates, so $q_j \neq q_l$ implies $D_j \neq D_l$. The degree and atom estimates used here follow from Proposition 2.1. Combining all these estimates gives positive probability that all three tests pass, all displayed quantities are bounded by B , and

$$|\sigma(Z_{\hat{p}} - Z_p)| \geq r, \quad |\sigma(D_j - D_l)| \geq r \quad (j < l). \quad (4.14)$$

Fix such an outcome and suppress σ from the notation. Successful acceptance provides, for each j , a genuine neighbor of q_j in the same code as $\hat{p}$. Since the latter has coordinates bounded by B , this is a nonoverflow code, and the two complex coordinate errors relative to $\hat{p}$, denoted by $\varepsilon_Z, \varepsilon_W$, each have modulus at most $\sqrt{2}\tau$. Expanding its exact edge equation around the predicted coordinates gives

$$\begin{aligned} |(Z_{\hat{p}} - Z_{q_j})(W_{\hat{p}} - W_{q_j}) - 1| &\leq 2B(|\varepsilon_Z| + |\varepsilon_W|) + |\varepsilon_Z \varepsilon_W| \\ &\leq 4\sqrt{2}B\tau + 2\tau^2 \leq 8B\tau. \end{aligned}$$

The last inequality uses $B \geq 2$ and $\tau < 1$. Set $\Delta Z = Z_{\hat{p}} - Z_p$ and $\Delta W = W_{\hat{p}} - W_p$. Subtracting the exact equation for (p, q_j) yields

$$\Delta Z/D_j + \Delta W D_j + \Delta Z \Delta W = E_j, \quad |E_j| \leq 8B\tau.$$

Subtract the equation with $j = 1$ from those with $j = 2, 3$. The coefficient matrix for $(\Delta Z, \Delta W)$ has determinant

$$\det \begin{pmatrix} D_2^{-1} - D_1^{-1} & D_2 - D_1 \\ D_3^{-1} - D_1^{-1} & D_3 - D_1 \end{pmatrix} = \frac{(D_1 - D_2)(D_1 - D_3)(D_2 - D_3)}{D_1 D_2 D_3}. \quad (4.15)$$

Its absolute value is at least r^3/B^3 by (4.14). Each matrix entry is at most $2B$, and each right-hand side is at most $16B\tau$. Cramer's rule therefore gives

$$|\Delta Z| \leq \frac{64B^5\tau}{r^3} < r/2,$$

contradicting (4.14). This excludes every bounded subsequence of S and proves the proposition. $\square$

4.4 Coordinate states on the original levels

We now return to the original coordinates z, w , intervals I_v , and thresholds n_e . By [Proposition 4.3](#), $S \rightarrow \infty$. At a level (v, s) give each vertex x a state

$$y_x(v, s) = (A_x(v, s), B_{1,x}(v, s)).$$

At a nonarchimedean entry, A_x is the equivalence class for $|z_x - z_{x'}|_v \leq e^{-s}$, and $B_{1,x}$ is that for $|w_x - w_{x'}|_v \leq e^s$. At an archimedean entry associated with σ , A_x is the half-open square-grid cell containing $\sigma(z_x)$ at side length e^{-s} , and $B_{1,x}$ is the cell containing $\sigma(w_x)$ at side length e^s . For each of these two grids choose a fractional shift in $[0, 1)^2$ and use that same fractional shift, scaled with the side length, at every level of the entry.

For a state $y = (a, b)$ recall the set

$$R_y = \{(a', b') : \text{exactly one of } a' = a \text{ and } b' = b \text{ holds}\}. \quad (4.16)$$

Call an edge *consistent* at a level if $Y_e = 1$ and exactly its A -labels agree, or if $Y_e = 0$ and exactly its B_1 -labels agree. Thus consistency implies that the endpoint states satisfy $y_p \in R_{y_q}$, and also specifies which coordinate agrees. Define

$$m(v, s) = \mathbb{P}_E\{e \text{ is inconsistent at } (v, s)\}, \quad m_e = \int \mathbf{1}_{\{e \text{ is inconsistent}\}}. \quad (4.17)$$

Lemma 4.5 (State consistency). *The fractional shifts can be fixed so that*

$$\int m = \mathbb{E}_E m_e = O(1), \quad (4.18)$$

with an absolute constant. At each nonarchimedean entry every edge is consistent outside finitely many threshold levels. At a level (v, s) , any two vertices $x, x' \in P \cup Q$ with matching A -labels satisfy

$$|z_x - z_{x'}|_v \leq c_v e^{-s};$$

matching B_1 -labels similarly implies $|w_x - w_{x'}|_v \leq c_v e^s$, where $c_v = 1$ at nonarchimedean entries and $c_v = \sqrt{2}$ at archimedean entries. All the level functions can be chosen measurable.

Proof. At a nonarchimedean entry, $n_e > s$ means $|d_e|_v < e^{-s}$ and $|d_e^{-1}|_v > e^s$, so exactly A agrees. For $n_e < s$ the two conclusions reverse. Only $s = n_e$ is excluded, a null set for the level measure.

At infinity, put $r = |s - n_e|$. The smaller coordinate jump divided by its grid scale is e^{-r} , so its mismatch probability under a uniform shift is at most $\sqrt{2}e^{-r}$. The larger jump divided by its scale is e^r ; it cannot lie in the same cell once $r \geq \log \sqrt{2}$, by the diameter of a half-open square. Consequently the expected inconsistency indicator is at most

$$\mathbf{1}_{\{|s - n_e| \leq \log \sqrt{2}\}} + \sqrt{2}e^{-|s - n_e|}.$$

Its integral over the full real line is bounded by an absolute constant, and hence so is its integral over I_v . Sum over entries using total archimedean weight one, then average over edges. There are only finitely many archimedean shifts to choose at each index, so averaging fixes shifts satisfying (4.18).

The match implications follow directly from the equivalence classes and the square diameter. For measurability, label nonarchimedean classes by the first vertex in a fixed ordering of $P \cup Q$, and use integer cell indices at infinity. All class comparisons are threshold comparisons and all cell indices are measurable functions of s . $\square$

We have obtained a space of tests of finite measure with minority mass $S \rightarrow \infty$, while the integrated discrepancy between an edge's threshold bit and its two coordinate agreements remains $O(1)$. This makes that discrepancy negligible relative to S .

5 Level decomposition and bounds for integrated scores

We compare two edges sharing a point or a center by assigning a score that rewards agreement of specified coordinate labels at their opposite endpoints and subtracts disagreement of their bits. The product formula bounds the integrated score from above. A lower bound records how often a vertex differs from the label shared by most vertices in the selected coordinate. Both bounds also hold when the samples at a point are conditioned on a short list of histories.

We use the graph satisfying (2.4), the threshold bits $Y_e(v, s) = \mathbf{1}_{\{n_e(v) > s\}}$, and the states $y_x = (A_x, B_{1,x})$ of Lemma 4.5. Consistency means that exactly the A labels match when the bit is 1, and exactly the B_1 labels match when it is 0. The inconsistency probability has integral $\int m = O(1)$, negligible relative to $S \rightarrow \infty$ by Proposition 4.3.

All level integrals use the weighted measure of Section 4.2. In particular $f = \min(\mathbb{P}_E(Y = 0), \mathbb{P}_E(Y = 1))$ and $S = \int f$. Write $d_{P,\min} = \min_{p \in P} d(p)$ and $d_{Q,\min} = \min_{q \in Q} d(q)$.

5.1 Parameters, level sets, and histories

The parameters must make prediction errors small even after taking several samples, while keeping the number of possible histories small relative to a point degree.

For the incidence pieces, information D and variation S fixed above, there are positive integer sequences $K, L, k \rightarrow \infty$ such that

$$K \log L = o(\log t), \quad D/K = o(1), \quad (5.1)$$

$$kK \log L = o(\log t), \quad k^{20} \left((D/K)^{1/4} + L^{-1} + S^{-1} + d_{Q,\min}^{-1} \right) = o(1). \quad (5.2)$$

The incidence estimates give $D = o(\log t)$ and $d_{Q,\min} \rightarrow \infty$. For example, first take $K = \lceil \sqrt{(D+1) \log t} \rceil$ and then choose $L \rightarrow \infty$ with $\log L \sim \sqrt{\log t / K}$. These choices give (5.1). Both $K \log L / \log t$ and the parenthesized expression in (5.2) tend to zero. Choosing $k \rightarrow \infty$ sufficiently slowly gives (5.2). In particular $k = o(\log t)$ since $K \log L \rightarrow \infty$.

Fix these sequences and set $\epsilon = k^{-1/16}$. The hierarchy gives $k^3/S = o(1)$, $k^{3+1/16}((D/K)^{1/4} + L^{-1}) = o(1)$, and $k = t^{o(1)}$.

Recall that f is the less frequent bit's probability and m is the probability of an inconsistent edge at a level. Partition the level space by

$$\mathcal{C} = \{f \geq \epsilon\}, \quad \mathcal{B} = \{f < \epsilon, m > 1/k\}, \quad \mathcal{G} = \{f < \epsilon, m \leq 1/k\}.$$

Their first two widths satisfy

$$W_c := \int \mathbf{1}_{\mathcal{C}} \leq S/\epsilon, \quad W_b := \int \mathbf{1}_{\mathcal{B}} \leq k \int m = O(k) = o(S). \quad (5.3)$$

Prediction will be used on $\mathcal{C}$, where both bits have frequency at least ϵ . The measure of $\mathcal{B}$ is negligible compared with S ; on $\mathcal{G}$, both the minority frequency and inconsistency are small enough to select a label shared by most vertices. At each absolute-value entry, $\mathcal{C}$ is an interval, possibly empty or of zero length: the function $s \mapsto \mathbb{P}_E(n_e(v) > s)$ is nonincreasing.

At a level in $\mathcal{G}$ let $r \in \{0, 1\}$ be the minority bit, which is unique for all sufficiently large t . On every consistent majority-bit edge the B_1 labels match if $r = 1$, and the A labels match if $r = 0$. Thus the mismatch probability of this coordinate is at most $f + m$. By Lemma 2.4 it has a label ℓ whose two marginal exception probabilities have sum $O(f + m)$. Fix one such label measurably, and let $I_x(v, s)$ indicate that the relevant coordinate of x differs from ℓ . Write

$$h_P(v, s) = \mathbb{E}_\mu I_P(v, s), \quad h_Q(v, s) = \mathbb{E}_\kappa I_Q(v, s), \quad H_x^* = \int_{\mathcal{G}} I_x(v, s).$$

Here and below I_x is used only on $\mathcal{G}$. Fixed orderings of the finite vertex sets settle all choices of labels, so measurability causes no additional restriction.

Lemma 5.1 (Exception mass and joint exceptions). *For the indicators I_x of exceptions to the chosen coordinate label, uniformly over levels in $\mathcal{G}$,*

$$h_P + h_Q \leq C(f + m) = o(1), \quad f \leq h_P + h_Q + m. \quad (5.4)$$

Consequently

$$\mathbb{E}_\mu H_P^* + \mathbb{E}_\kappa H_Q^* = O(S). \quad (5.5)$$

Furthermore, with (p, q, r) sampled according to (2.11),

$$\int_{\mathcal{G}} \mathbb{P}_E(I_p = I_q = 1) = o(S), \quad \int_{\mathcal{G}} \mathbb{P}(I_p = I_r = 1) = o(S). \quad (5.6)$$

Proof. The first inequality was just proved, and its right side is uniformly $O(\epsilon + 1/k) = o(1)$. If both endpoints have the dominant label, a consistent edge has the majority bit. Every minority-bit edge therefore has an exceptional endpoint or is inconsistent, proving the second inequality. Integrating and using $\int m = O(1)$ proves (5.5) because $S \rightarrow \infty$.

For the joint errors, apply Lemma 2.5 with the uniform marginal bound $a_t = C(\epsilon + k^{-1}) \rightarrow 0$ from (5.4). Its edge estimate gives $\mathbb{P}_E(I_p = I_q = 1) = o(1)(h_P + h_Q)$, and its estimate for two neighbors of a center gives $\mathbb{P}(I_p = I_r = 1) = o(1)h_P$. These are exactly the edge and two-step walk laws used here. Both multipliers are uniform over all levels in $\mathcal{G}$, so they can be taken outside the integrals. Equation (5.5) then gives (5.6). $\square$

We next apply the prediction lemma only to levels in $\mathcal{C}$. When $W_c > 0$, sample a fresh level from their normalized measure and use Corollary 3.3. An inconsistent edge accounts for every true failure of one-coordinate agreement. We obtain a fixed history map H with $|H(P)| \leq L^{2K}$ such that, writing ν_h for the true-state law of a point of marginal μ conditional on history h at the fresh level,

$$k^3 \int_{\mathcal{C}} \mathbb{E}_E[1 - \nu_{H(p)}(R_{y_q})] = o(S). \quad (5.7)$$

Indeed the integral before multiplication by k^3 is at most $\int_{\mathcal{C}} m + O(W_c((D/K)^{1/4} + L^{-1}))$, and the hierarchy (5.2) and width bound (5.3) apply. If $W_c = 0$, take the empty history, and the assertion is vacuous.

For each center q , form an ordered list of histories $\theta = (h_1, \dots, h_k)$ obtained by sampling k independent neighbors of q and recording their histories. Let $a(q, \theta) = \mathbb{P}(\theta \mid q)$. The number M of possible history lists satisfies

$$M \leq L^{2Kk} = t^{o(1)}. \quad (5.8)$$

The history map is already fixed. Thus $a(q, \theta)$ uses no fresh level. In the joint law

$$\omega(p, q, \theta) = \frac{\mathbf{1}_{\{(p, q) \in E\}}}{|E|} a(q, \theta), \quad (5.9)$$

the point p and the history list are independent conditional on q .

5.2 Three fixed pair experiments

We use three ways to sample a base and two opposite endpoints:

- (i) Q : sample $q \sim \kappa$, then p, r independently from the conditional edge law at q ;
- (ii) P : sample $p \sim \mu$, then q, q' independently from the conditional edge law at p ;
- (iii) $P\theta$: sample (p, θ) from its marginal under ω , then q, q' independently from $\omega(q \mid p, \theta)$.

Only positive-mass bases are used. In each experiment write x for the base vertex, y_1, y_2 for its opposite endpoints, and Y_1, Y_2 for the corresponding edge bits. The base in the third experiment includes θ . At a level let $\rho = \mathbb{P}(Y_1 = 1 \mid \text{base})$ under the full indicated law. These laws are fixed across all levels. Define

$$C_i = \int_{\mathcal{C}} \mathbb{E}_{\text{base}}[2\rho(1 - \rho)], \quad i \in \{Q, P, P\theta\}. \quad (5.10)$$

Lemma 5.2 (Marginals and diagonal pairs). *In each of the experiments $Q, P, P\theta$, either edge slot has the uniform law on E after averaging over its base. In experiment $P\theta$, for every $p \in P$,*

$$\mathbb{E}_{\theta|p} \mathbb{P}(q = q' \mid p, \theta) \leq \frac{M}{d(p)},$$

where M is the number of possible history lists in (5.8).

Proof. The first assertion is immediate for the unconditioned pair samples. For the third, the (p, q) marginal of (5.9) is the edge law since $\sum_{\theta} a(q, \theta) = 1$; sampling a conditional slot preserves this marginal. More explicitly, a slice of positive probability $w_{\theta} = \sum_{p,q} \omega(p, q, \theta)$ has one-edge law $\omega(p, q, \theta)/w_{\theta}$ and must be averaged with weight w_{θ} . To compare with the Q experiment in the same slice, attach the history list at its center. The full law is

$$\mathbb{P}(p, r, q, \theta) = \kappa(q) a(q, \theta) \mathbb{P}_E(p \mid q) \mathbb{P}_E(r \mid q).$$

Given (q, θ) , the two points retain their original independent neighbor laws. Summing either point out gives $\omega(p, q, \theta)$ for the other slot. Conditioning on θ therefore gives the same single-slot law as for $P\theta$, and summing over θ recovers the original Q pair experiment.

Fix p and put $a_p(\theta) = \mathbb{P}(\theta \mid p)$. For $a_p(\theta) > 0$,

$$\omega(q \mid p, \theta) = \frac{\mathbf{1}_{\{(p,q) \in E\}} a(q, \theta)}{d(p) a_p(\theta)} \leq \frac{1}{d(p) a_p(\theta)}.$$

Its collision probability is at most its largest atom. Multiplying by $a_p(\theta)$ and summing over at most M positive-mass history lists gives the asserted bound pointwise in p . $\square$

We now define scores that reward matching opposite endpoints and charge differing bits. If $Y_1 = Y_2 = 1$, let N be the event that the B_1 labels of y_1, y_2 match; if $Y_1 = Y_2 = 0$, let N instead mean that their A labels match; set N to be false when the bits differ. Thus N concerns the coordinate which a consistent edge of that bit does not share with its base. For independent bits with parameter ρ , equality has probability $\rho^2 + (1 - \rho)^2$, whereas disagreement has probability $2\rho(1 - \rho)$. The former is larger except at $\rho = 1/2$. Thus, when equal bits force N , a reward slightly above one near balance and slightly below one elsewhere can give a positive expected score. The

decrease away from balance also makes room for the increase in the integrated upper bound below. [Section 6](#) will identify the configurations in which equal bits force the required match. Set

$$\delta = \frac{1}{100}, \quad \lambda = \frac{\delta^2}{2}, \quad c(\rho) = \begin{cases} 1 + \lambda, & |\rho - \frac{1}{2}| \leq \delta, \\ 1 - \lambda, & |\rho - \frac{1}{2}| > \delta. \end{cases}$$

The scalar λ is a score coefficient; it is unrelated to the edge ratios λ_{pr} introduced later. At a level in $\mathcal{G}$ write r for the minority bit. The score is

$$\text{score} = \begin{cases} 2I_x \mathbf{1}_{\{Y_1=Y_2=r, N\}}, & \text{at a level in } \mathcal{G}, \\ c(\rho) \mathbf{1}_{\{Y_1=Y_2, N\}}, & \text{at a level in } \mathcal{C}, -\mathbf{1}_{\{Y_1 \neq Y_2\}}. \\ 0, & \text{at a level in } \mathcal{B} \end{cases}$$

For each $i \in \{Q, P, P\theta\}$, let D_i be the expectation of this score integrated over all levels in experiment i .

5.3 The two score bounds

We bound the scores both when the common base is a center and when it is a point. Outside $\mathcal{C}$, the lower bounds for Q and $P\theta$ contain the same difference between the point and center expectations of H_x^* , with opposite signs. Adding them cancels that difference. Comparing the remaining lower bound on $\mathcal{C}$ with the upper bounds will show that $C_Q + C_{P\theta} = o(S)$.

Proposition 5.3 (Upper bound for the integrated scores). *The integrated score D_i and variation C_i of each fixed pair experiment $i \in \{Q, P, P\theta\}$ satisfy*

$$D_i \leq o(S) + 10\delta^2 \lambda C_i. \quad (5.11)$$

Proposition 5.4 (Lower bound outside $\mathcal{C}$). *For each fixed pair experiment $i \in \{Q, P, P\theta\}$, let x be the base vertex and y either opposite endpoint. Its expected score, integrated over $\mathcal{L} \setminus \mathcal{C}$, is at least*

$$2\mathbb{E}H_x^* - 2\mathbb{E}H_y^* - o(S), \quad (5.12)$$

where both expectations use the original edge marginals on the appropriate sides. In particular this is $2\mathbb{E}_\kappa H_q^* - 2\mathbb{E}_\mu H_p^* - o(S)$ for Q , and its opposite main term for either P or $P\theta$.

The upper bound is proved in [Section 5.4](#), and the lower bound in [Section 5.5](#).

5.4 The product-formula upper bound

At a fixed absolute-value entry, positive score terms lie in two intervals. The next lemma bounds how much their integral can exceed the allowance $2T$ when their coefficient rises from $1 - \lambda$ to $1 + \lambda$ on a central interval.

Lemma 5.5 (Two intervals and a central window). *Let $b \leq b'$, $T \geq 0$, and J be an interval of finite length, with any choice of included endpoints. Put $I_- = [b - T, b]$ and $I_+ = [b', b' + T]$, and let $|\cdot|$ denote length. For a real number x , write $[x]_+ = \max(x, 0)$. For $0 < \lambda < 1$,*

$$2\lambda[|J \cap I_-| + |J \cap I_+| - T]_+ \leq \lambda|J| \mathbf{1}_{\{b, b' \in \text{int } J\}}. \quad (5.13)$$

If X has any probability law on $\mathbb{R}$, and $|\mathbb{P}(X > s) - 1/2| \leq \delta$ throughout J , then

$$\mathbb{P}(X \in \text{int } J) \leq 2\delta. \quad (5.14)$$

Proof. Put $a = |J \cap I_-| + |J \cap I_+|$. The left side of (5.13) is $[2\lambda(a - T)]_+$. For it to be positive both intersections must have positive length, since either has length at most T . As J is an interval, this puts both b, b' in its open interior, also when $b = b'$. Moreover $a \leq 2T$ and $a \leq |J|$, since I_- and I_+ have disjoint interiors. Hence $2(a - T) \leq a \leq |J|$. For (5.14), take $u < v$ in the interior. Then

$$\mathbb{P}(u < X \leq v) = \mathbb{P}(X > u) - \mathbb{P}(X > v) \leq 2\delta.$$

Exhausting the open interior by such intervals proves the claim. The empty-interior case is immediate. This proof makes no assumption about atoms at either endpoint of J . $\square$

Figure 1 illustrates why positive excess in Lemma 5.5 requires both thresholds to lie inside the central window.

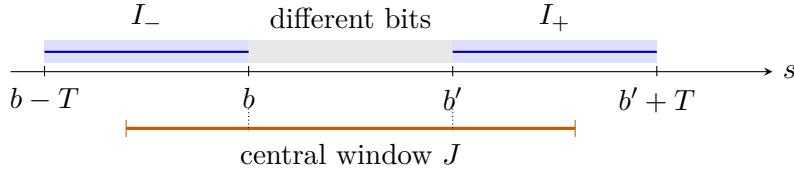


Figure 1: The intervals $I_- = [b - T, b]$ and $I_+ = [b', b' + T]$, with $b \leq b'$ and $T \geq 0$, lie on either side of $[b, b']$, where the two threshold bits differ. In Lemma 5.5, the quantity $2\lambda[|J \cap I_-| + |J \cap I_+| - T]_+$ is positive only if the overlap with J exceeds T ; then both b and b' lie strictly inside J .

Proof of Proposition 5.3. First fix a base and two distinct opposite endpoints, and write d, d' for their incident jumps. Distinct real endpoints have distinct z coordinates, so $d \neq d'$. At an absolute-value entry v let

$$b = \min(n_e(v), n_{e'}(v)), \quad b' = \max(n_e(v), n_{e'}(v)), \quad u = -\log |d - d'|_v, \quad \alpha = u - b.$$

The two opposite-endpoint coordinate differences are $d - d'$ and $1/d - 1/d'$ up to signs. Therefore

$$-\log |(d - d')(1/d - 1/d')|_v = 2\alpha - (b' - b). \quad (5.15)$$

All thresholds lie within the chosen level interval, so the integrated bit-difference charge at this entry is exactly $b' - b$.

Let $b_v = \log 2$ at archimedean entries and $b_v = 0$ otherwise. The triangle inequality gives $\alpha \geq -b_v$, hence $T_v := \alpha + b_v \geq 0$. Equality of a grid label at infinity bounds the corresponding difference by $\sqrt{2}$ times the grid scale; at a nonarchimedean entry a matching label bounds it by the scale itself. If both bits are 1, a positive charge requires $s < b$ and a B_1 match, which gives

$$s \geq b' - \alpha - b_v \geq b - T_v.$$

If both bits are 0, the analogous A match gives $s \geq b'$ and $s \leq b + \alpha + b_v \leq b' + T_v$. Thus all positive charges lie in

$$I_- = [b - T_v, b], \quad I_+ = [b', b' + T_v]. \quad (5.16)$$

Enlarging to these full intervals is legitimate even if part of them lies outside the chosen level interval.

Suppose a positive term from $\mathcal{G}$ is counted with minority bit 1 at a threshold s_0 . Then $s_0 < b$ and $\mathbb{P}_E(n_e(v) > s_0) < \epsilon$. By monotonicity every level in $\mathcal{C}$ and every level with minority bit 0

lies below s_0 ; none can produce a positive charge in I_+ . All minority-1 positive charges lie in I_- by definition. The case of a positive term from $\mathcal{G}$ with minority bit 0 is symmetric. Thus, if any positive term from $\mathcal{G}$ is counted, all positive charges are on one side, their coefficients are at most 2, and their total is at most $2T_v$.

Otherwise only positive terms from $\mathcal{C}$ remain. For this fixed base put

$$J_v = \{s : (v, s) \in \mathcal{C}, |\rho(v, s) - 1/2| \leq \delta\}.$$

This is an interval because the conditional edge law is fixed across levels and both restrictions are intervals. Extend the coefficient $1 - \lambda$ across both full intervals in (5.16), increasing it by 2λ on J_v . Put $a_v = |J_v \cap I_-| + |J_v \cap I_+|$. The extended positive integral is $2(1 - \lambda)T_v + 2\lambda a_v$; its excess over $2T_v$ is therefore $2\lambda(a_v - T_v)$. Lemma 5.5 bounds the positive part of this excess by

$$\lambda |J_v| \mathbf{1}_{\{n_e(v), n_{e'}(v) \in \text{int } J_v\}}.$$

This expression is also a valid excess bound when a positive term from $\mathcal{G}$ is counted, when no excess is needed. Combining it with the negative charge and (5.15), then summing over absolute-value entries, cancels the logarithm by the product formula for the fixed nonzero element $(d - d')(1/d - 1/d')$. The additive allowance is $2\sum_v^* b_v = 2\log 2$. Entries of zero level width can be included in this sum: their positive and negative charges vanish and the same upper bound still holds.

Now average this bound over the fixed-base independent pair law, restricting to distinct pairs. Dropping that restriction only in the nonnegative excess term, (5.14) bounds its expectation by $4\delta^2\lambda|J_v|$. Boundary atoms are not counted; distinct jumps with equal thresholds cause no exception to this argument. On J_v , $2\rho(1 - \rho) \geq 2(1/4 - \delta^2)$. Averaging over bases and absolute-value entries therefore bounds the distinct-pair contribution by

$$O(1) + \frac{4\delta^2\lambda}{2(1/4 - \delta^2)} C_i \leq O(1) + 10\delta^2\lambda C_i.$$

It remains to bound equal endpoints, to which the product formula was not applied. Their positive score is at most $2(W_c + H_x^*)$ for a base vertex x , and their negative score is zero. The ordinary collision probabilities are $1/d(q)$ and $1/d(p)$. For samples conditioned on a history list Lemma 5.2 gives the bound $M/d(p)$ pointwise in p . Consequently their total contributions are at most, respectively,

$$\frac{2}{d_{Q,\min}}(W_c + \mathbb{E}_\kappa H_q^*), \quad \frac{2}{d_{P,\min}}(W_c + \mathbb{E}_\mu H_p^*), \quad \frac{2M}{d_{P,\min}}(W_c + \mathbb{E}_\mu H_p^*).$$

Each is $o(S)$ by (5.2), Lemma 5.1, (5.8), and $d_{P,\min} = t^{1+o(a)}$. In particular multiplication by an unbounded H_p^* requires only its first moment, since the collision estimate holds for every p . Finally $O(1) = o(S)$. $\square$

5.5 The lower bound outside $\mathcal{C}$

Proof of Proposition 5.4. At a level in $\mathcal{G}$ denote the inconsistency indicator in slot j by M_j . If $I_x = 1$, both other endpoints have their dominant label, and both edges are consistent, their dominant-coordinate labels match each other but differ from the base's. Both bits must therefore be the minority bit and the required nonshared coordinate matches. The positive score is consequently at least

$$2I_x - 2I_x I_{y_1} - 2I_x I_{y_2} - 2M_1 - 2M_2.$$

If both opposite endpoints have their dominant label and both edges consistent, their bits agree regardless of whether the base is exceptional. Hence the negative charge is at most $I_{y_1} + I_{y_2} + M_1 + M_2$. Average these two pointwise bounds. Each slot has the original edge marginal by [Lemma 5.2](#), even with history lists. The integrals of the joint terms $I_x I_{y_j}$ are $o(S)$ by (5.6); the inconsistency integrals are $O(1)$. This gives (5.12) on $\mathcal{G}$. On $\mathcal{B}$ there are no positives and the negative score is at least -1 , whose total cost is $W_b = o(S)$. $\square$

6 Variation on $\mathcal{C}$ and positive exception mass

We use the fixed history map, sampled lists of histories, and three scores constructed in [Section 5](#). The first objective is to show that levels in $\mathcal{C} = \{f \geq \epsilon\}$, where both bits are frequent, contribute only $o(S)$ to the integrated probability that two independently sampled edges at a center have different bits. The second is to show that $\mathbb{E}_\mu H_p^*$, the mean measure of levels in $\mathcal{G}$ where a point lacks the chosen label, retains a positive fraction of S . Here $\mathcal{G} = \{f < \epsilon, m \leq 1/k\}$ has a rare minority bit and small inconsistency probability. Both conclusions come from comparing the upper and lower integrated score bounds, but the comparison first requires a precise description of the states permitted by a history list.

Throughout this section, $S \rightarrow \infty$, the parameters are chosen as in [Section 5.1](#), and H is the fixed history map from [Equation \(5.7\)](#). At a level, ν_h denotes the distribution of the true state (A, B_1) of a point sampled from μ conditional on $H = h$. Dependence on the level is suppressed. For a state $y = (a, b)$, recall that

$$R_y = \{(u, v) : (u = a, v \neq b) \text{ or } (u \neq a, v = b)\}.$$

A row means a set with fixed first coordinate; a column has fixed second coordinate. These conventions do not refer to a geometric direction in the plane.

6.1 Score estimates and the two deductions

The following three estimates isolate the inputs to the deductions. The first compares the two orientations. The other two apply when the point exception mean is temporarily assumed small in the proof of [Proposition 6.5](#).

Lemma 6.1. *The contribution from $\mathcal{C}$ to $D_Q + D_{P\theta}$ is at least*

$$\frac{\lambda}{4}(C_Q + C_{P\theta}) - o(S).$$

To prove this bound, [Section 6.2](#) shows that concentration of the prediction mixture on a row, column, or set of at most two states also constrains the actual point under the joint sampling law. [Section 6.3](#) classifies the resulting states and bounds their scores, keeping the original conditional bit probabilities even when some sampled edges are bad.

For a probability measure ν on states, let ν_A, ν_{B_1} denote its coordinate marginals. Their squared norms are the sums of squares of their atom masses, hence the match probabilities for two independent samples from the corresponding marginal.

Lemma 6.2. *If $\mathbb{E}_\mu H_p^* = o(S)$, then*

$$T := \int_{\mathcal{C}} \mathbb{E}_\mu \left[\|(\nu_{H(p)})_A\|_2^2 + \|(\nu_{H(p)})_{B_1}\|_2^2 - 1 \right] \leq o(S). \quad (6.1)$$

Lemma 6.3. Assume $\mathbb{E}_\mu H_p^* = o(S)$. Then

$$(\text{contribution from } \mathcal{C} \text{ to } D_P) + T \geq \frac{\lambda}{4} C_P - o(S).$$

The proof of Lemma 6.2 is in Section 6.4; the lower estimate in Lemma 6.3 is proved in Section 6.5. The hypothesis in the latter two estimates will be used only on the temporary subsequence in the proof of Proposition 6.5.

Proposition 6.4 (Variation on $\mathcal{C}$). *The variations of the fixed experiments Q and P_θ , defined in (5.10), satisfy*

$$C_Q + C_{P_\theta} = o(S). \quad (6.2)$$

Moreover their exception means over $\mathcal{G}$ satisfy

$$\mathbb{E}_\kappa H_q^* \leq \mathbb{E}_\mu H_p^* + o(S). \quad (6.3)$$

Proof. By Lemma 6.1, the contribution from $\mathcal{C}$ to $D_Q + D_{P_\theta}$ is at least $\lambda(C_Q + C_{P_\theta})/4 - o(S)$. The lower bounds in Proposition 5.4 outside levels in $\mathcal{C}$ cancel their main terms, because the two base vertices are on opposite sides. Hence

$$\frac{\lambda}{4}(C_Q + C_{P_\theta}) - o(S) \leq D_Q + D_{P_\theta} \leq 10\delta^2 \lambda(C_Q + C_{P_\theta}) + o(S),$$

where the last inequality is Proposition 5.3. Since $10\delta^2 < 1/4$, this proves (6.2).

Finally the contribution from $\mathcal{C}$ to D_Q is at least $-C_Q$. Proposition 5.4 and Proposition 5.3 therefore give

$$2\mathbb{E}_\kappa H_q^* - 2\mathbb{E}_\mu H_p^* - C_Q - o(S) \leq D_Q \leq 10\delta^2 \lambda C_Q + o(S).$$

Use $C_Q = o(S)$ to obtain (6.3). $\square$

Proposition 6.5 (Positive exception mass on $\mathcal{G}$). *For the exception indicators I_p and their integrated masses H_p^* chosen in Section 5,*

$$\liminf_{t \rightarrow \infty} \frac{\mathbb{E}_\mu H_p^*}{S} > 0. \quad (6.4)$$

Proof. If the conclusion fails, restrict to a subsequence on which $\mathbb{E}_\mu H_p^* = o(S)$. By (6.3), also $\mathbb{E}_\kappa H_q^* = o(S)$, and Lemma 6.2 gives $T \leq o(S)$.

By Lemma 6.3,

$$(\text{contribution from } \mathcal{C} \text{ to } D_P) + T \geq \frac{\lambda}{4} C_P - o(S).$$

The contribution from outside $\mathcal{C}$ to D_P is at least $-o(S)$ by Proposition 5.4, under the present assumptions on both exception means. Therefore

$$\frac{\lambda}{4} C_P - o(S) \leq D_P + T \leq 10\delta^2 \lambda C_P + o(S),$$

and again $10\delta^2 < 1/4$ yields $C_P = o(S)$.

To finish, let V_P, V_Q be the variations for the two unconditioned experiments at any fixed level. Assign to each vertex its conditional majority bit, breaking ties in a fixed way. The probability that the bit on an incident edge differs from this label is at most the corresponding average variation, because $\min(\rho, 1 - \rho) \leq 2\rho(1 - \rho)$. Thus the two endpoint labels disagree on an edge

with probability at most $V_P + V_Q$. When that sum is small, the matching-label consequence of expansion in Lemma 2.4 gives one bit label with total marginal exception probability $O(V_P + V_Q)$. The edge bit differs from it with probability $O(V_P + V_Q)$ as well. When the sum is not small the same bound follows from $f \leq 1/2$, with a larger absolute constant. Consequently

$$f \leq C(V_P + V_Q)$$

uniformly for all sufficiently large t . Integrating on levels in $\mathcal{C}$ gives $\int_{\mathcal{C}} f \leq C(C_P + C_Q) = o(S)$.

On levels in $\mathcal{G}$, the minority-bit probability is bounded by the two marginal exception probabilities and the inconsistency probability, as in Lemma 5.1. Its integral is $o(S)$ under the present assumptions. Levels in $\mathcal{B}$ contribute at most $W_b/2 = o(S)$. Thus $S = \int f = o(S)$, a contradiction. This proves (6.4). $\square$

6.2 From sampled histories to the actual point

For a center q , take the expectation over its conditional neighbor law in E and define the mixture of predictions

$$\tilde{\nu}_q = \mathbb{E}_{p|q} \nu_{H(p)}.$$

For its history list $\theta = (h_1, \dots, h_k)$, define the finite set of candidate center states

$$\mathcal{Y}_\theta = \{y \text{ a center state} : \nu_{h_j}(R_y) \geq 1 - k^{-2} \text{ for every } 1 \leq j \leq k\}. \quad (6.5)$$

The actual state y_q is usually a candidate: the union bound and Markov's inequality give, after integration on $\mathcal{C}$,

$$\int_{\mathcal{C}} \mathbb{P}(y_q \notin \mathcal{Y}_\theta) \leq k^3 \int_{\mathcal{C}} \mathbb{E}_E[1 - \nu_{H(p)}(R_{y_q})] = o(S). \quad (6.6)$$

We need every candidate's agreement set R_y to have $\tilde{\nu}_q(R_y)$ close to one. The next lemma proves this uniformly over candidates by applying the ε -net double-sampling argument to these sets [9, Lemmas 3.4–3.5]: a second independent sample detects mass outside an agreement set, and bounding the number of subsets cut out on the pooled samples gives control uniformly over candidates. The proof below supplies this bound for R_y and handles the weighted sampling law. We then show when concentration under the prediction mixture also constrains the actual point state.

Lemma 6.6 (Candidate states and the prediction mixture). *For sufficiently large integers k , the candidates $\mathcal{Y}_\theta$ defined in (6.5) satisfy*

$$\mathbb{P}(\exists y \in \mathcal{Y}_\theta : \tilde{\nu}_q(R_y) < 1 - k^{-1/4} \mid q) \leq 4(2k+1)^2 2^{-k^{3/4}/2} \quad (6.7)$$

at every level, uniformly over centers q . Here the k entries of θ are independent neighbor histories conditional on q , as in (5.9).

Proof. We first prove a sampling bound. Let ν be any probability measure on a finite product set and let $X_1, \dots, X_k$ be independent samples from ν . For $0 < \alpha < 1$ and $k\alpha \geq 8$,

$$\mathbb{P}(\exists y : \nu(R_y) < 1 - \alpha, X_1, \dots, X_k \in R_y) \leq 2(2k+1)^2 2^{-k\alpha/2}.$$

Adjoin independent samples $X_{k+1}, \dots, X_{2k}$. If the first k samples admit a set R_y as in the event, select one by a fixed ordering. For this selected set R_y , the number of misses among the new samples has mean $k\beta$, where $\beta = \nu(R_y^c) > \alpha$, and variance at most $k\beta$. Thus the probability of

fewer than $k\alpha/2$ misses is at most $4/(k\beta) \leq 1/2$. It suffices to bound twice the probability that some R_y contains the first half and misses at least $k\alpha/2$ of the second half.

Condition on the pooled $2k$ samples and split their indices uniformly into two halves. A set R_y induces at most $(2k+1)^2$ different subsets of these indices: its first coordinate can agree with one of at most $2k$ observed coordinates, or none, and the same holds for its second coordinate. If a pattern has m misses, the probability that they all fall in the second half is zero for $m > k$, and otherwise is

$$\frac{k(k-1)\cdots(k-m+1)}{(2k)(2k-1)\cdots(2k-m+1)} \leq 2^{-m}.$$

The union bound proves this sampling bound. Repeated sampled states cause no problem, since the argument concerns the indexed samples.

Now, after sampling the histories independently given q , independently sample a state from each corresponding ν_{h_j} . Before the histories are revealed these states are independent with law $\tilde{\nu}_q$. If a candidate violating the required conclusion exists, select one as a function of the history list. Conditional on that history list, all the synthetic states lie in R_y with probability at least $1 - k \cdot k^{-2} = 1 - 1/k$. Apply the sampling bound with $\alpha = k^{-1/4}$ and divide by $1 - 1/k$. $\square$

Since $W_c \leq k^{1/16}S$, the integral of the failure probability in (6.7) is $o(S)$. A candidate therefore constrains the mixture of predicted states. To constrain the actual point, we will use only rows, columns and sets of at most two states. This restriction is essential to the next argument.

Lemma 6.7 (Transfer to rows, columns, or two states). *Fix a level and use the joint law (5.9) of edges and history lists,*

$$\mathbb{P}(p, q, \theta) = \mathbb{P}_E(p, q) \mathbb{P}(\theta \mid q).$$

Let $0 < \eta < 1/2$. As a function of (q, θ) , select a set $U(q, \theta)$ that is a row, a column, or a set of at most two states. On an event G determined by (q, θ) , suppose

$$\tilde{\nu}_q(U(q, \theta)) \geq 1 - 3\eta^2.$$

Then the actual point state y_p satisfies

$$\mathbb{P}(G, y_p \notin U(q, \theta)) \leq 8\eta.$$

The form of the set and its coordinates may depend on both q and θ .

Proof. Conditional on (q, θ) , the history of the actual point has its ordinary conditional law given q , because the history list is independent of the point given q . Hence on G ,

$$\mathbb{E}[\nu_{H(p)}(U(q, \theta)^c) \mid q, \theta] \leq 3\eta^2.$$

The event that this set has $\nu_{H(p)}$ -mass less than $1 - \eta$ therefore has probability at most 3η .

For each history h , define three exceptional sets using only ν_h . If a row has mass at least $1 - \eta$, it is unique, and the first exceptional set is its complement. If no such row exists, the first exceptional set is empty. Define the second set in the same way for columns. Their masses are at most η each.

If a set of at most two states has mass at least $1 - \eta$, put

$$M_h = \{x : \nu_h(x) > \eta\}$$

and take M_h^c as the third exceptional set. Every set U of at most two states and mass at least $1 - \eta$ contains M_h : omitting an atom of mass greater than η would violate that mass bound. Choosing one such U also gives

$$\nu_h(M_h^c) \leq \nu_h(U^c) + \nu_h(U \setminus M_h) \leq \eta + 2\eta = 3\eta.$$

If no such U exists, the third exceptional set is empty. The union of the three exceptional sets thus has mass at most 5η .

Let E_h denote this union of exceptional states. It is determined by the history and the fixed level. Summing the joint law over centers and history lists leaves the point marginal μ , so

$$\mathbb{P}(y_p \in E_{H(p)}) = \sum_h \mu(H = h) \nu_h(E_h) \leq 5\eta.$$

This bound holds for the entire joint law, even though the selected set $U(q, \theta)$ depends on both the center and its history list. Outside the event $y_p \in E_{H(p)}$ and the earlier nonconcentration event, the actual state belongs to whichever permitted set was selected. Adding the two error bounds proves the result. $\square$

In the application, at most three candidates will be chosen by a fixed ordering at each level and history list. Outside the failure event bounded in (6.7), the intersection of their sets R_y has mixture mass at least $1 - 3k^{-1/4}$. When this intersection is contained in a row or a column, use that row or column as U ; when it has size at most two, use the intersection itself. With $\eta = k^{-1/8}$, Lemma 6.7 costs at most

$$8k^{-1/8}W_c \leq 8k^{-1/16}S = o(S).$$

An empty intersection cannot have the asserted mixture mass for large k . The proof has used the actual point marginal only for exceptional events defined by the history and level. It has made no assertion about the law of the actual state after additionally conditioning on a selected set.

6.3 The lower score bound from candidate states

We now apply this transfer to the finite geometry of the sets R_y . Two distinct center states in one row, say (a, b) and (a, b') , have intersection

$$R_{(a,b)} \cap R_{(a,b')} = \{(a, v) : v \notin \{b, b'\}\}. \quad (6.8)$$

A third center state outside that row cuts this intersection down to at most one point. The corresponding assertions hold for columns. If two states (a, b) and (a', b') differ in both coordinates, their sets R_y intersect in

$$R_{(a,b)} \cap R_{(a',b')} = \{(a, b'), (a', b)\}. \quad (6.9)$$

Three states whose first coordinates are all distinct and whose second coordinates are all distinct have sets R_y with empty intersection. Indeed each of the two points in (6.9) fails to share a coordinate with the third state. These observations give an exhaustive classification: a candidate set is empty, is a singleton, contains two distinct states sharing a coordinate, or consists of states with pairwise distinct coordinates. In the shared-coordinate case, either all candidates lie in that row or column, or a third lies outside it. In the pairwise-distinct case there are either two candidates or at least three.

Figure 2 shows the two-corner case: either endpoint state together with the bit determines the state at the other endpoint.

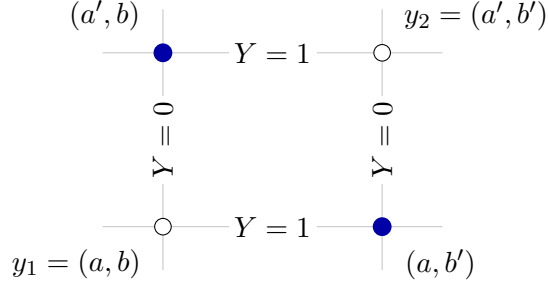


Figure 2: Two candidate center states differing in both coordinates admit exactly two point states in $R_{y_1} \cap R_{y_2}$. The filled points are those opposite corners. Within this four-state configuration, the bit of a consistent edge and the state at either endpoint determine the state at the other endpoint. Rows have fixed first label and columns have fixed second label. The diagram displays coordinate equalities of states, not geometric positions of the original plane points.

Declare a single edge with its history list bad at a level in $\mathcal{C}$ if its state test is inconsistent, if $y_q \notin \mathcal{Y}_\theta$, if the event inside the probability in (6.7) occurs, or if the transfer to the selected set fails for the witnesses specified by this classification. The witnesses are two states sharing a coordinate, with a third outside that row or column when one exists; two states differing in both coordinates; or three states with pairwise distinct coordinates. Every other edge with its history list is called good at this level. Empty candidate sets and empty witness intersections have no good edges. All choices are deterministic functions of the level and history list. The preceding estimates and the integrated inconsistency bound give

$$\int_{\mathcal{C}} \mathbb{P}(p, q, \theta \text{ is bad}) = o(S). \quad (6.10)$$

The comparisons below must use the full conditional bit probability ρ in the definition of the scores. We never remove bad edges and renormalize that law. In particular the discontinuity of $c(\rho)$ at $1/2 \pm \delta$ introduces no error.

Here is the numerical estimate used in the two-corner case. For every $0 \leq \rho \leq 1$, with $\delta = 0.01$, $\lambda = \delta^2/2$, and c as in Section 5,

$$\begin{aligned} c(\rho)(\rho^2 + (1 - \rho)^2) - 2\rho(1 - \rho) &= (2\rho - 1)^2 + (c(\rho) - 1)(\rho^2 + (1 - \rho)^2) \\ &\geq \lambda/2. \end{aligned} \quad (6.11)$$

For $|\rho - 1/2| \leq \delta$, discard the first square and use $\rho^2 + (1 - \rho)^2 \geq 1/2$. Outside that interval the expression is at least $4\delta^2 - \lambda > \lambda/2$.

Proof of Lemma 6.1. Fix a level in $\mathcal{C}$. For each history list of positive probability, let $\pi_\theta = \mathbb{P}(\theta)$ and $b_\theta = \mathbb{P}(\text{bad edge} \mid \theta)$, with both probabilities taken under the joint law in Lemma 6.7. In the experiment Q , condition on θ and then sample q by its actual conditional law. The two points are still independent with their ordinary law given q . In experiment $P\theta$, sample p conditional on θ , then two independent centers given (p, θ) . Each slot in either experiment has the same conditional single-edge law $\mathbb{P}(p, q \mid \theta)$. Therefore a pair with at least one bad slot has probability at most $2b_\theta$ in either experiment.

Write $V_Q(\theta)$ and $V_{P\theta}(\theta)$ for the corresponding averages of $2\rho(1 - \rho)$ over bases in this slice, and write $d_Q(\theta), d_{P\theta}(\theta)$ for its two expected scores. Each V is at most $1/2$, and either score is at

least the negative of its variation. We prove

$$d_Q(\theta) + d_{P\theta}(\theta) \geq \frac{\lambda}{4}(V_Q(\theta) + V_{P\theta}(\theta)) - Cb_\theta \quad (6.12)$$

with an absolute constant C . All coefficients in a score at a level of $\mathcal{C}$ are bounded by $1 + \lambda$, so charging a pair with a bad slot always costs at most an absolute multiple of its probability.

An empty candidate set. Every edge is bad, so $b_\theta = 1$, and the boundedness of the scores proves (6.12) on increasing C .

One candidate state. At a fixed base consisting of a point and history list, two good edges have the same center state. Consistency forces their bits to agree, and the nonshared coordinate of their center states matches. Thus this score is at least $1 - \lambda - O(b_\theta)$ after averaging over its bases. The score in experiment Q is at least $-1/2$. Their sum is at least $1/2 - \lambda - O(b_\theta)$, which implies (6.12) because the two variations sum to at most one.

Two candidates sharing a coordinate, and all candidates in that row or column. The transfer using (6.8) places every good point in the same row or column. If it is a row, every good edge has matching A -coordinates and hence bit 1, by consistency; in the column case every good edge has bit 0. A bit difference in either pair experiment therefore requires a bad slot. This gives $V_Q(\theta) + V_{P\theta}(\theta) \leq 4b_\theta$. The lower score bound $d_i \geq -V_i$ now proves (6.12).

Two candidates sharing a coordinate, and a third outside that row or column. The sets R_y for the three witnesses intersect in at most one state. If the intersection is empty, all edges are bad. Otherwise all good point states equal that one state. At a center base, two good edges consequently have equal bits and the required nonshared-coordinate match. Its score is at least $1 - \lambda - O(b_\theta)$, and the other score is at least $-1/2$. The same estimate as in the singleton case applies.

No coordinate shared by distinct candidates. If there are at least three candidates, their sets R_y have empty intersection and all edges are bad. Otherwise there are exactly two candidates. The good point states lie in their two opposite corners from (6.9). At either type of base, a good edge's bit determines the state at the other endpoint. Thus two good edges with equal bits have the required collision. Let e_b be the probability of a pair with a bad slot at the fixed base. The full independent pair law has equal-bit probability $\rho^2 + (1 - \rho)^2$, so

$$\mathbb{P}(Y_1 = Y_2, N \mid \text{base}) \geq \rho^2 + (1 - \rho)^2 - e_b.$$

Its disagreement probability is still $2\rho(1 - \rho)$. Multiplying by the unchanged $c(\rho)$ therefore gives the left side of (6.11), less at most $(1 + \lambda)e_b$. After averaging bases, each score is at least $\lambda/2 - O(b_\theta)$, which proves (6.12).

This completes the classification. Multiply (6.12) by its actual slice weight π_θ , sum, and integrate over levels in $\mathcal{C}$. The error is $C \int_{\mathcal{C}} \sum_\theta \pi_\theta b_\theta = o(S)$ by (6.10). This argument allows individual b_θ to be large; no inverse slice weight or number of history lists enters. The contributions from $\mathcal{C}$ to $D_Q + D_{P\theta}$ are therefore at least

$$\frac{\lambda}{4}(C_Q + C_{P\theta}) - o(S).$$

□

6.4 The prediction-pair bound

The integrand defining T is the sum of the two coordinate match probabilities, less one, for two independent predictions conditional on the same history. Its full-level integral is bounded by the

product formula. Under the temporary hypothesis in [Lemma 6.2](#), removing levels outside $\mathcal{C}$ costs only $o(S)$.

Proof of Lemma 6.2. Choose a history with its marginal probability, then two independent points p', p'' from μ conditional on that history. This is one fixed coupling throughout all levels, and each point marginal is μ . At a fixed level (v, s) , write $A(x) = A_x(v, s)$ and $B_1(x) = B_{1,x}(v, s)$. The expected value of

$$F(p', p'') = \mathbf{1}_{\{A(p')=A(p'')\}} + \mathbf{1}_{\{B_1(p')=B_1(p'')\}} - 1$$

is the integrand defining T .

For a distinct pair put $Z = z_{p'} - z_{p''}$ and $W = w_{p'} - w_{p''}$, both nonzero. At an absolute-value entry with level interval $[l, u]$, the endpoint conditions from (4.7) say $l \leq -\log |Z|_v$ and $u \geq \log |W|_v$. Put $c_v = 0$ at nonarchimedean entries and $c_v = \log \sqrt{2}$ at infinity. By [Lemma 4.5](#), an A -match requires $s \leq -\log |Z|_v + c_v$, and a B_1 -match requires $s \geq \log |W|_v - c_v$. Put $a_v = -\log |Z|_v$ and $b_v = \log |W|_v$. The two sets of matching levels inside $[l, u]$ have measures at most $a_v - l + c_v$ and $u - b_v + c_v$, respectively: these upper bounds remain valid when the containing threshold interval is truncated by an endpoint, because $l \leq a_v$ and $u \geq b_v$. Subtracting the full width gives

$$\begin{aligned} \int_l^u F(p', p'') ds &\leq (a_v - l + c_v) + (u - b_v + c_v) - (u - l) \\ &= -\log |Z|_v - \log |W|_v + 2c_v. \end{aligned}$$

The product formula, and total archimedean weight one, show that the integral over all levels is at most an absolute constant.

At a level in $\mathcal{G} = \{f < \epsilon, m \leq 1/k\}$ where neither point is exceptional, their dominant coordinate labels match, so $F \geq 0$. In general at such a level

$$F \geq -\mathbf{1}_{\{p' \text{ exceptional}\}} - \mathbf{1}_{\{p'' \text{ exceptional}\}}.$$

The contribution on levels in $\mathcal{B}$ is at least $-W_b$. Removing all levels outside $\mathcal{C}$ from the full integral consequently gives, for each distinct pair,

$$\int_{\mathcal{C}} F(p', p'') \leq O(1) + H_{p'}^* + H_{p''}^* + W_b.$$

On the diagonal the integral over $\mathcal{C}$ is exactly $W_{\mathcal{C}}$. Its probability under the selected coupling is

$$\sum_{p \in P} \frac{\mu(p)^2}{\mu(H^{-1}(H(p)))} \leq |H(P)| \max_p \mu(p) = t^{-1+o(1)}.$$

Here $|H(P)| \leq L^{2K} = t^{o(1)}$ and the marginal degree bounds were proved in [Section 2](#). Since $W_{\mathcal{C}} \leq k^{1/16}S$ and $k = t^{o(1)}$, its contribution is $o(S)$. Average the bound for distinct pairs, using the two marginals μ , $W_b = o(S)$, and $S \rightarrow \infty$. This proves $T \leq 2\mathbb{E}_{\mu} H_p^* + o(S) = o(S)$. $\square$

6.5 The second lower score bound

For the ordinary point-based experiment, the actual point already lies in the one-coordinate agreement set of each consistent neighbor. The same state classification can therefore compare its score with the two coordinate match probabilities defining T , without a sampled list of histories.

Proof of Lemma 6.3. At a level in $\mathcal{C}$ declare an ordinary edge bad if its test is inconsistent or if $\nu_{H(p)}(R_{y_q}) < 1 - k^{-2}$. Its integrated probability is $o(S)$, because Markov's inequality bounds the second contribution by $k^2 \int_{\mathcal{C}} \mathbb{E}_E[1 - \nu_{H(p)}(R_{y_q})] = o(S)$. Fix a point p and this level. Write b_p for its conditional bad-edge probability, $\nu = \nu_{H(p)}$, $\zeta = k^{-2}$, and

$$\tau(\nu) = \|\nu_A\|_2^2 + \|\nu_{B_1}\|_2^2 - 1.$$

Let $\mathcal{Y}_p$ be the states of its good neighbors. Every such state has $\nu(R_y) \geq 1 - \zeta$, and the actual point state belongs to each R_y , by consistency. Let $d_P(p)$ be the score at a level in $\mathcal{C}$ at this sample based at a point, using its full bit probability ρ_p . We claim

$$d_P(p) + \tau(\nu) \geq \frac{\lambda}{4} 2\rho_p(1 - \rho_p) - C(b_p + \zeta). \quad (6.13)$$

As before, a pair with a bad slot costs only $O(b_p)$; its probability is at most $2b_p$. The classification of intersections of the sets R_y proves the claim as follows.

If $\mathcal{Y}_p$ is empty, $b_p = 1$, while $d_P(p) \geq -1/2$ and $\tau(\nu) \geq -1$, so the claim holds for an absolute C .

If $\mathcal{Y}_p = \{(a, b)\}$, put $r = \nu_A(a)$, $s = \nu_{B_1}(b)$. Since $r + s \geq \nu(R_{(a,b)}) \geq 1 - \zeta$,

$$\tau(\nu) \geq r^2 + s^2 - 1 \geq \frac{1}{2}(1 - \zeta)^2 - 1 \geq -\frac{1}{2} - \zeta.$$

Two good neighbors have the same state and bit, so $d_P(p) \geq 1 - \lambda - O(b_p)$. Their sum has the positive lower bound needed for (6.13).

If two states share a row, the intersection of their sets R_y lies in that row, with ν -mass at least $1 - 2\zeta$. Thus $\tau(\nu) \geq (1 - 2\zeta)^2 - 1 \geq -4\zeta$. If every state is in that row, the actual point lies there too, since it belongs to R_y for both distinct states. All good edges consequently have bit 1. Their full-law variation is at most $2b_p$, and $d_P(p) \geq -2b_p$, proving the claim in this case. Columns give bit 0 and the same estimates. If there is a third state outside the row or column, the three sets R_y have intersection of size at most one and mass at least $1 - 3\zeta$. It must be a singleton for large k . Both marginals then have an atom of at least that mass, giving

$$\tau(\nu) \geq 2(1 - 3\zeta)^2 - 1 \geq 1 - 12\zeta.$$

Together with $d_P(p) \geq -1/2$, this is more than sufficient.

Finally suppose no coordinate is shared by distinct states in $\mathcal{Y}_p$. Three such states are impossible, since the actual point belongs to all their sets R_y and their intersection is empty. In the remaining case there are two states differing in both coordinates. The two opposite corners carry ν -mass at least $1 - 2\zeta$. Each coordinate marginal therefore has at least that mass on a set of size two, so

$$\tau(\nu) \geq 2 \cdot \frac{(1 - 2\zeta)^2}{2} - 1 \geq -4\zeta.$$

The actual point is one of those corners. Its bit determines the good neighbor state; hence (6.11) gives $d_P(p) \geq \lambda/2 - O(b_p)$. This completes the proof of (6.13).

Average (6.13) over p and integrate on $\mathcal{C}$. The bad-edge integral is $o(S)$, and $\zeta W_c \leq k^{-31/16} S = o(S)$. The result is

$$(\text{contribution from } \mathcal{C} \text{ to } D_P) + T \geq \frac{\lambda}{4} C_P - o(S).$$

□

The bound on $\mathbb{E}_\mu H_p^*$ alone would allow its entire contribution to come from very rare points with exceptionally large H_p^* . [Section 7](#) bounds that contribution first. It then discards pairs whose approximation error is large relative to $H_p^* + H_r^*$, bounding the lost expectation of this sum before restricting the remaining pairs to one interval $[J, 2J]$. Throughout that selection, the bound $C_Q = o(S)$ for variation in experiment Q remains available under the original two-step walk law.

7 Extracting pairs with contrasting height bounds

We now use [Propositions 6.4](#) and [6.5](#): the variation in experiment Q satisfies $C_Q = o(S)$, whereas the point exception mass satisfies $\liminf \mathbb{E}_\mu H_p^*/S > 0$. Here $S \rightarrow \infty$, and the parameters, levels, and exception indicators are those of [Section 5](#). Our objective is a graph on two copies of P whose edges carry ratios of unit jumps. Each ratio will have large height. Whenever all four pairs $(p, r), (p, r'), (p', r), (p', r')$ are edges, the quotient $\lambda_{pr}\lambda_{p'r'}/(\lambda_{pr'}\lambda_{p'r})$ will have negligible height on the same scale. Such four edges form a rectangle in the bipartite graph; no geometric rectangle is required.

Proposition 7.1 (Height bounds for ratios and their four-edge quotients). *For the fixed sequence of incidence graphs and levels, there are $J \rightarrow \infty$ and a simple bipartite graph $\mathcal{E}$ between two copies of P , each of size at most $t^{1+o(1)}$, with at least $t^{2-o(1)}$ edges, having the following properties. Each edge (p, r) has distinct endpoints and an assigned center $q = q(p, r) \in Q$ with $pq, rq \in E$. Set*

$$\lambda_{pr} = \frac{z_p - z_{q(p,r)}}{z_r - z_{q(p,r)}}.$$

No edge has both endpoints on any real unit circle containing at least $t^{9/10}$ points of P . Uniformly over the edges and complete rectangles of $\mathcal{E}$,

$$(z_p - z_r)(w_p - w_r) = 2 - \lambda_{pr} - \lambda_{pr}^{-1}, \quad (7.1)$$

$$h(\lambda_{pr}) \geq J/3, \quad (7.2)$$

$$h\left(\frac{\lambda_{pr}\lambda_{p'r'}}{\lambda_{pr'}\lambda_{p'r}}\right) = o(J). \quad (7.3)$$

In the last line all four indicated edges are required to be present; their assigned centers need not agree.

Throughout this section a two-step walk is sampled by taking q with law κ and then taking p, r independently and uniformly among its neighbors in E . Its two point marginals are μ . We retain the notation W_c for the width of levels in $\mathcal{C}$, W_b for the width of levels in $\mathcal{B}$, and m_{pq} for the integrated inconsistency indicator of the edge pq . In particular,

$$\mathbb{E}_\mu H_p^* + \mathbb{E}_\kappa H_q^* = O(S), \quad \mathbb{E}_E m_{pq} = O(1), \quad W_c \leq S/\epsilon, \quad W_b = o(S), \quad \epsilon = k^{-1/16}. \quad (7.4)$$

All integrals over levels include the prescribed weights of the absolute-value entries.

7.1 From a height band to the pair graph

The deduction of [Proposition 7.1](#) uses an approximation to the logarithm of each ratio by a difference of endpoint quantities. Each endpoint then appears once with each sign in a four-edge

quotient, so its contribution cancels. We define those quantities before stating the band estimate that makes the approximation uniform.

On a level $(v, s) \in \mathcal{G}$, use the exception indicator $I_x(v, s)$ from [Section 5](#), and put $\sigma(v, s) = 1$ when one is the minority bit and -1 when zero is the minority bit. Define a real-valued profile, indexed by absolute-value entries, by

$$a_p(v) = \int_{\{s: (v, s) \in \mathcal{G}\}} \sigma(v, s) I_p(v, s) ds. \quad (7.5)$$

For a real-valued profile $b = (b(v))_v$ with finite support, write $\|b\|_1 = \sum_v^* |b(v)|$. For an edge (p, q) , the profile notation means $n_{pq}(v) = -\log|z_p - z_q|_v$; as before, $[x]_+ = \max(x, 0)$.

Lemma 7.2 (A height band with small profile errors). *Along the fixed sequence, there are a constant $c > 0$, a sequence $\delta_t \rightarrow 0$, and scales $J \rightarrow \infty$ for which, at every sufficiently large index, a set of two-step walks has probability at least*

$$\frac{c}{k^{1/8} \log(2k^{1/8})} = t^{-o(1)} \quad (7.6)$$

and satisfies, uniformly on the set,

$$\begin{aligned} J &\leq H_p^* + H_r^* \leq 2J, \\ \|(n_{pq} - n_{rq}) - (a_p - a_r)\|_1 + [H_p^* + H_r^* - \|n_{pq} - n_{rq}\|_1]_+ &\leq \delta_t(H_p^* + H_r^*). \end{aligned} \quad (7.7)$$

Every such two-step walk has $p \neq r$.

The band is obtained in [Section 7.3](#), using the bound on large values of H_p^* proved in [Section 7.2](#). Before choosing a common center for each retained pair, we remove pairs on unit circles containing many points. If both later algebraic loci were that circle, the radicands in [Section 8](#) would all be squares.

Lemma 7.3 (Deleting pairs on circles containing many points). *Under (2.11), the probability that p, r lie on a common real unit circle containing at least $t^{9/10}$ points of P is at most*

$$t^{-3/5+o(1)} + t^{-9/5+o(1)} + \frac{2}{\min_{q \in Q} d(q)}. \quad (7.8)$$

This is $o(1/(k^{1/8} \log(2k^{1/8})))$. In particular, removing such two-step walks from the set in [Lemma 7.2](#) preserves the lower probability bound (7.6), after decreasing its constant.

The counting proof of this deletion estimate is in [Section 7.4](#).

Proof of Proposition 7.1. Take the band from [Lemma 7.2](#), and remove the two-step walks excluded by [Lemma 7.3](#). Its probability remains at least $c/(k^{1/8} \log(2k^{1/8})) = t^{-o(1)}$, and all its endpoint pairs are distinct. For any fixed off-diagonal pair (p, r) , its two-step walk probability is

$$\sum_{q: pq, rq \in E} \frac{1}{|E| d(q)} \leq \frac{2}{|E| \min_q d(q)}.$$

Here the sum has at most two terms, since two distinct real points have at most two unit-circle centers. The retained two-step walks therefore contain at least

$$\frac{c|E| \min_q d(q)}{2k^{1/8} \log(2k^{1/8})} \geq t^{2-o(1)}$$

distinct ordered endpoint pairs, using $|E| = t^{2+2a+o(a)}$ and $\min_q d(q) = t^{a+o(a)}$. These pairs form $\mathcal{E}$. Choose one surviving witnessing center for each pair.

For its two jumps $d = z_p - z_q$, $d' = z_r - z_q$, the exact unit equations give $w_p - w_q = 1/d$ and $w_r - w_q = 1/d'$. Hence

$$(z_p - z_r)(w_p - w_r) = (d - d')(1/d - 1/d') = 2 - d/d' - d'/d,$$

proving (7.1). The product formula gives

$$2h(\lambda_{pr}) = \sum_v^* |\log |\lambda_{pr}|_v| = \|n_{pq} - n_{rq}\|_1 \geq (1 - \delta_t)J,$$

so (7.2) holds for all sufficiently large indices.

Finally, put

$$b_{pr}(v) = \log |\lambda_{pr}|_v + a_p(v) - a_r(v).$$

By (7.7), $\|b_{pr}\|_1 \leq 2\delta_t J$, uniformly on all retained edges. On a complete rectangle, the endpoint profiles cancel, leaving

$$\log \left| \frac{\lambda_{pr}\lambda_{p'r'}}{\lambda_{p'r}\lambda_{p'r'}} \right|_v = b_{pr}(v) + b_{p'r'}(v) - b_{p'r}(v) - b_{p'r'}(v).$$

Its weighted absolute sum is at most $8\delta_t J$. Another use of the product formula bounds its height by $4\delta_t J = o(J)$, uniformly over all complete rectangles, proving (7.3). The required circle exclusion holds by Lemma 7.3. $\square$

7.2 The pointwise bound and height tail

The first-moment bound in (7.4) alone does not prevent all exception mass from being carried by extremely rare points. We obtain the needed tail control by applying the product formula separately to each distinct pair of neighbors of a point.

Lemma 7.4 (Pointwise estimate for exceptions on $\mathcal{G}$). *The exception masses and inconsistency integrals satisfy, for every $p \in P$ and distinct $q, q' \in Q$ with $pq, pq' \in E$,*

$$2H_p^* \leq 3(H_q^* + H_{q'}^* + m_{pq} + m_{pq'}) + W_c + W_b + 2 \log 2. \quad (7.9)$$

Proof. At a level in $\mathcal{G}$, let X, U, V be the exception indicators of p, q, q' , and let I, I' be the inconsistency indicators of the two edges. Write P_+ for the positive term on $\mathcal{G}$ in the score for experiment P : it equals two when p is exceptional, both bits are minority bits, and the two neighbors have matching nonshared coordinates; otherwise it is zero. Write $N = |Y_{pq} - Y_{pq'}|$. If $U = V = I = I' = 0$, the two bits are the same function of X : they are both majority bits for $X = 0$, and both minority bits for $X = 1$. In the latter case the nonshared coordinate of both neighbors has the dominant label, so $P_+ = 2$. Consequently, pointwise,

$$P_+ \geq 2X - 2(U + V + I + I'), \quad N \leq U + V + I + I',$$

and hence

$$P_+ - N \geq 2X - 3(U + V + I + I'). \quad (7.10)$$

In particular, a simultaneous exception at p and at a neighbor is charged directly to the neighbor indicator.

For completeness, the upper bound for this score with positive terms only on $\mathcal{G}$ is also pointwise. Put $d = z_p - z_q$, $d' = z_p - z_{q'}$. These jumps are distinct and nonzero. At an absolute-value entry v , sort their thresholds as $b \leq b'$, and set

$$u = -\log|d - d'|_v, \quad \alpha = u - b, \quad T_v = \begin{cases} \alpha + \log 2, & v \text{ archimedean,} \\ \alpha, & v \text{ nonarchimedean.} \end{cases}$$

The triangle inequality makes $T_v \geq 0$. Matching coordinates in a square cell of side length r gives distance at most $\sqrt{2}r$, and matching nonarchimedean labels gives distance at most r . Thus positive charges with both bits one lie within $[b - T_v, b]$, and positive charges with both bits zero lie within $[b', b' + T_v]$.

Only one of these two intervals can contain a positive term from $\mathcal{G}$. Indeed every level where zero is the minority bit precedes every level where one is the minority bit. A positive term from $\mathcal{G}$ with bit zero would require a threshold at least b' , while a positive term from $\mathcal{G}$ with bit one would require a later threshold strictly less than b , which is impossible. The total positive charge from $\mathcal{G}$ at this entry is therefore at most $2T_v$. The all-level negative charge is $b' - b$, because the level interval contains both thresholds. Moreover,

$$-\log|(d - d')(1/d - 1/d')|_v = 2\alpha - (b' - b).$$

Summing over absolute-value entries, the product formula cancels this last expression. The archimedean weights sum to one, so the positive terms from $\mathcal{G}$ minus the all-level negative charge total at most $2 \log 2$.

Integrating (7.10) gives a lower bound for that same score. Outside the levels in $\mathcal{G}$ there are no positive terms, and the negative charge is at most $W_c + W_b$. Comparing the lower and upper bounds proves (7.9). $\square$

Lemma 7.5 (Tail control). *The point exception masses satisfy*

$$\mathbb{E}_\mu[H_p^* \mathbf{1}_{\{H_p^* > k^{1/8}S\}}] = o(S). \quad (7.11)$$

Proof. Choose an absolute C_0 for which each of the two exception expectations in (7.4) is at most C_0S , for all sufficiently large indices. Put

$$T = \{p : H_p^* > k^{1/8}S\}, \quad A_t = \mathbb{E}_\mu[H_p^* \mathbf{1}_T], \quad d_{P,\min} = \min_{p \in P} d(p).$$

On T , the deterministic last three terms of (7.9) have sum at most H_p^* , eventually, since

$$\frac{W_c + W_b + 2 \log 2}{k^{1/8}S} \leq \epsilon + \frac{W_b + 2 \log 2}{k^{1/8}S} = o(1).$$

After absorbing them, we may cap each neighbor height at H_p^* :

$$H_p^* \leq 3(\min(H_p^*, H_q^*) + \min(H_p^*, H_{q'}^*) + m_{pq} + m_{pq'}), \quad p \in T, \quad (7.12)$$

for distinct q, q' . If a cap changes a term, that capped term already equals H_p^* and suffices for this inequality; otherwise the preceding inequality is unchanged.

Sample p with law μ and two neighbors independently and uniformly given p . Restrict the left side to distinct neighbors. The exact conditional probability of distinctness is $1 - 1/d(p)$. On the nonnegative right side we can drop this restriction. Thus

$$(1 - d_{P,\min}^{-1})A_t \leq 6\mathbb{E}_E[\min(H_p^*, H_q^*) \mathbf{1}_T] + C_m, \quad (7.13)$$

where C_m is an absolute upper bound for $6\mathbb{E}_E m_{pq}$. This calculation uses the distinctness probability at each individual point, so it remains valid even when its exception mass is very large.

We estimate the capped expectation by layer cake. For $u \geq 0$ let

$$F_u = \{p : H_p^* > \max(u, k^{1/8}S)\}, \quad G_u = \{q : H_q^* > u\},$$

and write $h_P(u) = \mu(F_u)$, $h_Q(u) = \kappa(G_u)$. The marginal version of the expansion estimate from [Section 2](#), with $\eta = \gamma - 1/2 > 0$, is

$$\mathbb{P}_E(F, G) \leq \sqrt{\mu(F)\kappa(G)} \left[\gamma^{-(1+2\eta)} (\mu(F)\kappa(G))^\eta + \gamma^{-1} \xi_t \right].$$

Because $h_P(u) \leq C_0/k^{1/8}$ and $h_Q(u) \leq 1$, the same multiplier

$$\vartheta_t = \gamma^{-(1+2\eta)} (C_0/k^{1/8})^\eta + \gamma^{-1} \xi_t = o(1)$$

works at every threshold u , including thresholds where the center set has full measure. Tonelli's theorem and Cauchy–Schwarz now give

$$\begin{aligned} \mathbb{E}_E[\min(H_p^*, H_q^*) \mathbf{1}_T] &= \int_0^\infty \mathbb{P}_E(F_u, G_u) \, du \\ &\leq \vartheta_t \left(\int_0^\infty h_P(u) \, du \right)^{1/2} \left(\int_0^\infty h_Q(u) \, du \right)^{1/2} \\ &= \vartheta_t \sqrt{A_t \mathbb{E}_\kappa H_q^*} \leq \vartheta_t \sqrt{C_0 A_t S}. \end{aligned}$$

The first integral equals the unknown tail first moment A_t , including its contribution below the threshold $k^{1/8}S$. Since $d_{P,\min} \rightarrow \infty$, [\(7.13\)](#) implies eventually

$$\frac{1}{2} A_t \leq 6\vartheta_t \sqrt{C_0 A_t S} + C_m \leq \frac{1}{4} A_t + 36C_0 \vartheta_t^2 S + C_m.$$

The last step is Young's inequality. Consequently

$$\frac{A_t}{S} \leq 144C_0 \vartheta_t^2 + \frac{4C_m}{S} \rightarrow 0,$$

as required. □

7.3 Profile errors and selection of the band

The tail bound shows that points with $H_p^* > k^{1/8}S$ contribute only $o(S)$ to $\mathbb{E}_\mu H_p^*$. We now approximate the profile $n_{pq} - n_{rq}$ of negative logarithms of the jump ratio by $a_p - a_r$. Before choosing a band, we remove two-step walks with large relative approximation error and bound the lost expectation of $H_p^* + H_r^*$.

Lemma 7.6 (Profiles of two neighbors of a center). *Under the two-step walk law [\(2.11\)](#), the endpoint profiles a_p, a_r satisfy*

$$\mathbb{E}[\|(n_{pq} - n_{rq}) - (a_p - a_r)\|_1] = o(S), \tag{7.14}$$

$$\mathbb{E}[H_p^* + H_r^* - \|n_{pq} - n_{rq}\|_1]_+ = o(S). \tag{7.15}$$

Here $n_{pq}(v) = -\log|z_p - z_q|_v$, and $[x]_+ = \max(x, 0)$.

Proof. Let I_{pq} be the levelwise inconsistency indicator. At a level in $\mathcal{G}$, write $b \in \{0, 1\}$ for the majority bit. Consistency gives

$$Y_{pq} = b + \sigma(I_p + I_q)$$

unless both endpoints are exceptional or the edge is inconsistent. In all cases the absolute residual in this identity is at most $2(I_p I_q + I_{pq})$. To see the bit rule, $\sigma = 1 - 2b$: zero exceptional endpoints give bit b , and exactly one gives $1 - b = b + \sigma$ on a consistent edge. The excluded cases account for the stated residual. Writing that residual as e_{pq} , subtraction gives

$$Y_{pq} - Y_{rq} = \sigma(I_p - I_r) + e_{pq} - e_{rq}.$$

The center indicator cancels because both edges use the same dominant coordinate and label at this level. At each absolute-value entry,

$$n_{pq}(v) - n_{rq}(v) = \int_{\ell_v}^{u_v} (Y_{pq}(v, s) - Y_{rq}(v, s)) \, ds,$$

since both bits are single thresholds inside the level interval. Thus the norm inside the expectation in (7.14) is bounded by

$$E_1 = \int_{\mathcal{C}} |Y_{pq} - Y_{rq}| + W_b + 2 \int_{\mathcal{G}} (I_p I_q + I_r I_q + I_{pq} + I_{rq}).$$

Its mean is $o(S)$: the first term averages to $C_Q = o(S)$, the joint edge exceptions have integrated expectation $o(S)$, and the remaining bounds are in (7.4).

There is also the exact identity

$$\|n_{pq} - n_{rq}\|_1 = \int |Y_{pq} - Y_{rq}|.$$

On a level in $\mathcal{G}$ with exactly one exceptional point endpoint, a center with its dominant label, and two consistent edges, the bits differ. More explicitly, at every level in $\mathcal{G}$,

$$|Y_{pq} - Y_{rq}| \geq I_p + I_r - 2(I_p I_r + I_p I_q + I_r I_q + I_{pq} + I_{rq}).$$

If none of the terms subtracted is present, this is the preceding observation; if any is present, the right side is nonpositive. After integration, the positive part in (7.15) is bounded by

$$E_2 = 2 \int_{\mathcal{G}} (I_p I_r + I_p I_q + I_r I_q + I_{pq} + I_{rq}).$$

The integrated joint point exceptions on a two-step walk and joint edge exceptions are $o(S)$, by Lemma 5.1. Thus $\mathbb{E}E_2 = o(S)$ as well. $\square$

The next selection makes these average error bounds uniform before choosing a dyadic band. This order matters because the probability of the eventual band tends to zero.

Proof of Lemma 7.2. Set $W = H_p^* + H_r^*$, and let Z be the sum of the norm and positive part on the left of (7.7). By Lemma 7.6, $\mathbb{E}Z = o(S)$. By Proposition 6.5, choose $c_0 > 0$, independent of t along the fixed sequence, such that $\mathbb{E}W \geq 4c_0 S$ eventually. Choose $\delta_t \rightarrow 0$ sufficiently slowly that $\mathbb{E}Z/\delta_t = o(S)$. Before selecting any band, discard two-step walks with $Z > \delta_t W$. Their lost weight is at most

$$\mathbb{E}[W \mathbf{1}_{\{Z > \delta_t W\}}] \leq \frac{\mathbb{E}Z}{\delta_t} = o(S).$$

Also discard $W < c_0 S$, losing at most $c_0 S$. If $W > 2k^{1/8} S$, its larger summand exceeds $k^{1/8} S$ and is at least $W/2$. Thus [Lemma 7.5](#) gives

$$\mathbb{E}[W \mathbf{1}_{\{W > 2k^{1/8} S\}}] \leq 2\mathbb{E}[H_p^* \mathbf{1}_{\{H_p^* > k^{1/8} S\}}] + 2\mathbb{E}[H_r^* \mathbf{1}_{\{H_r^* > k^{1/8} S\}}] = o(S).$$

The surviving weighted mass is at least $2c_0 S$, whereas every surviving two-step walk has $W \leq 2k^{1/8} S$. Its probability is therefore at least $c_0/k^{1/8}$.

Partition $[c_0 S, 2k^{1/8} S]$ into at most $1 + \lceil \log_2(2k^{1/8}/c_0) \rceil = O(\log(2k^{1/8}))$ dyadic bands, using half-open endpoints except at the final endpoint. One band has the probability in [\(7.6\)](#); if J is its lower endpoint, then $J \geq c_0 S \rightarrow \infty$, and [\(7.7\)](#) holds. The parameter hierarchy gives $k = t^{o(1)}$, so the lower bound in [\(7.6\)](#) is indeed $t^{-o(1)}$. Finally, if $p = r$, then $\|n_{pq} - n_{rq}\|_1 = 0$, so [\(7.7\)](#) would imply $W \leq \delta_t W$. This contradicts $W \geq J > 0$ once $\delta_t < 1$. $\square$

7.4 Deleting pairs on common unit circles

The algebraic argument will require that a retained pair does not lie on a common unit circle containing many points of P . This exclusion must concern all real unit circles, including those whose centers do not belong to Q .

Proof of [Lemma 7.3](#). Let $\mathcal{K}_t$ be the family of real unit circles containing at least $t^{9/10}$ points of P , and write M for its size. Every pair of distinct real points lies on at most two unit circles. Counting pairs on circles in $\mathcal{K}_t$ therefore gives

$$M \binom{\lceil t^{9/10} \rceil}{2} \leq 2 \binom{|P|}{2}, \quad M \leq t^{1/5+o(1)},$$

using $|P| = t^{1+o(1)}$. The set P_{mult} of points on at least two circles in $\mathcal{K}_t$ has size at most $2 \binom{M}{2} \leq t^{2/5+o(1)}$, because distinct circles intersect in at most two points. The degree bounds give

$$\max_p \mu(p) \leq t^{-1+o(1)}, \quad \max_q \kappa(q) \leq t^{-2+o(1)}.$$

Consequently two-step walks hitting P_{mult} at either point endpoint have probability at most $t^{-3/5+o(1)}$. Centers in Q whose own unit circle belongs to $\mathcal{K}_t$ number at most M ; two-step walks with such a center have probability at most $t^{-9/5+o(1)}$.

To bound the remaining bad two-step walks, fix $p \notin P_{\text{mult}}$ and a center q whose own circle does not belong to $\mathcal{K}_t$, in the original sampling law. The point p belongs to at most one circle in $\mathcal{K}_t$. If there is one, it differs from the unit circle centered at q , since the latter does not belong to $\mathcal{K}_t$. Their intersection has at most two points. In this original law, r is a uniform neighbor of q , so the probability that it shares a circle in $\mathcal{K}_t$ with p is at most $2/d(q)$, even before requiring $r \notin P_{\text{mult}}$. This proves [\(7.8\)](#).

For the comparison with the selected band, recall $k = t^{o(1)}$ and the explicit requirement $k^{20}/\min_q d(q) \rightarrow 0$ in [Section 5.1](#). The first two terms of [\(7.8\)](#), multiplied by $k^{1/8} \log(2k^{1/8})$, tend to zero by a fixed power of t . For the third term,

$$\frac{k^{1/8} \log(2k^{1/8})}{\min_q d(q)} = \frac{k^{1/8} \log(2k^{1/8})}{k^{20}} \frac{k^{20}}{\min_q d(q)} \rightarrow 0.$$

This establishes the required relative loss bound. $\square$

8 The algebraic obstruction

The preceding section produced many pairs whose edge ratios have large height, although their alternating products around rectangles have negligible height. We now show that this is impossible. The argument uses a fixed complete bipartite pattern. After passing to an ultraproduct, we will choose its positions so that no nonempty product of certain radicands is a square over the field containing the fixed coefficients. Kummer theory then permits one square root to change sign while the others remain fixed. The equations obtained by differentiating the distance identities and the four-edge quotients will contradict this sign change.

Here and below a unit circle means a circle of radius exactly one in the original real plane. For a point $p = (x_p, y_p)$, retain the coordinates $z_p = x_p + iy_p$ and $w_p = x_p - iy_p$.

Proposition 8.1 (Incompatibility of the ratio and quotient height bounds). *There is no sequence with the following properties. At index n , let $t_n \rightarrow \infty$, let $J_n \rightarrow \infty$, and let $P_n \subset \mathbb{R}^2$ be a finite set of algebraic points with $|P_n| \leq t_n^{1+o(1)}$. Let $G_n \subset P_n \times P_n$ be a bipartite graph, using two copies of P_n , with at least $t_n^{2-o(1)}$ edges. To each edge (p, r) assign a nonzero algebraic number λ_{pr} . Suppose that, for an absolute $c > 0$,*

$$(z_p - z_r)(w_p - w_r) = 2 - \lambda_{pr} - \lambda_{pr}^{-1}, \quad (8.1)$$

$$h(\lambda_{pr}) \geq cJ_n, \quad (8.2)$$

$$h\left(\frac{\lambda_{pr}\lambda_{p'r'}}{\lambda_{pr'}\lambda_{p'r}}\right) = o(J_n) \quad (8.3)$$

uniformly over edges and complete rectangles, respectively. Finally, suppose that no edge has both endpoints on a unit circle containing at least $t_n^{9/10}$ points of P_n . Here h is the absolute logarithmic height from [Lemma 4.2](#).

To prepare the contradiction in [Section 8.5](#), assume the proposition's data exist and suppress the index when this causes no ambiguity. At each index choose a finite normal number field F_n containing all coordinates, ratios and i . By [Lemma 4.2](#), this choice does not affect the height. In the application from [Proposition 7.1](#), take the field $\mathcal{N}$ already used there: every ratio is a quotient of coordinate differences.

8.1 Negligible height as a field of constants

Fix a nonprincipal ultrafilter $\mathbf{u}$ on $\mathbb{N}$: a collection of subsets closed under finite intersection and supersets, containing the cofinite sets and exactly one part of every finite partition. Such a collection is obtained by extending the cofinite filter to a maximal proper filter. Call a set of indices *large* if it belongs to $\mathbf{u}$. Two sequences are identified when they agree on a large set. Write

$$F = \prod_{n \rightarrow \mathbf{u}} F_n, \quad \mathcal{U} = \prod_{n \rightarrow \mathbf{u}} \mathbb{C},$$

where F_n is the number field just chosen. These are characteristic-zero fields, with $F \subset \mathcal{U}$: the inverse of a nonzero class is defined coordinatewise on its large set of nonzero entries. The field $\mathcal{U}$ is algebraically closed. Indeed, a polynomial of fixed positive degree has that degree on a large set, and choosing one complex root at each such index supplies a root in $\mathcal{U}$. Complex conjugation also acts coordinatewise on $\mathcal{U}$.

We use limits only along $\mathbf{u}$. Thus $a_n = o(J_n)$ means that $\{n : |a_n| < \varepsilon J_n\}$ is large for every $\varepsilon > 0$. Edge-count bounds of the form $t_n^{2-o(1)}$ are also understood along $\mathbf{u}$. Define

$$k_0 = \{[u_n] \in F : h(u_n) = o(J_n)\}.$$

In height estimates for $u = [u_n] \in F$, the expression $h(u)$ denotes the sequence $(h(u_n))_n$, taken up to agreement on a large set; comparisons and the notation $o(J_n)$ are interpreted along $\mathbf{u}$ as above. This convention does not assert that $h(u)$ is a real number.

Lemma 8.2. *The elements of F with height $o(J_n)$ form a subfield k_0 that is relatively algebraically closed in F : every element of F algebraic over k_0 already lies in k_0 .*

Proof. The height inequalities

$$h(uv) \leq h(u) + h(v), \quad h(u+v) \leq h(u) + h(v) + \log 2, \quad h(u^{-1}) = h(u) \quad (u \neq 0)$$

prove the field assertion, since $J_n \rightarrow \infty$. These statements do not depend on representatives, and $h(0) = 0$.

Suppose that $u \in F$ satisfies $u^m + a_{m-1}u^{m-1} + \cdots + a_0 = 0$ with $a_r \in k_0$ and fixed $m \geq 1$. At each finite index where the equation holds, the nonarchimedean triangle inequality gives $|u|_v \leq \max(1, \max_r |a_r|_v)$. At infinity the same estimate holds with the additional factor m : if $|u|_v > 1$, divide the equation by $|u|_v^{m-1}$. Since the total weight of the archimedean entries is one,

$$h(u) \leq \sum_{r=0}^{m-1} h(a_r) + \log m.$$

This estimate holds on a large set of indices. Its right side is $o(J_n)$, proving the claim. $\square$

In particular, any edge ratio selected from the graphs is transcendental over k_0 , by (8.2). The alternating products on complete rectangles in (8.3), in contrast, belong to $k_0^\times$. We next select finitely many edges whose positions retain enough algebraic independence to exploit this distinction.

8.2 A complete grid with generic positions

We need points whose coordinates satisfy no extra polynomial relations beyond those of their loci. Here a locus is either the (z, w) -plane, of dimension two, or an irreducible plane curve, of dimension one. The next lemma chooses both the loci and the points. Throughout this construction, adjoining a point means adjoining both coordinates. Over a field containing i , the (x, y) and (z, w) coordinates generate the same field, by the invertible change $z = x + iy$, $w = x - iy$.

Lemma 8.3. *There are two sequences of loci, each either a plane or an irreducible curve of bounded degree, to which the graphs can be restricted while retaining at least $t_n^{2-o(1)}$ edges. Their limiting loci $V_1, V_2 \subset \mathcal{U}^2$, of dimensions d_1, d_2 , have the following properties. Every curve is irreducible over $\mathcal{U}$, is invariant under $(z, w) \mapsto (\bar{w}, \bar{z})$, and has both coordinate projections nonconstant. The two loci are not both the curve $(z - u)(w - \bar{u}) = 1$ for any $u \in \mathcal{U}$. Let $B_0 \subset \mathcal{U}$ be the countable field generated over $\mathbb{Q}$ by i and the coefficients of their curve equations, each normalized by setting one nonzero coefficient to one.*

For every fixed $N, j \geq 1$, the restricted graphs contain a complete N -by- j grid on a large set of indices whose limiting vertices $p_1, \dots, p_N \in V_1$ and $r_1, \dots, r_j \in V_2$ satisfy the following

independence assertion. Select both coordinates of each vertex on a plane locus and its z -coordinate on a curve locus. The resulting $Nd_1 + jd_2$ coordinates are algebraically independent over B_0 . The remaining coordinate of each point on a curve is algebraic over its selected coordinate and B_0 . In particular,

$$\text{trdeg}_{B_0(r_1, \dots, r_j)} B_0(r_1, \dots, r_j, p_1, \dots, p_N) = Nd_1. \quad (8.4)$$

All vertex coordinates and assigned edge ratios of the grid belong to F ; the coefficient field B_0 need not lie in F .

For a plane locus V over a field $\mathcal{B}$, define its function field to be $\mathcal{B}(V) = \mathcal{B}(z, w)$. For a curve with irreducible equation f , define it to be the fraction field of $\mathcal{B}[z, w]/(f)$. For the loci in [Lemma 8.3](#) and a field $B_0 \subseteq \mathcal{B} \subseteq \mathcal{U}$, a point on V is called generic over $\mathcal{B}$ if its coordinates have transcendence degree $\dim V$ over that field.

Evaluation at a point $x \in V$ generic over $B_0 \subseteq \mathcal{B} \subseteq \mathcal{U}$ identifies $\mathcal{B}(V)$ with $\mathcal{B}(x)$. For a curve, the z -coordinate is transcendental and the curve equation accounts for every polynomial relation, so no nonzero function-field denominator vanishes at x . The grid's joint independence will let us apply this identification after adjoining derivative values: one left point still has transcendence degree d_1 . It also ensures that the zero chosen for one radicand is not a zero of another. The full grid construction and evaluation proof are both in [Section 8.6](#).

8.3 Differentiation and the surviving generic point

Take $N = 20$ and $j = 5$ in [Lemma 8.3](#), and write $\lambda_{i\ell}$ for the edge ratios. A k_0 -derivation of F is an additive map $F \rightarrow F$ satisfying the product rule and vanishing on k_0 .

Lemma 8.4. *The grid ratios admit a common k_0 -derivation $\partial : F \rightarrow F$ with $\partial\lambda_{i\ell} \neq 0$ for every $1 \leq i \leq N$ and $1 \leq \ell \leq j$.*

Proof. Each $\lambda_{i\ell}$ is transcendental over k_0 by [Lemma 8.2](#). Put it in a transcendence basis of F/k_0 , prescribe its derivative to be one and those of the other basis elements to be zero, and extend by the quotient rule to the rational function field. Since the characteristic is zero, this derivation extends uniquely over the algebraic extension to F [[6](#), Theorems 1.5 and 4.1]. More explicitly, if $g(X)$ is the minimal polynomial of an algebraic element α over an intermediate field carrying the derivation, let g^∂ mean differentiation of its coefficients. For every polynomial f over that intermediate field, the extension satisfies

$$\partial\alpha = -\frac{g^\partial(\alpha)}{g'(\alpha)}, \quad \partial(f(\alpha)) = f^\partial(\alpha) - f'(\alpha)\frac{g^\partial(\alpha)}{g'(\alpha)}.$$

Separability gives $g'(\alpha) \neq 0$; this formula preserves the intermediate field with α adjoined and is independent of the polynomial representative [[6](#), Theorem B.8 and Corollary B.10]. For the full algebraic extension, take these extensions on finite subextensions. Their restrictions agree on overlaps by uniqueness, so their union is an F -valued derivation.

For each ratio, this construction gives a derivation nonzero on that ratio. To obtain one derivation working for all ratios, take an F -linear combination of these finitely many derivations. Such combinations still satisfy the product rule and vanish on k_0 . For each ratio, vanishing of its derivative is one nonzero linear equation in the combination coefficients. Avoiding the zeros of the product of these linear forms therefore gives the required common derivation. This product is a nonzero polynomial over the infinite field F , so such coefficients exist. $\square$

The rectangle condition gives

$$\frac{\lambda_{i\ell}\lambda_{11}}{\lambda_{i1}\lambda_{1\ell}} \in k_0^\times. \quad (8.5)$$

Repeated rows or columns would violate the algebraic independence in [Lemma 8.3](#); all vertices on either side are distinct.

Put

$$A_i = \frac{\partial \lambda_{i1}}{\lambda_{i1}}, \quad A'_\ell = \frac{\partial \lambda_{11}}{\lambda_{11}} - \frac{\partial \lambda_{1\ell}}{\lambda_{1\ell}}, \quad H_i = (\partial x_{p_i}, \partial y_{p_i}), \quad U_\ell = (\partial x_{r_\ell}, \partial y_{r_\ell}).$$

The derivation kills i , which is algebraic over $\mathbb{Q} \subset k_0$. Define

$$L_\ell(z, w) = (z - z_{r_\ell})(w - w_{r_\ell}), \quad R_{i\ell} = \lambda_{i\ell}^{-1} - \lambda_{i\ell}.$$

For a point x , write $L_\ell(x) = L_\ell(z_x, w_x)$. Taking the logarithmic derivative of (8.5) gives

$$\frac{\partial \lambda_{i\ell}}{\lambda_{i\ell}} = \frac{\partial \lambda_{i1}}{\lambda_{i1}} + \frac{\partial \lambda_{1\ell}}{\lambda_{1\ell}} - \frac{\partial \lambda_{11}}{\lambda_{11}} = A_i - A'_\ell.$$

The quotient rule also gives $\partial(2 - \lambda - \lambda^{-1}) = (\lambda^{-1} - \lambda)\partial\lambda/\lambda$. In the ordinary coordinates, $L_\ell(p_i) = (x_{p_i} - x_{r_\ell})^2 + (y_{p_i} - y_{r_\ell})^2$, so the product rule on both position variables now yields

$$2(p_i - r_\ell) \cdot (H_i - U_\ell) = R_{i\ell}(A_i - A'_\ell), \quad (8.6)$$

$$R_{i\ell}^2 = L_\ell(p_i)(L_\ell(p_i) - 4), \quad (8.7)$$

$$A_i - A'_\ell \neq 0. \quad (8.8)$$

Here the dot denotes the bilinear expression $xu + yv$ in the ordinary two coordinates, with no conjugation. The second identity follows by squaring $\lambda^{-1} - \lambda$, and the third by [Lemma 8.4](#).

We now form a coefficient field containing the right-hand positions r_ℓ , their coordinate derivatives U_ℓ , and the scalars A'_ℓ from the logarithmic derivatives of the ratios. Let K_0 be the subfield of $\mathcal{U}$ consisting of all elements algebraic over

$$B_0(r_1, \dots, r_j, U_1, \dots, U_j, A'_1, \dots, A'_j).$$

This field is algebraically closed: roots in $\mathcal{U}$ of polynomials over it are still algebraic over the displayed field. We do not adjoin k_0 to K_0 .

Lemma 8.5. *At least one of the selected left vertices $p = p_i$ is generic on V_1 over the specified field K_0 . Evaluation at this point identifies $K_0(V_1)$ with $K_0(p)$.*

Proof. Over $B_0(r_1, \dots, r_j)$, all the left positions jointly have transcendence degree Nd_1 , by (8.4). Put $B = B_0(r_1, \dots, r_j)$, and let $\mathbf{d}$ be the tuple consisting of the two coordinates of each U_ℓ and each scalar A'_ℓ , containing at most $3j$ elements. The tower rule for transcendence degree gives

$$\begin{aligned} \text{trdeg}_{B(\mathbf{d})} B(\mathbf{d}, p_1, \dots, p_N) &= \text{trdeg}_B B(\mathbf{d}, p_1, \dots, p_N) - \text{trdeg}_B B(\mathbf{d}) \\ &\geq Nd_1 - 3j. \end{aligned}$$

Passing from $B(\mathbf{d})$ to its algebraic closure K_0 does not change this transcendence degree: the corresponding compositum with the positions is also an algebraic extension. Therefore

$$\text{trdeg}_{K_0} K_0(p_1, \dots, p_N) \geq Nd_1 - 3j. \quad (8.9)$$

If no individual point had degree d_1 , subadditivity would bound the left side by $N(d_1 - 1)$. This contradicts $N > 3j$. The function-field assertion is the evaluation statement following [Lemma 8.3](#), with $\mathcal{B} = K_0$; its proof is in [Section 8.6](#). $\square$

The coordinate derivatives U_ℓ and scalars A'_ℓ may depend arbitrarily on the left positions. The transcendence-degree bound above allows those dependencies. No derivative of a locus equation has been used, and none of its coefficients is being asserted constant for the original derivation.

We have reduced the proof to one generic point p . Write $M = K_0(V_1) = K_0(p)$, $R_\ell = R_{i_\ell}$, and

$$D_\ell = L_\ell(L_\ell - 4) \in M.$$

Here each L_ℓ is regarded as an element of M through the evaluation identification just established. The next step proves that the radicands are independent modulo squares over this field.

8.4 No nonempty product of the radicands is a square

Independence in $M^\times/(M^\times)^2$ means linear independence over $\mathbf{F}_2$: no nonempty product of the indicated classes is the identity.

An integer-valued valuation of M is a map $v : M^\times \rightarrow \mathbb{Z}$ with $v(ab) = v(a) + v(b)$ and $v(a+b) \geq \min\{v(a), v(b)\}$, where $v(0) = +\infty$.

Lemma 8.6. *The radicands $D_1, \dots, D_j$ are nonzero and have independent classes in $M^\times/(M^\times)^2$, for the field $M = K_0(V_1) = K_0(p)$ fixed above.*

The proof in [Section 8.7](#) constructs, for each radicand, a valuation that takes value one on it and zero on every other radicand. These valuations are constructed over the actual enlarged field K_0 . We can now finish the obstruction before giving the generic-grid and valuation constructions.

8.5 The sign-change contradiction

We use the quadratic case of Kummer theory over the characteristic-zero field M . If $D_1, \dots, D_j \in M^\times$ have independent classes modulo squares and $R_\ell^2 = D_\ell$, then

$$[M(R_1, \dots, R_j) : M] = 2^j,$$

and every choice of signs of the R_ℓ is induced by an automorphism fixing M [[15](#), Theorem 5.30 and Remark 5.32]. The hypotheses hold because M contains both square roots of unity and has characteristic different from two. Independence of the square classes is required in this same field M , as proved in [Lemma 8.6](#).

Apply this fact with the radicands of [Lemma 8.6](#), and let $\mathcal{M} = M(R_1, \dots, R_j) \subset \mathcal{U}$, using the actual roots from (8.7). Write $H = H_i$, $A = A_i$ for the surviving point. We have the equations

$$2(p - r_\ell) \cdot H - R_\ell A = 2(p - r_\ell) \cdot U_\ell - R_\ell A'_\ell, \quad 1 \leq \ell \leq j. \quad (8.10)$$

Their right-hand positions, velocities and scalars belong to K_0 .

First suppose the three rows $p - r_1, p - r_2, p - r_3$ have rank two. Put $v_\ell = p - r_\ell \in M^2$. The coefficient matrix of the first three equations in the unknowns (H_x, H_y, A) has rows $(2v_{\ell,x}, 2v_{\ell,y}, -R_\ell)$, and its determinant is

$$-4(R_1 \det(v_2, v_3) - R_2 \det(v_1, v_3) + R_3 \det(v_1, v_2)).$$

At least one of these minors is nonzero because the position rows have rank two. For any relation $\sum_{\ell=1}^3 c_\ell R_\ell = 0$, $c_\ell \in M$, changing only the sign of R_m and subtracting gives $2c_m R_m = 0$. Hence all c_m vanish, so the single radicals are linearly independent over M and the determinant is nonzero. Cramer's rule places the existing solution H_x, H_y, A in $M(R_1, R_2, R_3)$. The automorphism of $\mathcal{M}$

that changes only the sign of R_4 fixes these unknowns. Every coefficient of the fourth equation other than R_4 is then fixed: the positions, U_4 , and A'_4 lie in M , and the unknowns lie in $M(R_1, R_2, R_3)$. Subtracting the transformed equation from the original gives $2R_4(A - A'_4) = 0$, up to an overall sign. Since $R_4 \neq 0$, this gives $A = A'_4$, contradicting (8.8).

We verify when the rank condition can fail. If V_2 is a plane or a curve other than a line, the three independent generic anchors are noncollinear. Indeed the first two are distinct, and a generic third point can lie on their line only if its locus is contained in that line, by generic evaluation. Three noncollinear anchors make the three position rows have rank two for every p . If V_2 is a line, its first two anchors are distinct. The rows have rank two unless p belongs to that line. Since p is generic on V_1 over K_0 , this last possibility requires $V_1 = V_2$ to be the same line.

In that remaining case choose a nonzero direction vector $v \in K_0^2$ for the line, and write $p - r_\ell = s_\ell v$, with $s_\ell \in M^\times$. Nonzero row factors follow from distinctness of the points. Only the scalar $h = v \cdot H$ enters (8.10). The first two equations in h, A have determinant

$$2(s_2 R_1 - s_1 R_2) \neq 0,$$

again by independence of the single radicals. Thus $h, A \in M(R_1, R_2)$. Changing only the sign of R_3 in the third equation gives $A = A'_3$, the same contradiction. The line has nonconstant projections, as established above; writing the line as $w = ez + e'$ with $e \neq 0$, as in (8.14), its direction has nonzero squared length proportional to e .

This proves Proposition 8.1. All sign changes were applied inside the finite radical field $\mathcal{M}$, after Cramer's rule located the unknowns there. No automorphism of the whole field F , or compatibility of an automorphism with ∂ , is required.

8.6 Proof of the generic-grid statement

We choose loci of minimum total dimension among those retaining $t_n^{2-o(1)}$ edges. This choice bounds the number of vertex tuples satisfying any fixed nonzero polynomial equation by t_n^{N+j-b} for some $b > 0$, while there are at least $t_n^{N+j-o(1)}$ complete grids. A grid therefore remains after excluding the zeros of any fixed finite list of polynomials. Choosing grids that avoid successively longer lists makes every nonzero polynomial over B_0 nonzero on the selected coordinates in the ultraproduct.

Proof of Lemma 8.3. We first choose the loci and control additional polynomial relations; then we select the grid by counting.

We use two standard algebraic facts. First, for every field K , $K[Z, W]$ is a unique factorization domain [24, Lemma 10.120.10]. A polynomial in W with coefficients in $K[Z]$ is primitive if its coefficients have no common nonconstant factor. Gauss's lemma [16, Proposition 1.27 and Lemma 1.28] then identifies factorization of primitive polynomials in $K[Z][W]$ with factorization over $K(Z)$: clear denominators and cancel common prime factors of the coefficients. In particular, an irreducible polynomial in $K[Z, W]$ of positive W -degree remains irreducible in $K(Z)[W]$. Two primitive polynomials that are scalar multiples over $K(Z)$ differ by an element of $K^\times$.

Second, two distinct irreducible plane curves of degrees at most D , over an algebraically closed field, have at most D^2 common points. Indeed their projective closures have the same degrees and no common component, so Bézout's theorem bounds the sum of their intersection multiplicities by D^2 [16, Theorem 6.37]. Every affine intersection contributes at least one. We will apply this both over $\mathbb{C}$ at finite indices and over the algebraically closed field $\mathcal{U}$.

By a locus at an index we mean either the whole (z, w) -plane or the zero set of a nonconstant irreducible polynomial in $\mathbb{C}[z, w]$ whose degree has a fixed bound along the sequence. Its dimension is two or one, respectively. Among pairs of such locus sequences retaining at least $t_n^{2-o(1)}$ graph edges, choose one of minimal total dimension. A minimum exists since the pair of planes is available and the possible total dimensions are 2, 3, 4. Finite partitions let us fix all degrees and normalize one nonzero coefficient of each curve equation on a large set. Denote the limiting loci in $\mathcal{U}^2$ by V_1, V_2 , their dimensions by d_1, d_2 , and the field generated over $\mathbb{Q}$ by their normalized equation coefficients and i by B_0 . There are finitely many such coefficients, so B_0 is countable. It is a subfield of $\mathcal{U}$, and need not be a subfield of F .

Each limiting curve equation is irreducible over $\mathcal{U}$. Otherwise a factorization into positive-degree factors would give, from finitely many coefficient sequences, a factorization at a large set of indices. The degrees of its factors are bounded by the degree of the equation, so there is no unbounded factorization to consider.

Each finite curve contains $t_n^{1-o(1)}$ retained real points: divide the $t_n^{2-o(1)}$ edge count by the upper bound for the other vertex set. Its image under $(z, w) \mapsto (\bar{w}, \bar{z})$ shares all these points. The bounded intersection fact forces the two curves to be equal on a large set. The limiting curve has the same symmetry. Neither coordinate projection is constant, since a curve with constant z would be a line $z = c$, whose symmetric image $w = \bar{c}$ is a different line. Thus every curve equation has positive degree in both variables.

The loci cannot both be the same unit circle. To check the exact meaning of this assertion, suppose their limiting equations, after normalization, were

$$(z - u)(w - v) = 1. \quad (8.11)$$

Equality of its finitely many coefficients holds at a large set of finite indices. The symmetry gives $v = \bar{u}$, again an exact coordinatewise identity on a large set. Hence the finite loci there are the same real unit circle. Both vertex sets on it have size $t_n^{1-o(1)} > t_n^{9/10}$, contradicting the deletion hypothesis in [Proposition 8.1](#). This reasoning concerns equality in the ultraproduct; ordinary convergence of a radius to one would not imply (8.11).

Restrict the graph to these loci and prune it to minimum degree $t_n^{1-o(1)}$ on both sides. Explicitly, if its initial edge and total vertex counts are e_n, v_n , repeatedly remove vertices of degree less than $e_n/(2v_n)$. Fewer than $e_n/2$ edges are removed. The remaining graph still has at least $t_n^{2-o(1)}$ edges, both vertex sets have size $t_n^{1+o(1)}$, and the asserted minimum degree holds.

First, on a side whose locus is the plane, for each fixed degree bound D there is a $b_D > 0$ such that every nonzero polynomial of degree at most D , even with coefficients depending on n , vanishes at at most $t_n^{1-b_D}$ pruned vertices on a large set of indices. If this failed, choose at each index a polynomial maximizing its number of zeros among the finite vertex set. A maximum is attained because there are only finitely many subsets of that set. Failure of every fixed saving means that its zero count is $t_n^{1-o(1)}$ along $\mathbf{u}$. One of at most D irreducible factors then contains $t_n^{1-o(1)}$ vertices. Their minimum degree supplies $t_n^{2-o(1)}$ edges after restricting this side to a bounded-degree curve, contradicting minimality. This proves the uniform assertion. On a curve side, a nonzero polynomial of degree D in the selected z -coordinate has at most D zeros among the vertices, since real points have distinct $z = x + iy$.

Let the two pruned set sizes be m, n' and their edge density be δ . Independent uniform vertices $p_1, \dots, p_N$ have a common-neighbor fraction X on the other side. If $d(r)$ is the degree of a right

vertex, then

$$\mathbb{E}X = \frac{1}{n'} \sum_r \left(\frac{d(r)}{m} \right)^N \geq \left(\frac{1}{n'} \sum_r \frac{d(r)}{m} \right)^N = \delta^N.$$

Conditional on the left tuple, $(n'X)^j$ ordered right tuples complete it. A second application of Jensen, to X^j , therefore gives the number of ordered complete grids, initially allowing repetitions, is

$$m^N (n')^j \mathbb{E}X^j \geq m^N (n')^j \delta^{Nj} = t_n^{N+j-o(1)}. \quad (8.12)$$

Consider a fixed nonzero polynomial in the selected coordinates with coefficients in B_0 . Choose representatives of its coefficients. Its vanishing among all ordered vertex tuples, before requiring edges, has a power saving:

$$\#\{\text{tuples at which the test vanishes}\} \leq t_n^{N+j-b} \quad (8.13)$$

on a large set, for some $b > 0$ depending on the test. Here is an induction that also handles moving coefficients. View the polynomial as a polynomial in the last point block, and select one nonzero coefficient polynomial in the earlier blocks. The earlier tuples annihilating this coefficient have a power saving by induction. At every other earlier tuple, the remaining polynomial is nonzero, with a fixed degree bound. The uniform plane estimate, or the bounded root count for a curve block, bounds its zeros. Multiply by the upper bounds $t_n^{1+o(1)}$ for the other block sizes and slightly decrease the saving. The starting case is exactly the one-block estimates. The finitely many coefficient identities and nonzero denominators involved hold on a large set.

There are countably many nonzero polynomial tests over B_0 . Enumerate them as $T_1, T_2, \dots$. For each fixed s , (8.12) and (8.13), with a finite union bound, show that on a large set some grid avoids all the first s tests. At index n , choose the largest feasible prefix length $s_n \leq n$, and choose a grid avoiding that prefix. Include the validity of all representatives and denominators of a prefix in its feasibility requirement. If no prefix is feasible, use arbitrary number-field coordinates at that index. For every fixed s , the set where $n \geq s$ and the first s tests are feasible is large; hence $s_n \geq s$ on a large set. Each fixed polynomial test is therefore nonzero in the limiting grid. Only finite intersections of large sets were used.

For a curve equation $f(z, w) = 0$, its positive w -degree and the transcendence of z show that w is algebraic over $B_0(z)$. This proves all the transcendence assertions. The vertices and the assigned ratios at each chosen finite grid lie in F_n , so their classes lie in F . $\square$

Evaluation at a generic point. If $B_0 \subset \mathcal{B} \subset \mathcal{U}$ and a point on a locus V has transcendence degree $\dim V$ over $\mathcal{B}$, then evaluation at that point embeds $\mathcal{B}(V)$ into $\mathcal{U}$. For the plane this is algebraic independence. For a curve, its equation f is irreducible even over $\mathcal{U}$, hence over $\mathcal{B}$. Its coefficients when viewed as a polynomial in w have no common factor of positive degree in z . They cannot all vanish at an abscissa in $\mathcal{U}$: a common root would give a linear common factor over $\mathcal{U}$, contradicting irreducibility. If z were algebraic over $\mathcal{B}$, the nonzero specialized equation would make w algebraic as well. Thus z is transcendental. Gauss's lemma makes f irreducible over $\mathcal{B}(z)$, so every polynomial relation at the point is a multiple of f . Every nonzero denominator in the function field consequently evaluates to a nonzero element. The same argument applies to a point after adjoining the other grid points to $\mathcal{B}$, using Lemma 8.3. $\square$

8.7 A separate odd valuation for each radicand

We must prove that no nonempty product of the radicands is a square in $M = K_0(V_1)$. For each radicand, we construct an order of vanishing that equals one on that radicand and zero on all the others. Any product containing it then has odd order and cannot be a square. The independence of the selected coordinates over B_0 ensures that the chosen zero for one radicand is not a zero of another.

Proof of Lemma 8.6. For each ℓ we construct a valuation v_ℓ satisfying

$$v_\ell(D_\ell) = 1, \quad v_\ell(D_m) = 0 \quad (m \neq \ell).$$

These finite values also prove that every radicand is nonzero. A square has even valuation, so these valuations will imply the claimed independence. They count orders of zeros at points or along irreducible polynomials, directly over K_0 in each possible locus case. Before constructing the valuations over K_0 , we use algebraic independence of the selected coordinates over B_0 to show that L_m and $L_m - 4$ do not vanish at the zero chosen for D_ℓ , whenever $m \neq \ell$.

The left locus is the plane. The polynomial

$$Q_\ell = (z - z_{r_\ell})(w - w_{r_\ell}) - 4$$

is irreducible in $K_0[z, w]$. It is primitive as a polynomial in w , because its leading coefficient $z - z_{r_\ell}$ is coprime to its constant coefficient, and is linear over $K_0(z)$. The polynomials Q_ℓ are distinct, as comparison of the coefficients of z and w recovers their centers. None divides any $L_m = (z - z_{r_m})(w - w_{r_m})$. The exponent of Q_ℓ in a rational function is therefore the required valuation.

The left locus is a curve, and either the loci differ or its z -projection has degree at least two. Let $f(z, w)$ be its irreducible equation, write $f_w = \partial f / \partial w$, and set $a = z_{r_\ell}$. The coordinate a is transcendental over B_0 . Consequently the leading coefficient of f in w does not vanish at a , so specialization retains its w -degree. Characteristic zero and irreducibility make f and f_w coprime in $B_0(z)[w]$. Clearing denominators in a Bézout identity in that polynomial ring gives

$$A(z, w)f(z, w) + B(z, w)f_w(z, w) = c(z), \quad A, B \in B_0[z, w], \quad 0 \neq c \in B_0[z].$$

Since $c(a) \neq 0$, every root b of $f(a, w)$ has $f_w(a, b) \neq 0$. Thus all roots in this fiber are simple.

If $V_2 \neq V_1$, the anchor r_ℓ is not on V_1 . Indeed it is generic on V_2 , which is either the plane or a different irreducible curve. Choose any root b of $f(a, w)$. Then $b \neq w_{r_\ell}$. If $V_2 = V_1$ and the projection degree is at least two, choose a root other than w_{r_ℓ} , again obtaining $b \neq w_{r_\ell}$. In both cases

$$s_\ell = (a, b) \in V_1, \quad f_w(a, b) \neq 0, \quad b \neq w_{r_\ell}.$$

The ordinate b is algebraic over $B_0(a)$. Since K_0 contains a and is algebraically closed, $b \in K_0$.

For $m \neq \ell$, original joint genericity implies that r_m remains generic on V_2 over $B_0(s_\ell)$. Since b is algebraic over $B_0(a)$, adjoining $s_\ell = (a, b)$ instead of a does not reduce the transcendence degree of r_m . Neither coordinate of r_m equals the corresponding coordinate of s_ℓ : both projections of a curve are nonconstant, and the assertion for a plane is immediate. Thus $L_m(s_\ell) \neq 0$. Also $L_m(s_\ell) \neq 4$. Otherwise generic evaluation would force V_2 to lie in

$$(z - a)(w - b) = 4.$$

The plane cannot do so. If V_2 is a curve, this irreducible circle equation must be proportional to its defining equation. Normalize the coefficient of zw to one. The coefficient of w then recovers $-a$ from coefficients in B_0 , contradicting the transcendence of a over B_0 . These are nonvanishing statements in $\mathcal{U}$; placing more of their entries in K_0 cannot change them.

To obtain the valuation over K_0 , introduce an indeterminate T and solve

$$f(a + T, b + c_1T + c_2T^2 + \cdots) = 0$$

recursively in $K_0[[T]]$. After $c_1, \dots, c_{r-1}$ have been chosen, the coefficient of T^r is $f_w(a, b)c_r$ plus a known element of K_0 . Solving this linear equation determines c_r uniquely, since $f_w(a, b) \neq 0$. The resulting series lies in $K_0[[T]]$ and gives an embedding $K_0(V_1) \hookrightarrow K_0((T))$. Indeed z maps to the transcendental element $a + T$, and irreducibility of f over $K_0(z)$ shows that its only polynomial kernel is (f) . Now

$$L_\ell = T(b - w_{r_\ell} + O(T))$$

has order one, while $L_\ell - 4$ and all $L_m(L_m - 4)$, $m \neq \ell$, have nonzero constant terms. The order in T is the required valuation.

Both loci are the same curve with projection degree one. The symmetry of this curve shows that the other projection has degree one as well. Its irreducible equation therefore has degree one in each variable. Solving for w gives either

$$w = ez + e', \quad e \neq 0, \quad (8.14)$$

or

$$(z - u)(w - v) = b, \quad b \neq 0. \quad (8.15)$$

The constants here belong to B_0 . To see the classification directly, write the equation as $azw + cz + dw + e_0 = 0$. If $a = 0$, both linear coefficients are nonzero. If $a \neq 0$, complete the product; its constant is nonzero by irreducibility. In (8.15), $b \neq 1$, by the exact unit-circle exclusion (8.11).

In the line case, writing $a_\ell = z_{r_\ell}$,

$$L_\ell = e(z - a_\ell)^2.$$

Thus $L_\ell - 4$ has two distinct simple zeros. Choose one, say $z = a_\ell + c$ with $c^2 = 4/e$, and let s_ℓ be the corresponding point on the line. Its coordinates are algebraic over $B_0(a_\ell)$. For an independent anchor r_m , the conditions $L_m(s_\ell) = 0$ or 4 would make a_m one of finitely many elements algebraic over this field. This contradicts the joint genericity. Since $L_\ell = 4$ at the chosen zero, order at that abscissa in the rational function field $K_0(z)$ gives the required valuation.

In the fractional-linear case, put $\zeta = (z - u)/(a_\ell - u)$, where $a_\ell - u \neq 0$. The curve equation gives

$$w - w_{r_\ell} = \frac{b}{z - u} - \frac{b}{a_\ell - u} = -\frac{b(z - a_\ell)}{(z - u)(a_\ell - u)}.$$

Multiplying by $z - a_\ell$ and substituting $z - u = (a_\ell - u)\zeta$ gives

$$L_\ell = -b \frac{(\zeta - 1)^2}{\zeta}. \quad (8.16)$$

The equation $L_\ell - 4 = 0$ is equivalent to

$$b\zeta^2 + (4 - 2b)\zeta + b = 0.$$

Its discriminant is $16(1 - b) \neq 0$, and its two roots are nonzero because $b \neq 0$. Hence they give simple zeros of $L_\ell - 4$ at finite points with $z \neq u$. The corresponding point s_ℓ is algebraic over $B_0(a_\ell)$. As a function of the other anchor abscissa a_m , its squared distance to that anchor is

$$-b \frac{(z_{s_\ell} - a_m)^2}{(z_{s_\ell} - u)(a_m - u)}.$$

This is a nonconstant rational function: its numerator has degree two and its denominator degree one, and $z_{s_\ell} \neq u$. Its values cannot be zero or four at the transcendental a_m over $B_0(s_\ell)$. Thus the simple zero is shared by no other radicand, and order at its abscissa again gives the valuation over $K_0(z)$.

These cases exhaust the possibilities. In each, a nonempty product $\prod_{\ell \in I} D_\ell$ has valuation one at any v_ℓ with $\ell \in I$, and therefore is not a square. $\square$

The excluded value $b = 1$ illustrates why the deletion of pairs on unit circles containing at least $t_n^{9/10}$ points of P_n was necessary. Using $\zeta = (z - u)/(z_{r_\ell} - u)$ as in (8.16), when $b = 1$ we have

$$L_\ell(L_\ell - 4) = \left(\frac{\zeta^2 - 1}{\zeta} \right)^2.$$

Each radicand is then a square, so every integer-valued valuation takes an even value on it; the value-one valuations required above cannot exist on this common unit circle. For all the retained cases they were constructed after passing to K_0 ; no preservation of nonsquares under enlargement of the coefficient field has been assumed.

References

- [1] Péter Ágoston and Dömötör Pálvölgyi. An improved constant factor for the unit distance problem. *Studia Scientiarum Mathematicarum Hungarica*, 59(1):40–57, 2022.
- [2] Rudolf Ahlswede. An elementary proof of the strong converse theorem for the multiple-access channel. *Journal of Combinatorics, Information and System Sciences*, 7(3):216–230, 1982.
- [3] M. Ajtai, V. Chvátal, M. M. Newborn, and E. Szemerédi. Crossing-free subgraphs. In Alexander Rosa, Gert Sabidussi, and Jean Turgeon, editors, *Theory and Practice of Combinatorics*, volume 60 of *North-Holland Mathematics Studies*, pages 9–12. North-Holland, Amsterdam, 1982.
- [4] Noga Alon, Thomas F. Bloom, W. T. Gowers, Daniel Litt, Will Sawin, Arul Shankar, Jacob Tsimerman, Victor Wang, and Melanie Matchett Wood. Remarks on the disproof of the unit distance conjecture. <https://arxiv.org/abs/2605.20695v1>, 2026. arXiv:2605.20695v1, submitted 20 May 2026.
- [5] Kenneth L. Clarkson. New applications of random sampling in computational geometry. *Discrete & Computational Geometry*, 2:195–222, 1987.
- [6] Keith Conrad. Separability II. <https://kconrad.math.uconn.edu/blurbs/galoistheory/separable2.pdf>. Undated author notes; accessed 23 September 2026.
- [7] Thomas M. Cover and Joy A. Thomas. *Elements of Information Theory*. Wiley-Interscience, 2 edition, 2006.

- [8] P. Erdős. On sets of distances of n points. *The American Mathematical Monthly*, 53(5):248–250, May 1946.
- [9] David Haussler and Emo Welzl. ε -nets and simplex range queries. *Discrete & Computational Geometry*, 2:127–151, 1987.
- [10] Nets Katz and Olivine Silier. Structure of cell decompositions in extremal Szemerédi–Trotter examples. <https://arxiv.org/abs/2303.17186v2>, 2026. Version 2, 1 June 2026; first submitted 2023.
- [11] Nets Hawk Katz and Gábor Tardos. A new entropy inequality for the Erdős distance problem. In János Pach, editor, *Towards a Theory of Geometric Graphs*, volume 342 of *Contemporary Mathematics*, pages 119–126. American Mathematical Society, Providence, RI, 2004.
- [12] Salma Kuhlmann. Real algebraic geometry lecture notes: Lecture 09. <https://www.math.uni-konstanz.de/algebra/WS0910/Notes09.pdf>, 2009. 17 November 2009, Theorems 3.1 and 4.1.
- [13] F. Thomson Leighton. *Complexity Issues in VLSI: Optimal Layouts for the Shuffle-Exchange Graph and Other Networks*. Foundations of Computing. MIT Press, Cambridge, MA, 1983.
- [14] James S. Milne. Algebraic number theory. <https://www.jmilne.org/math/CourseNotes/ANTc.pdf>, 2020. Version 3.08, 19 July 2020.
- [15] James S. Milne. Fields and galois theory. <https://www.jmilne.org/math/CourseNotes/FT.pdf>, 2022. Version 5.10, September 2022.
- [16] James S. Milne. Algebraic geometry. <https://www.jmilne.org/math/CourseNotes/AG.pdf>, 2024. Version 6.10, 11 November 2024.
- [17] OpenAI. Planar point sets with many unit distances. <https://cdn.openai.com/pdf/74c24085-19b0-4534-9c90-465b8e29ad73/unit-distance-proof.pdf>, 2026. Public manuscript, 20 May 2026.
- [18] János Pach, Orit E. Raz, and József Solymosi. Erdős’s unit distance problem and rigidity. In Hee-Kap Ahn, Michael Hoffmann, and Amir Nayyeri, editors, *42nd International Symposium on Computational Geometry (SoCG 2026)*, volume 367 of *Leibniz International Proceedings in Informatics*, pages 83:1–83:9. Schloss Dagstuhl–Leibniz-Zentrum für Informatik, 2026.
- [19] Will Sawin. An explicit lower bound for the unit distance problem. <https://arxiv.org/abs/2605.20579v1>, 2026. Version 1, 20 May 2026.
- [20] Joseph H. Silverman. Lecture notes on arithmetic dynamics. <https://swc-math.github.io/aws/2010/2010SilvermanNotes.pdf>, 2010. Arizona Winter School lecture notes, 13–17 March 2010, Section 4.1.
- [21] J. Spencer, E. Szemerédi, and W. T. Trotter, Jr. Unit distances in the Euclidean plane. In B. Bollobás, editor, *Graph Theory and Combinatorics*, pages 293–303. Academic Press, London, 1984. Proceedings of the Cambridge Conference, 1983.
- [22] László A. Székely. Crossing numbers and hard Erdős problems in discrete geometry. *Combinatorics, Probability and Computing*, 6(3):353–358, 1997.

- [23] Gábor Tardos. On distinct sums and distinct distances. *Advances in Mathematics*, 180(1):275–289, 2003.
- [24] The Stacks Project Authors. The Stacks Project. <https://stacks.math.columbia.edu/tag/0BC1>, 2026. Lemma 10.120.10, Tag 0BC1; accessed 23 September 2026.