Electronic Colloquium on Computational Complexity, Revision 1 of Report No. 244 (2026)

$\Omega((\log n/\log\log n)^2)$ Lower Bounds for Dynamic Graph Problems

Young Kun Ko*

Abstract

We prove an $\Omega\big((\log n/\log\log n)^2\big)$ unconditional lower bound on the maximum of the query time and update time for dynamic data structures supporting reachability queries in n-node directed acyclic graphs under edge insertions. This improves the $\widetilde{\Omega}(\log^{3/2} n)$ lower bound of Larsen and Yu [SICOMP 2025], and matches the strongest lower bound known for any dynamic problem. The same bound holds for incremental undirected shortest paths and subgraph connectivity. To prove it, we bring the recent framework of Ko [FOCS 2026], which gave this bound for Pătraşcu's multiphase problem [STOC 2010] with Inner Product, to the multiphase problem with Disjointness. Our main technical contribution is to show that verification in Ko's 2.5-round multiphase communication game makes a one-sided corruption bound suffice in place of discrepancy, which Disjointness lacks.

*Department of Computer Science and Engineering, Pennsylvania State University, University Park, PA, USA. ykko@psu.edu

ISSN 1433-8092

1 Introduction

Dynamic graph problems, such as reachability, shortest paths and connectivity, are among the most natural and well-studied problems in theoretical computer science. A data structure maintains a graph on n vertices under edge updates, and answers queries about the current graph; in dynamic directed reachability, for instance, a query asks "is v reachable from u?". The goal is to make both updates and queries fast. For directed reachability, every known fully dynamic algorithm needs polynomial time for its updates or for its queries. Is this necessary?

There are two very different ways to argue that it is. A *conditional* lower bound shows that a fast data structure would give a fast algorithm for a problem conjectured to be hard, such as 3SUM, all-pairs shortest paths (APSP) or CNF satisfiability [Pat10, AW14, HKNS15]. It is a reduction: it shows that reachability is at least as hard as these problems, but not that any of them is hard. An *unconditional* lower bound proves, from first principles, that no data structure can be fast, whatever algorithms are discovered in the future. Conditional lower bounds have been remarkably successful at charting which problems are as hard as which, but only unconditional lower bounds tell us that a problem is hard.

This distinction recently became concrete: two of the three central conjectures of fine-grained complexity, the 3SUM and APSP conjectures, were refuted [AW26], with truly subquadratic and truly subcubic algorithms, respectively, and the result has been formalized in Lean.[1] Every lower bound conditioned on them, including Pătraşcu's original evidence for the multiphase problem that this paper builds on, lost its support at once. Lower bounds based on other conjectures, such as the OMv conjecture, still stand, but this is a good moment for healthy skepticism.[2] Unconditional lower bounds are sometimes seen as quantitatively weak next to conditional ones, but this compares statements of different kinds: a conditional lower bound can disappear overnight, while an unconditional one, however modest, is eternal. This is why we must always strive for unconditional lower bounds.

Unconditional lower bounds for data structures are proved in the cell-probe model [Yao81]. Memory consists of cells of $w = \Theta(\log n)$ bits, a data structure is charged one unit for each cell it reads or writes, and computation is free, so a lower bound in this model applies to every realistic architecture. In the dynamic setting, the update time t_u and the query time t_q are the numbers of cells probed by an update and by a query, respectively. Because the model is so strong, cell-probe lower bounds are among the hardest unconditional lower bounds to prove: for static data structures, even slight improvements of the known bounds would imply new circuit lower bounds [Vio18, DGW19, KPI25], and strong ones face a natural-proofs barrier [KLMS26]. For dynamic directed reachability, the best unconditional lower bound, due to Larsen and Yu [LY25], is $\max(t_u, t_q) = \widetilde{\Omega}(\log^{3/2} n)$.

In this work, we prove an unconditional $\Omega\big((\log n/\log\log n)^2\big)$ lower bound for dynamic directed reachability. More precisely, every zero-error data structure with worst-case update time t_u and expected query time t_q has $t_q = \Omega\big((\log n/\log(t_u w))^2\big)$, already for incremental reachability on depth-2 DAGs (Theorem 1.1). This surpasses the bound of Larsen and Yu by a factor of about $\sqrt{\log n}$ (Fig. 1), and matches the strongest lower bound known for any dynamic problem [Lar12, Ko26], which [Ko26] argues is likely the limit of the Chronogram framework. The same $\Omega\big((\log n/\log\log n)^2\big)$ bound holds for incremental undirected shortest paths and for subgraph connectivity. We obtain these bounds by bringing the recent framework of [Ko26], which proved this bound for the multiphase problem with Inner Product, to the multiphase problem with Disjointness, from which Pătraşcu's reductions [Pat10] give the graph problems.

[1] See the `3sum-apsp` folder of https://github.com/anthropics/formal-math.

[2] SETH is in a different position: refuting it would itself yield new circuit lower bounds [Wil13, JMV15].

1.1 Previous work

Upper bounds. Fast updates or fast queries alone are easy: recording the updates and searching the graph at query time makes updates fast, and maintaining the answers to all queries explicitly makes queries fast, but each makes the other operation polynomially slow. A long line of seminal work in data structure design gives algorithms that balance the two. In undirected graphs, connectivity can be maintained with polylogarithmic update and query time [HDLT01, KKM13]. For directed reachability and transitive closure, there are fully dynamic algorithms [Kin99, San04, VDBNS19], including algebraic ones based on dynamic matrix inverse [San04, VDBNS19], and for shortest paths, fully dynamic algorithms for all pairs [DI04, Tho05], all with polynomial update or query time.

Conditional lower bounds. To argue that these algorithms are close to optimal, a series of conditional lower bounds has been developed since 2010. This line of work grew out of the frustration of not being able to prove strong unconditional lower bounds: at the time (2010), the highest unconditional lower bound for any explicit dynamic problem was $\Omega(\log n)$ [Pat10], far from the polynomial bounds believed to hold. Pătraşcu [Pat10] initiated the systematic study of conditional lower bounds for dynamic problems. He introduced the multiphase problem, showed that it requires polynomial time under the 3SUM conjecture, and reduced it to dynamic reachability, shortest paths and subgraph connectivity, among others. Building on this, polynomial lower bounds for dynamic reachability and many other dynamic graph problems now follow from a handful of popular conjectures: the 3SUM and APSP conjectures and the Strong Exponential Time Hypothesis (SETH) [IPZ01, AW14], the hardness of combinatorial Boolean matrix multiplication [AW14], and the Online Matrix-Vector multiplication (OMv) conjecture [HKNS15]. For some trade-offs they match the known algorithms [VDBNS19]. Together, these results suggest that dynamic reachability requires polynomial update or query time, but only as long as the conjectures behind them hold, and as discussed above, the 3SUM and APSP conjectures do not [AW26].

Unconditional lower bounds. If we turn to unconditional lower bounds, however, the progress has been very slow, yet steady. In their seminal work, Fredman and Saks [FS89] introduced the Chronogram method and proved a lower bound of $\Omega(\log n/\log\log n)$, which transfers to dynamic connectivity by a simple reduction [MSVT94, HF98], and Pătraşcu and Demaine [PD06] improved this to $\Omega(\log n)$. For almost two decades, this remained the strongest unconditional lower bound for any natural graph problem. Stronger techniques were developed in the meantime. Larsen [Lar12] proved $\Omega\big((\log n/\log\log n)^2\big)$, but only for weighted problems, whose queries return $\Omega(\log n)$ bits. For problems whose queries return a single bit, such as reachability, there was no technique for proving even an $\omega(\log n)$ lower bound until the breakthrough of Larsen, Weinstein and Yu [LWY20], who proved $\widetilde{\Omega}(\log^{3/2} n)$. Building on their technique, Larsen and Yu [LY25] finally obtained the first super-logarithmic lower bound for a natural graph problem: for n-vertex DAGs under edge insertions, with worst-case update time t_u, expected query time t_q and word size $w = \Omega(\log n)$, they show $t_q = \Omega\big(\log^{3/2} n/\log^2(t_u w)\big)$ [LY25, Theorem 1], so, for $w = \Theta(\log n)$, $\max(t_u, t_q) = \widetilde{\Omega}(\log^{3/2} n)$. Fig. 1 summarizes this progress.

Each of these improvements may look marginal, but each required a conceptual shift in proof technique. In the Chronogram [FS89], the updates are split into epochs whose sizes grow geometrically going back in time, and a query must read information about each epoch from the cells written during that epoch. Pătraşcu and Demaine [PD06] replaced the epochs by a tree of time intervals and bounded the information transferred between adjacent intervals, which removes the $\log\log n$ loss of the geometric epochs. Larsen [Lar12] showed, by cell sampling, that a query must read

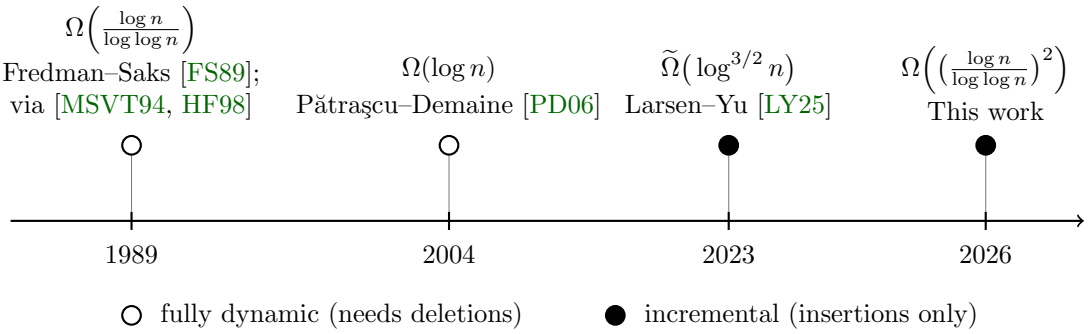


Figure 1: Unconditional cell-probe lower bounds for dynamic directed reachability, with $w = \Theta(\log n)$, stated as bounds on $\max(t_u, t_q)$. The two fully dynamic bounds are for undirected connectivity and carry over to directed reachability. The first is Fredman and Saks's bound for prefix parity [FS89], which transfers to dynamic connectivity by a simple reduction observed independently by Miltersen et al. [MSVT94] and by Henzinger and Fredman [HF98].

$\Omega(\log n/\log(t_u w))$ cells from each epoch, not just one: if it read fewer, a small sample of the cells of an epoch would answer many queries, and hence encode the updates of that epoch in fewer bits than their entropy. However, Larsen's original argument [Lar12] needs queries with long answers. Larsen, Weinstein and Yu [LWY20] extended it to Boolean queries, by reducing to a one-way communication game in which Bob only needs to guess the answer with a small advantage over random guessing, which they analyze via a peak-to-average lemma based on Chebyshev polynomials.

Despite this progress, it has long been conjectured that the true lower bound is much higher. Besides the conditional lower bounds above, Pătraşcu [Pat10] conjectured that the multiphase problem requires polynomial time, and identified a three-party communication game, now known as the multiphase communication game, whose hardness would imply this even in the cell-probe model, i.e. unconditionally; by his reductions, this would give polynomial unconditional lower bounds for dynamic reachability. Toward this conjecture, lower bounds for the multiphase problem are known only in restricted settings: polynomial bounds for non-adaptive data structures [BL15], threshold bounds in which very fast queries force polynomial update time [CGL15], lower bounds for semi-adaptive data structures, via two-round communication lower bounds [KW20, DL20], and lower bounds for a 1.5-round version of Pătraşcu's communication game [CEEP16].

A new landscape. Now the landscape is different. Ko [Ko26] recently paved a new way to $\Omega\big((\log n/\log\log n)^2\big)$ lower bounds for problems with Boolean answers, and proved this bound for the multiphase problem with Inner Product over $\mathbb{F}_2$. The key change there is to replace the one-way communication game behind previous bounds with a 2.5-round game, in which Bob verifies Alice's simulation, so that the players know when the simulation fails. This bound matches the strongest lower bound known for any dynamic problem [Lar12]. As discussed in [Ko26], it is likely the limit of the Chronogram framework, so Boolean problems can now be brought up to the best bounds that the Chronogram framework can give. Inner Product, however, is not known to capture natural dynamic graph problems: Pătraşcu's reductions to reachability and related problems start from Disjointness. In this work, we bring this framework to dynamic reachability (Theorem 1.1).

1.2 Our results

Theorem 1.1 (Directed reachability, informal)**.** *Every zero-error cell-probe data structure for incremental directed reachability on N vertices, with word size $w = \Theta(\log N)$, worst-case insertion time t_u and expected query time t_q, has $t_q = \Omega\big((\log N/\log(t_u w))^2\big)$; in particular $\max(t_u, t_q) = \Omega\big((\log N/\log\log N)^2\big)$. This holds already on depth-2 DAGs whose queries ask whether a fixed sink is reachable.*

The formal statement is Theorem A.1. Theorem 1.1 holds in the setting of Larsen and Yu with $w = \Theta(\log N)$, and needs only depth-2 DAGs. A fully dynamic data structure also handles insertion-only sequences, so the bound applies to fully dynamic directed reachability as well. The same bound holds for incremental undirected shortest paths and for incremental subgraph connectivity (Theorem A.1), by the same reduction. Since this reduction is standard, we leave the formal reduction to Section A for completeness.

The route: multiphase Disjointness. As our hard problem, we consider the multiphase problem of Pătraşcu [Pat10] with Disjointness, in its insert-only form.

- *Phase I.* Sets $S_1, \ldots, S_m \subseteq [n]$ arrive and are preprocessed arbitrarily.
- *Phase II.* Elements are inserted one at a time into a set $X \subseteq [n]$, initially empty; each insertion probes at most t_u cells.
- *Phase III.* An index $q \in [m]$ arrives, and the data structure outputs whether $S_q \cap X = \emptyset$, reading at most t_q cells in expectation; it does not write.

Problem 1: Insert-only multiphase Disjointness

As mentioned above, Pătraşcu conjectured that $\max(t_u, t_q) \geq n^{\Omega(1)}$ when $m = n^{1+\Omega(1)}$, and reduced the problem to several dynamic problems, including dynamic reachability. Our main technical result is the following.

Theorem 1.2 (Main technical theorem, informal)**.** *Every zero-error cell-probe data structure for insert-only multiphase Disjointness with $m = n$ query sets, word size $w = \Theta(\log n)$, worst-case update time t_u per insertion and expected query time t_q has $t_q = \Omega\big((\log n/\log(t_u w))^2\big)$.*

The formal statement is Theorem 3.1. Theorem 1.1 follows by Pătraşcu's reduction (Section A), in which each insertion into X is one edge insertion. Only insertions, i.e. setting a bit of X to one, are charged, as in Pătraşcu's reductions to dynamic problems. As in [LY25], a data structure may use random coins, drawn once; zero error means that it answers correctly for every input and every outcome of the coins, the update time is worst case over inputs and coins, and the query time is expected over the coins, for every input.

1.3 Proof overview

We sketch the proof of Theorem 1.2, which follows the framework of [Ko26]. We can interpret Ko's framework as working in two layers. The outer layer is the Chronogram method of Fredman and Saks [FS89]: we split the insertions into $\Theta(\log n/\log(t_u w))$ epochs, and show that a query must probe $\Omega(\log n/\log(t_u w))$ cells written in each of a constant fraction of them. The inner layer proves this bound for a single epoch, following [Ko26]: if a query probes fewer cells of the epoch, we

turn the data structure into a too-good-to-be-true protocol for a 2.5-round communication game. Section 3 gives the details.

1.3.1 The outer layer: the Chronogram method

At a high level, the Chronogram method [FS89] divides the insertions into ℓ epochs whose sizes grow geometrically going back in time. We number the epochs $\ell, \ell-1, \ldots, 1$ in the order in which they occur, so that the larger epochs come first and epoch 1 is the most recent. The main observation of Fredman and Saks is that we can decompose the cells by the epoch of their last write: let C_i be the set of cells last written in epoch i. Each probe goes to at most one of the disjoint sets $C_1, \ldots, C_\ell$, so for most epochs i, a random query probes only $O(t_q/\ell)$ cells of C_i in expectation. The goal then is to show that for a constant fraction of the epochs i, a random query must probe $\Omega(\log n/\log(t_u w))$ cells of C_i. With $\ell = \Theta(\log n/\log(t_u w))$, this gives $t_q = \Omega\big((\log n/\log(t_u w))^2\big)$.

The hard distribution. Recall that in multiphase Disjointness (Problem 1), the query sets $S_1, \ldots, S_m$ arrive first, the elements of X are then inserted, and finally a query asks whether $S_q \cap X = \emptyset$. We split the universe $[n]$ into blocks $B_1, \ldots, B_\ell$. In epoch i, $s_i = (t_u w)^{2i}$ random elements of B_i are inserted, and every query set has s_i random elements in B_i. In other words, this is the hard distribution of Babai, Frankl and Simon [BFS86] for Disjointness, two independent random sets of size about the square root of the universe, spread across the epochs: each block carries one copy of it, made slightly sparser so that the ℓ copies together meet with constant probability. Two features of this distribution matter.

- *Recent epochs cannot record older ones.* The epoch sizes are geometric: going back in time, they grow by a factor $(t_u w)^2$. So the insertions after epoch i number at most $s_{i-1} + s_{i-2} + \cdots + s_1 \le 2s_{i-1} = 2s_i/(t_u w)^2$, and since each insertion writes at most t_u cells of w bits, they write only $O(s_i/(t_u w))$ bits. This is far too few to describe the s_i elements inserted in epoch i, so a query that needs these elements must read cells of C_i. This is the key property of the Chronogram method.
- *Every epoch matters.* Block B_i has ℓs_i^2 elements. By the birthday paradox, two random sets of size s_i in a universe of size ℓs_i^2 meet with probability about $1/\ell$. So the answer is "disjoint" with constant probability, and each block can change it.

1.3.2 The inner layer: a 2.5-round multiphase communication game

Suppose, for contradiction, that $t_q = o\big((\log n/\log(t_u w))^2\big)$, and fix an epoch i in which a random query probes only $k = O(t_q/\ell)$ cells of C_i in expectation. The idea of cell sampling, introduced to the Chronogram method by Larsen [Lar12], is that a small random sample of C_i then already suffices to answer a nontrivial fraction of the queries. If each cell of C_i is sampled independently with probability $p = \Theta(1/(t_u w))$, then all k cells of C_i probed by a query are sampled with probability about p^k. Both $1/\ell$ and $\log(1/p)$ contribute a factor of $\log(t_u w)$, so

$$p^k \;\ge\; \exp\Big(-O\Big(\frac{t_q \log^2(t_u w)}{\log n}\Big)\Big) \;=\; n^{-o(1)}.$$

This is the source of the quadratic form of the bound. To exploit it, we use the following variant of the 2.5-round multiphase communication game of [Ko26]. Alice holds the query sets $S_1, \ldots, S_m$, Bob holds X, and a prover, Merlin, holds both.

- *Round 0 (Merlin to Bob).* Merlin sends Bob a message U.
- *Round 0.5 (Bob to Alice).* Bob sends Alice a message Φ, which may depend on U, X and his private randomness, but not on the query.
- *Round 1 (Alice to Bob).* A random query index $Q \in [m]$ is revealed to Alice only, and she sends Bob a message A_Q.
- *Round 2 (Bob).* Bob outputs whether $S_Q \cap X = \emptyset$, or FAIL.

Protocol 1: The 2.5-round multiphase communication game (cf. [Ko26, Protocol 3])

This game is a restriction of Pătraşcu's multiphase communication game, in which Alice and Bob both know the query and may then communicate freely. Pătraşcu's game has resisted all attempts at a lower bound, while the more restricted games for which lower bounds are known [CEEP16, KW20] do not imply lower bounds for general data structures. The 2.5-round game sits in between. Bob's message Φ does not depend on the query, which makes the game amenable to information-theoretic analysis, and yet the game can still simulate a data structure.

Simulation. We turn the data structure into a protocol for the game as follows (Step 1, Section 3.2).

- Merlin's message U is the list of all cells written in epochs $i, \dots, 1$, with their final contents.
- Bob's message Φ consists of the elements of X outside B_i, the contents of all cells written after epoch i, and each cell of C_i independently with probability $p = \varepsilon/(32t_u w)$, for a small constant $\varepsilon > 0$. With Φ, Alice knows the whole memory except the unsampled cells of C_i.
- Alice's message A_Q is her simulation of query Q on the memory she knows.
- Bob checks A_Q against U. Alice's simulation can only go wrong on unsampled cells of C_i, which are listed in U, so Bob detects every mistake. He outputs Alice's answer if A_Q is consistent with U, and FAIL otherwise.

So every answer that Bob outputs is correct. Moreover, Alice's simulation is correct whenever all k cells of C_i that the query probes are sampled, which happens with probability about $p^k = n^{-o(1)}$.

Verification is what distinguishes this protocol from the one-way games behind earlier bounds [Lar12, LY25]. There, Alice answers on her own and cannot tell whether her simulation is correct, since an unsampled cell of C_i looks to her like a cell that was not written in epoch i at all; a failed simulation is indistinguishable from a successful one. Here, Bob knows exactly when the simulation fails.

On the other hand, the protocol reveals little about block i. Bob's message reveals at most $\varepsilon s_i/4$ bits about $X \cap B_i$, a small fraction of its entropy, and Alice's simulation has $O(t_q w)$ bits. Merlin's list U is long, but it is shared by all m query sets, so it reveals little about the random query set S_Q. Yet Bob learns, with probability $n^{-o(1)}$, that S_Q and X are disjoint. This is what makes the protocol too good to exist.

1.3.3 Our key insight

In [Ko26], Bob outputs a random guess when his verification fails, and the protocol is shown to be too good through its advantage over a random guess. Conditioned on the messages of Alice and Bob, the inputs still have nearly full min-entropy and are nearly independent, and for Inner

Product the proof ends with a discrepancy bound: under every product distribution with nearly full min-entropy, the answer is nearly unbiased. As remarked in [Ko26], we do not expect the direct application of [Ko26] to work for Disjointness.

Our key observation is that verification makes a much weaker, one-sided property enough. As noted in [Ko26], the fact that the players know when the simulation fails is not used there. We use it: Bob outputs FAIL instead of a guess, so every answer he outputs is correct. We can then condition on the answer "disjoint", and only need that every product distribution with nearly full min-entropy puts noticeable mass on intersecting pairs. This one-sided corruption bound does hold for Disjointness: it is the rectangle bound of Babai, Frankl and Simon [BFS86], which we adapt to our sparser sets in Section 4. More generally, the argument uses Disjointness only through the answers of accepted runs and the fact that intersection is an OR over the blocks, so any OR over the blocks of functions with an analogous one-sided corruption bound would give the same lower bound for its multiphase problem.

1.3.4 The complete argument

Our proof proceeds in four steps. As in [Ko26], the first is a simulation, and the other three prove a communication lower bound.

1. **Simulation** (Section 3.2). A too-good-to-be-true data structure yields a protocol for the game of Protocol 1 in which Bob outputs "disjoint" with probability $n^{-o(1)}$, and every answer he outputs is correct.

2. **A monochromatic near-rectangle** (Section 3.3). Condition on Bob's message, on Alice's messages for all queries, and on Bob outputting "disjoint". Then $S_Q \cap B_i$ and $X \cap B_i$ still have nearly full min-entropy, and they are nearly independent. Since every answer Bob outputs is correct, the two sets are always disjoint.

3. **A nearly monochromatic rectangle** (Section 3.3.4). By Pinsker's inequality, replacing this distribution by the product of its two marginals moves little mass. So the product, a rectangle, puts little mass on intersecting pairs.

4. **Corruption** (Section 3.3.4). Every product of two distributions on sparse subsets of B_i with nearly full min-entropy puts mass $\Omega(1/\ell)$ on intersecting pairs, a contradiction.

The data structure may also be randomized, with expected query time, as in [LY25]: we fix its coins, and Alice stops each simulation after $8t_q$ probes, which costs only a constant factor.

Organization. Section 2 collects facts on information theory, min-entropy and rectangles. Section A gives Pătraşcu's reduction to dynamic graph problems and derives Theorem 1.1 from Theorem 3.1. Section 3 defines the hard distribution, proves the simulation theorem (Step 1) and the communication lower bound (Steps 2–4), and completes the proof of Theorem 3.1. Section 4 proves the corruption bound for sparse sets used in Step 4.

2 Preliminaries

2.1 Information theory

We recall the notions from information theory used in this paper; see [CT06] for background. Logarithms are base 2 unless written ln.

Definition 2.1 (Entropy, mutual information, divergence)**.** The entropy of a random variable X is $H(X) := \sum_x \Pr[X = x] \log \frac{1}{\Pr[X=x]}$, and the conditional entropy is $H(X|Y) := \mathbb{E}_{y\sim Y}\left[H(X|Y = y)\right]$. The mutual information between X and Y, conditioned on Z, is $I(X;Y|Z) := H(X|Z) - H(X|Y,Z)$; when we need to specify the distribution μ of (X,Y), we write $I_\mu(X;Y)$. The KL divergence and the total variation distance between two distributions P and Q are

$$D(P \,\|\, Q) := \sum_x P(x) \log \frac{P(x)}{Q(x)} \qquad \text{and} \qquad \|P - Q\|_{TV} := \frac{1}{2}\|P - Q\|_1.$$

For a random variable X and an event E, we write $X|_E$ for the distribution of X conditioned on E, for instance $X|_{Y=y}$. A random variable in place of a distribution, as in $D(X \,\|\, Y)$, stands for its distribution.

Fact 2.2 (Basic properties)**.** *For any random variables A, B, C and D:*

(a) If A takes at most 2^λ values, then $H(A) \le \lambda$. Moreover, $0 \le H(A|B) \le H(A)$.

(b) $I(A;B|C) = I(B;A|C)$ and $0 \le I(A;B|C) \le H(A|C)$. If A and B are independent given C, for instance if A is a function of C, then $I(A;B|C) = 0$.

(c) $I(A;B|C) = \mathbb{E}_{c\sim C}\left[I(A;B|C = c)\right]$.

(d) $D(P \,\|\, Q) \ge 0$ for any two distributions P and Q.

(e) If B' is a function of B, then $I(A;B'|C) \le I(A;B|C)$.

(f) If D is independent of (A,B,C), then $I(A;B|C,D) = I(A;B|C)$.

Fact 2.3 (Chain rules)**.** *(a) For any random variables $A_1, \ldots, A_k$, B and C,*

$$I(A_1 \cdots A_k; B|C) \;=\; \sum_{j=1}^{k} I\big(A_j; B \;\big|\; CA_1 \cdots A_{j-1}\big).$$

(b) For any two pairs of random variables (X,Y) and (X',Y'),

$$D\big(XY \,\|\, X'Y'\big) \;=\; D\big(X \,\|\, X'\big) + \mathop{\mathbb{E}}_{x\sim X}\Big[D\big(Y|_{X=x} \,\big\|\, Y'|_{X'=x}\big)\Big].$$

Fact 2.4. *For any random variables A, B, C and D, if $I(B;D|AC) = 0$, then $I(A;B|CD) \le I(A;B|C)$.*

Proof. By the chain rule (Fact 2.3(a)) and Fact 2.2(b),

$$I(A;B|CD) \;\le\; I(AD;B|C) \;=\; I(A;B|C) + I(B;D|AC) \;=\; I(A;B|C). \qquad \square$$

Fact 2.5 (KL divergence and mutual information)**.** *For any random variables A and B, $I(A;B) = \mathbb{E}_{b\sim B}\left[D\big(A|_{B=b} \,\|\, A\big)\right]$.*

Fact 2.6 (Pinsker's inequality)**.** *For any two distributions P and Q,*

$$\|P - Q\|_{TV} \;\le\; \sqrt{(\ln 2/2)\, D(P \,\|\, Q)}.$$

2.2 Min-entropy

The min-entropy of a random variable A is $H_\infty(A) := -\log \max_a \Pr[A = a]$, and for a distribution ν we write $H_\infty(\nu)$ likewise. We use the average min-entropy of [DORS08],

$$\widetilde{H}_\infty(A|B) := -\log \mathop{\mathbb{E}}_{b\sim B}\Big[\max_a \Pr[A = a|B = b]\Big].$$

Fact 2.7 ([DORS08])**.** *Let A, B, C be random variables.*

(a) If B has at most 2^λ possible values, then $\widetilde{H}_\infty(A|B, C) \geq \widetilde{H}_\infty(A|C) - \lambda$.

(b) For every $r > 0$, $\Pr_{b\sim B}\left[H_\infty(A|B = b) \geq \widetilde{H}_\infty(A|B) - r\right] \geq 1 - 2^{-r}$.

(c) If B' is a function of B, then $\widetilde{H}_\infty(A|B') \geq \widetilde{H}_\infty(A|B)$.

Fact 2.8. *Let $E \in \{0, 1\}$ be a random variable with $\Pr[E = 1] > 0$. Then $\widetilde{H}_\infty(X|Y, E{=}1) \geq \widetilde{H}_\infty(X|Y) - \log(1/\Pr[E = 1])$, where the left side is with respect to the distribution conditioned on $E = 1$.*

Proof. For every y with $\Pr[E = 1|Y = y] > 0$, we have $\Pr[X = x|Y = y, E = 1] \leq \Pr[X = x|Y = y]/\Pr[E = 1|Y = y]$ and, by Bayes' rule, $\Pr[Y = y|E = 1]/\Pr[E = 1|Y = y] = \Pr[Y = y]/\Pr[E = 1]$. Summing over such y, and then over all y,

$$\mathop{\mathbb{E}}_{y\sim Y|E=1} \max_x \Pr[X = x|Y = y, E = 1] \;\leq\; \sum_y \frac{\Pr[Y = y]\max_x \Pr[X = x|Y = y]}{\Pr[E = 1]} \;=\; \frac{2^{-\widetilde{H}_\infty(X|Y)}}{\Pr[E = 1]}.$$

□

Fact 2.9 (Conditioning on one side)**.** *Let A, B be random variables, and let $E \in \{0, 1\}$ be a function of A with $\pi := \Pr[E = 1] > 0$. Then*

$$I(A; B|E{=}1) \leq \frac{I(A; B)}{\pi} \qquad \text{and} \qquad H_\infty(B|E{=}1) \geq H_\infty(B) - \log\frac{1}{\pi},$$

where the left sides are with respect to the distribution conditioned on $E = 1$.

Proof. Since E is a function of A, $I(B; E|A) = 0$, so Fact 2.4 and Fact 2.2(b) and (c) give

$$I(A; B) \;\geq\; I(A; B|E) \;\geq\; \pi \cdot I(A; B|E{=}1).$$

For the second bound, $\Pr[B = b|E = 1] \leq \Pr[B = b]/\pi$ for every b. □

The next two facts pass between min-entropy and Shannon entropy. For a finite set $\mathcal{S}$, $\mathcal{U}_\mathcal{S}$ denotes the uniform distribution on $\mathcal{S}$; note that $D(\mu \,\|\, \mathcal{U}_\mathcal{S}) = \log|\mathcal{S}| - H(\mu)$.

Fact 2.10 (Flat decomposition; folklore, see e.g. [Vad12, Lemma 6.10])**.** *If $H_\infty(\mu) \geq \log K$ for an integer $K \geq 1$, then μ is a convex combination of uniform distributions $\mathcal{U}_T$ on K-element sets T.*

Fact 2.11 (Smoothing [Ko26]; see also [JRS09])**.** *If $D(\mu \,\|\, \mathcal{U}_\mathcal{S}) \leq t$ with $t > 1$, then for every $\alpha \geq 4$ there exists an event E such that $\mu(E) \geq 1 - 1/\alpha$ and $H_\infty(\mu|_E) \geq \log|\mathcal{S}| - 2\alpha t$.*

2.3 Rectangles

In this section, we would like to connect the size of a combinatorial rectangle $\mathcal{A} \times \mathcal{B}$ with the min-entropy. We work with distributional versions of these notions, in which the size of a set is replaced by the min-entropy of a distribution.

Definition 2.12 (Deficit)**.** Let $\mathcal{S}$ be a finite family of sets. The *deficit* of a distribution ν on $\mathcal{S}$ is $\log|\mathcal{S}| - H_\infty(\nu)$.

The deficit is zero exactly for the uniform distribution on $\mathcal{S}$, and the uniform distribution on a subfamily $\mathcal{A} \subseteq \mathcal{S}$ has deficit $\log(|\mathcal{S}| / |\mathcal{A}|)$. So a distribution of deficit d is as spread out as a subfamily of density 2^{-d}; intuitively, it is what remains of a uniformly random element of $\mathcal{S}$ after about d bits about it have been revealed. Below, $(A, B) \sim \mu$ for a distribution μ on $\mathcal{S} \times \mathcal{S}$ with marginals μ^A, μ^B.

Definition 2.13 (Rectangles, near-rectangles, monochromatic)**.** Let $d, \kappa, \delta \geq 0$.

(a) μ is a *d-rectangle* if $\mu = \mu^A \otimes \mu^B$ and both marginals have deficit at most d. That is,

$$H_\infty(\mu^A) \;\geq\; \log|\mathcal{S}| - d \qquad \text{and} \qquad H_\infty(\mu^B) \;\geq\; \log|\mathcal{S}| - d.$$

(b) μ is a *(d, κ)-near-rectangle* if both marginals have deficit at most d and $I_\mu(A; B) \leq \kappa$.

(c) μ is *δ-monochromatic* if $\Pr_\mu[A \cap B \neq \emptyset] \leq \delta$, and *monochromatic* if it is 0-monochromatic, i.e. supported on disjoint pairs.

A d-rectangle is the distributional analog of a large combinatorial rectangle. For $\mathcal{A}, \mathcal{B} \subseteq \mathcal{S}$, the uniform distribution $\mathcal{U}_{\mathcal{A}\times\mathcal{B}} = \mathcal{U}_\mathcal{A} \otimes \mathcal{U}_\mathcal{B}$ is a d-rectangle if and only if $|\mathcal{A}|, |\mathcal{B}| \geq 2^{-d}|\mathcal{S}|$, since $\mathcal{U}_\mathcal{A}$ has deficit $\log(|\mathcal{S}| / |\mathcal{A}|)$. Conversely, every d-rectangle is a convex combination of uniform distributions on combinatorial rectangles of almost this size.

Corollary 2.14 (Flat decomposition of rectangles)**.** *Let μ be a d-rectangle on $\mathcal{S} \times \mathcal{S}$, and let K be an integer with $1 \leq K \leq 2^{-d}|\mathcal{S}|$. Then there are subfamilies $\mathcal{A}_j, \mathcal{B}_k \subseteq \mathcal{S}$ of size K and weights $p_j, q_k \geq 0$ with $\sum_j p_j = \sum_k q_k = 1$ such that*

$$\mu \;=\; \sum_{j,k} p_j q_k \, \mathcal{U}_{\mathcal{A}_j \times \mathcal{B}_k}.$$

Proof. Both marginals of μ have min-entropy at least $\log|\mathcal{S}| - d \geq \log K$. So by Fact 2.10, $\mu^A = \sum_j p_j \mathcal{U}_{\mathcal{A}_j}$ and $\mu^B = \sum_k q_k \mathcal{U}_{\mathcal{B}_k}$ for such subfamilies and weights. Therefore

$$\mu \;=\; \mu^A \otimes \mu^B \;=\; \sum_{j,k} p_j q_k \, \mathcal{U}_{\mathcal{A}_j} \otimes \mathcal{U}_{\mathcal{B}_k}. \qquad \square$$

A near-rectangle relaxes the product condition: its two sides may be correlated, and κ bounds their mutual information. In particular, a d-rectangle is a $(d, 0)$-near-rectangle. Replacing a near-rectangle by the product of its marginals gives a rectangle that is close in total variation distance (see, e.g., [Ko26]).

Lemma 2.15 (From near-rectangles to rectangles)**.** *If μ is a δ-monochromatic (d, κ)-near-rectangle, then $\mu^A \otimes \mu^B$ is a $\left(\delta + \sqrt{(\ln 2/2)\,\kappa}\right)$-monochromatic d-rectangle.*

Proof. Let $R := \mu^A \otimes \mu^B$. It has the same marginals as μ, so it is a d-rectangle. By the chain rule for KL divergence (Fact 2.3(b)) and Fact 2.5, $D(\mu \,\|\, R) = \mathbb{E}_{b\sim\mu^B}\left[D(A|_{B=b} \,\|\, \mu^A)\right] = I_\mu(A; B) \leq \kappa$, so $\|\mu - R\|_{TV} \leq \sqrt{(\ln 2/2)\,\kappa}$ by Fact 2.6. Therefore

$$\Pr_{(A,B)\sim R}\left[A \cap B \neq \emptyset\right] \;\leq\; \Pr_{(A,B)\sim\mu}\left[A \cap B \neq \emptyset\right] + \|\mu - R\|_{TV} \;\leq\; \delta + \sqrt{(\ln 2/2)\,\kappa}. \qquad \square$$

3 Main proof

In this section, we prove our main theorem, the formal version of Theorem 1.2.

Theorem 3.1 (Main theorem)**.** *Every zero-error cell-probe data structure for insert-only multiphase Disjointness with $m = n$ query sets, word size $w = \Theta(\log n)$, worst-case update time t_u and expected query time t_q has*

$$t_q \;\geq\; \Omega\left(\left(\frac{\log n}{\log(t_u w)}\right)^2\right).$$

We first recall the outline from Section 1.3. After defining the hard distribution (Section 3.1), suppose, for contradiction, that we have a too-good-to-be-true data structure.

1. *A too-good 2.5-round protocol* (Section 3.2). The data structure yields a 2.5-round protocol in which Bob verifies Alice's simulation of the query, and accepts a run answering "disjoint" with noticeable probability.

2. *A monochromatic near-rectangle* (Section 3.3). Conditioning on these accepted runs gives a distribution on the pair $(S_Q^{(i)}, X^{(i)})$. This distribution is monochromatic, since accepted answers are correct. It is also a near-rectangle: both marginals have high min-entropy, and the two sides are almost independent.

3. *A nearly monochromatic rectangle* (Section 3.3.4). Replacing this distribution by the product of its marginals gives a rectangle with high min-entropy that is nearly monochromatic.

4. *Corruption* (Section 3.3.4, proved in Section 4). Finally, the corruption bound for sparse sets shows that no such rectangle exists, which is a contradiction.

As in [Ko26], Step 1 is a simulation theorem (Section 3.2), and Steps 2–4 prove a lower bound for the 2.5-round game (Section 3.3); together they prove Theorem 3.1.

3.1 The hard distribution

Let $\beta := (t_u w)^4$. As usual, we assume that $n = \sum_{i\in[\ell]} \ell\beta^i$ for an integer ℓ, the number of epochs. We partition $[n]$ into blocks $B_1, \ldots, B_\ell$ with

$$|B_i| = n_i := \ell\beta^i, \qquad \text{and let} \qquad s_i := \sqrt{n_i/\ell} = \beta^{i/2}.$$

Epochs are indexed backward in time: epoch ℓ comes first and epoch 1 comes last. Under the distribution $\mathcal{D}$ of $(\vec{S}, X)$, independently:

- each query set S_j, $j \in [m]$, consists of s_i uniformly random elements of each block B_i; write $S_j^{(i)} := S_j \cap B_i$ and $\vec{S} := (S_1, \ldots, S_m)$;
- in epoch i, s_i uniformly random elements of B_i are inserted into X, in increasing order. We write $X^{(i)} := X \cap B_i$ for the set inserted in epoch i, $X^{(>i)} := X \cap (B_{i+1} \cup \cdots \cup B_\ell)$, and $X^{(\neq i)} := X \setminus B_i$.

The query index Q is uniform in $[m]$ and independent of $(\vec{S}, X)$. We write $\Pr_{\mathcal{D},Q}$ and $\mathbb{E}_{\mathcal{D},Q}$ for probability and expectation over $(\vec{S}, X) \sim \mathcal{D}$ and Q. In each block, S_Q and X share $s_i^2/n_i = 1/\ell$ elements in expectation over $\mathcal{D}$ and Q, so they are disjoint with probability at least $1/4$ (Claim 3.2(d) and Fig. 2). The *Chronogram epochs* are $\mathcal{I} := \{i \in [\ell] : i \geq 2(\ell+1)/3\}$; any range of $\Omega(\ell)$ epochs with $s_i \geq n^{\Omega(1)}$ would do.

We prove the following useful (counting) facts about the properties of our hard distribution.

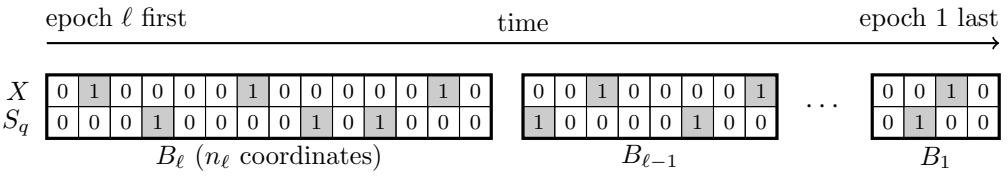


Figure 2: The hard distribution (not to scale), with X and S_q written as characteristic vectors. The ones of X in block B_i are inserted during epoch i. Block sizes grow by a factor $\beta = (t_u w)^4$ going back in time. In each block, the expected number of common ones is $s_i^2/n_i = 1/\ell$, so S_Q and X are disjoint with constant probability.

Claim 3.2 (Parameters)**.** *The hard distribution has the following properties.*

(a) $\ell = \Theta(\log n/\log(t_u w))$ *and* $|\mathcal{I}| \geq \ell/3 - 1$. *If moreover* $t_u w \leq n^{1/95}$, *then* $\ell \geq 20$.

(b) For $i \in \mathcal{I}$: $s_i \geq n^{1/3}$.

(c) $s_i/s_{i-1} = \sqrt{\beta} = (t_u w)^2$, $\sum_{j<i} s_j \leq 2s_{i-1}$, *and* $s_\ell \leq \sqrt{n}$.

(d) $\mathbb{E}_{\mathcal{D},Q}\left[\left|S_Q^{(i)} \cap X^{(i)}\right|\right] = s_i^2/n_i = 1/\ell$ *for every* i, *and* $\Pr_{\mathcal{D},Q}[S_Q \cap X = \emptyset] \geq (1-1/\ell)^\ell \geq 1/4$ *for* $\ell \geq 2$.

Proof. These are more-or-less standard calculations.

(a) Since $\beta \geq 2$, $\ell\beta^\ell \leq n \leq 2\ell\beta^\ell$. Hence $\ell \leq \log n$ and $\ell \log\beta \leq \log n \leq \ell\log\beta + \log\ell + 1$, so $\ell = \Theta(\log n/\log\beta)$. Moreover $|\mathcal{I}| \geq \ell - 2(\ell+1)/3 = (\ell-2)/3$. If $t_u w \leq n^{1/95}$, then $\log\beta \leq \frac{4}{95}\log n$, so $\ell \geq (\log n - \log\ell - 1)/\log\beta \geq \frac{95}{4}\left(1 - \frac{\log\log n + 1}{\log n}\right) \geq 20$.

(b) Since $w = \Theta(\log n)$ and $\ell \leq \log n$, $\beta = (t_u w)^4 \geq 2\ell$. For $i \geq 2(\ell+1)/3$, $s_i = \beta^{i/2} \geq (\beta^{\ell+1})^{1/3} \geq (\beta n/(2\ell))^{1/3} \geq n^{1/3}$.

(c) The ratio is immediate; $\sum_{j<i} s_j \leq s_{i-1}\sum_{r\geq 0}\beta^{-r/2} \leq 2s_{i-1}$ since $\beta \geq 4$; and $s_\ell^2 = n_\ell/\ell \leq n$.

(d) Each element of B_i lies in $S_Q^{(i)}$ with probability s_i/n_i and in $X^{(i)}$ with probability s_i/n_i, independently, so the expectation is $n_i(s_i/n_i)^2 = 1/\ell$. By Markov, block i is disjoint with probability at least $1 - 1/\ell$, and the blocks are independent. □

Setup. As usual, we fix the coins of the data structure by averaging, and we assume that $t_u w \leq n^{1/95}$ and $t_q \leq \log^2 n$, since otherwise the bound is trivial. So the data structure is deterministic, with zero error and worst-case update time t_u, and the number $T(q)$ of cells read by a query $q \in [m]$ is a function of $(\vec{S}, X)$ with $\mathbb{E}_{\mathcal{D},Q}[T(Q)] \leq t_q$. By Claim 3.2, $\ell \geq 20$ and $s_i \geq n^{1/3}$ for $i \in \mathcal{I}$. Since $w = \Theta(\log n)$, also $t_u w \geq 64/\varepsilon$, where ε is the constant of Lemma 4.1 (i.e. 1/40).

Epochs. As in all Chronogram-based proofs, the updates are processed in reverse order of the epoch index, so the larger epochs come first. We follow the notation and the setup of [LY25]. For $i \in \mathcal{I}$, let $M_{>i}$ be the memory after Phase I and epochs $\ell, \ldots, i+1$; it is a function of $(\vec{S}, X^{(>i)})$. For $j \in [\ell]$, let C_j be the set of cells last written in epoch j, and let $C_{<j} := \bigcup_{j'<j} C_{j'}$ be the set of

cells written after epoch j. The sets $C_1, \ldots, C_\ell$ are pairwise disjoint. Let $t_j(q)$ be the number of cells of C_j probed by query q, so that

$$\sum_{j\in[\ell]} t_j(q) \;\le\; T(q), \qquad \text{and let} \qquad \tau_j \;:=\; \mathop{\mathbb{E}}_{\mathcal{D},Q}\big[t_j(Q)\big].$$

3.2 Simulation theorem

First, we show the following simulation theorem, which argues that a too-good-to-be-true data structure yields a too-good-to-be-true protocol for the 2.5-round game of Protocol 1 (cf. [Ko26, Protocol 3]), in which Bob verifies Alice's simulation of the query against the actual memory. Our protocol is a variant and a tightening of the simulation protocol of [Ko26, Protocol 4], and its analysis follows [Ko26, Theorem III.2]. The main difference is Bob's output when the simulation fails: in [Ko26], he outputs FAIL together with an independent random guess, while here he outputs only FAIL. In addition, the epoch i is fixed in advance (Section 3.3.4) rather than chosen by Merlin, Merlin's message lists only the cells written in epochs $i, \ldots, 1$ rather than all updates, Bob sends Alice all elements of X outside B_i, and, since t_q bounds the query time only on average, Alice stops her simulation after $8t_q$ probes.

Let U consist of two lists, of the cells of C_i and of the cells of $C_{<i}$, each cell with its final content, that is, its content after Phase II. Let $p := \varepsilon/(32\, t_u w)$. Bob's private randomness R_{Bob} consists of an independent bit R_a with $\Pr[R_a = 1] = p$ for every cell a, and $(\vec{S}, X)$, Q and R_{Bob} are mutually independent. The length $|\cdot|$ of a message is measured in bits, padded to a fixed length, so that a message of length λ takes at most 2^λ values.

Theorem 3.3 (Simulation theorem)**.** *For every $i \in \mathcal{I}$ there is a 2.5-round protocol G_i for the game of Protocol 1, in which Merlin sends U and Bob's message has the form $\Phi = (X^{(\neq i)}, \Psi)$, with the following properties, where B_Q is the indicator that Bob outputs "disjoint" and $\rho_1 := \Pr_{\mathcal{D},R_{\text{Bob}},Q}[B_Q = 1]$.*

(a) (Zero error) *If Bob does not output* FAIL*, his output is the correct answer. In particular, $B_Q = 1$ implies $S_Q \cap X = \emptyset$.*

(b) (Lengths of U and Ψ) *$|U| \le 5\, t_u w\, s_i$ and $|\Psi| \le \varepsilon s_i/4$.*

(c) (Length of Alice's messages) *$|A_q| \le 17\, t_q w$ for every $q \in [m]$.*

(d) (Advantage) *$\rho_1 \ge \frac{1}{8}\, p^{8\tau_i} - e^{-n^{1/4}/3}$.*

(e) (Structure) *Ψ is a function of (U, R_{Bob}), each A_q is a function of $(\vec{S}, \Phi)$, and B_Q is a function of (Ψ, A_Q, U).*

Proof. We simulate the data structure by the 2.5-round protocol G_i of Protocol 2. Part (e) is immediate from Protocol 2: Bob computes Ψ from U and R_{Bob}, Alice computes A_q from $\vec{S}$ and Φ, and Bob's decision reads only Ψ, which records whether he aborted, A_Q, which records whether the run finished and its output, and the final contents listed in U.

Cost of the protocol. The following analysis of the cost is straightforward from Protocol 2. Each cell in a message is sent as an address and a content, $2w$ bits. Merlin's message U lists the cells of C_i and of $C_{<i}$. Here $|C_i| \le t_u s_i$, since each of the s_i insertions of epoch i writes at most t_u

- Alice holds $\vec{S}$. Bob holds X and R_{Bob}. Merlin holds $\vec{S}$ and X.
- Merlin sends U to Bob.
- As the 0.5-round message, Bob lets $\widetilde{C}_i := \{a \in C_i : R_a = 1\}$. If $\left|\widetilde{C}_i\right| > 2p\,t_u s_i$, Bob aborts: he sets $\Psi := \perp$, and outputs FAIL at the end. Otherwise Ψ lists the cells of $C_{<i} \cup \widetilde{C}_i$ with their final contents. Bob sends $\Phi := (X^{(\neq i)}, \Psi)$ to Alice. It is crucial that the message depends on U, X and R_{Bob}, but not on the query Q.
- The query $Q \in [m]$ is then revealed to Alice. Alice computes $M_{>i}$ from $\vec{S}$ and $X^{(>i)}$, and runs the query algorithm on Q. She answers a probe to a cell listed in Ψ with its listed content, and a probe to any other cell with its content in $M_{>i}$. She stops the run if it has not finished after $8t_q$ probes. She sends the transcript

$$A_Q \;:=\; \big((a_1, c_1), \ldots, (a_r, c_r), z_{out}\big), \qquad r \le 8t_q,$$

  where a_k is the k-th probed cell, c_k the content she used, and z_{out} the output, or a mark that the run was stopped.
- Bob *accepts* A_Q if (i) he did not abort, (ii) the run recorded in A_Q finished, and (iii) c_k is the final content of a_k for every k with $a_k \in C_i \cup C_{<i}$; he reads these final contents off U. If Bob accepts, he outputs z_{out}; otherwise he outputs FAIL.

Protocol 2: Simulation protocol G_i (cf. [Ko26, Protocol 4])

cells, and similarly $|C_{<i}| \le t_u \sum_{j<i} s_j \le 2t_u s_{i-1}$ by Claim 3.2(c), the property of the Chronogram epochs. Bob's list Ψ has at most $|C_{<i}| + 2p\,t_u s_i$ cells. Since $s_{i-1} = s_i/(t_u w)^2$ and $t_u w \ge 64/\varepsilon$,

$$\begin{aligned} |U| \;&\le\; 2w\big(\,|C_i| + |C_{<i}|\,\big) + 2 \;\le\; 2w\big(t_u s_i + 2t_u s_{i-1}\big) + 2 \;\le\; 5\,t_u w\, s_i, \\ |\Psi| \;&\le\; 2w\big(2t_u s_{i-1} + 2p\,t_u s_i\big) + 2 \;=\; \frac{4s_i}{t_u w} + \frac{\varepsilon s_i}{8} + 2 \;\le\; \frac{\varepsilon s_i}{4}. \end{aligned}$$

Alice's transcript records at most $8t_q$ probes, together with the number of probes and the output or the stop mark, so

$$|A_q| \;\le\; 16\,t_q w + \log(8t_q + 1) + 2 \;\le\; 17\,t_q w.$$

This proves (b) and (c).

Advantage. We now prove (a) and (d), using the following claim. It is the same argument as in [Ko26]: Bob accepts only the true execution of the query, and he accepts it whenever his sample contains every cell of C_i that the query probes.

Claim 3.4 (Acceptance)**.** *Fix $(\vec{S}, X)$ and $q \in [m]$. In (b) and (c), the probability is over R_{Bob}.*

(a) If Bob accepts A_q, then A_q is the true execution of query q. In particular, its output is $\text{DISJ}(S_q, X)$.

(b) $\Pr_{R_{\text{Bob}}}[\textit{Bob aborts}] \le e^{-n^{1/4}/3}$.

(c) If $T(q) \le 8t_q$, then $\Pr_{R_{\text{Bob}}}[\textit{Bob accepts } A_q] \ge p^{t_i(q)} - e^{-n^{1/4}/3}$.

Proof. (a) If Bob accepts A_q, then every content Alice used is the final content. For a cell of $C_i \cup C_{<i}$, Bob checked this. Any other cell is not listed in Ψ and is not written during epochs $i, \dots, 1$, so Alice used its content in $M_{>i}$, which is final. Hence her run is the true execution of query q, and its output is $\mathrm{DISJ}(S_q, X)$.

(b) $\left|\widetilde{C}_i\right|$ is a sum of $|C_i| \le t_u s_i$ independent indicators with mean p each. By the Chernoff bound, applied with the upper bound $p\,t_u s_i$ on the mean,

$$\Pr_{R_{\mathrm{Bob}}}\left[\left|\widetilde{C}_i\right| > 2p\,t_u s_i\right] \;\le\; e^{-p\,t_u s_i/3},$$

and $p\,t_u s_i = \varepsilon s_i/(32w) \ge n^{1/4}$.

(c) Let $P_q \subseteq C_i$ be the set of the $t_i(q)$ cells of C_i probed by the true execution of query q. If Bob does not abort and $P_q \subseteq \widetilde{C}_i$, then every cell that the true execution probes is either listed in Ψ or outside $C_i \cup C_{<i}$, so Alice reproduces the true execution, which finishes within $T(q) \le 8t_q$ probes, and Bob accepts. Since $\Pr_{R_{\mathrm{Bob}}}[P_q \subseteq \widetilde{C}_i] = p^{t_i(q)}$, a union bound with (b) gives (c). □

Part (a) of the theorem follows from Claim 3.4(a), since Bob outputs FAIL unless he accepts A_Q. It remains to show (d). Let F be the event that $S_Q \cap X = \emptyset$ and $T(Q) \le 8t_q$; it is determined by $(\vec{S}, X, Q)$. By Markov's inequality, $\Pr_{\mathcal{D},Q}[T(Q) > 8t_q] \le 1/8$, and by Claim 3.2(d), $\Pr_{\mathcal{D},Q}[S_Q \cap X = \emptyset] \ge 1/4$, so $\Pr_{\mathcal{D},Q}[F] \ge 1/8$. If F holds and Bob accepts A_Q, then the recorded output is "disjoint" by Claim 3.4(a), so $B_Q = 1$. Since R_{Bob} is independent of $(\vec{S}, X, Q)$, Claim 3.4(c), applied for each fixed $(\vec{S}, X, Q)$, and then the convexity of $x \mapsto p^x$ (with Jensen's inequality) give

$$\rho_1 \;\ge\; \mathop{\mathbb{E}}_{\mathcal{D},Q}\left[\mathbf{1}_F \cdot \left(p^{t_i(Q)} - e^{-n^{1/4}/3}\right)\right] \;\ge\; \Pr_{\mathcal{D},Q}[F] \cdot p^{\mathbb{E}_{\mathcal{D},Q}[t_i(Q)|F]} - e^{-n^{1/4}/3}.$$

Since $t_i \ge 0$, $\mathbb{E}_{\mathcal{D},Q}[t_i(Q)|F] \le \tau_i / \Pr_{\mathcal{D},Q}[F] \le 8\tau_i$, which gives (d). □

As in [Ko26, Remark III.4], the protocol errs only on an event the players detect: whenever Bob does not output FAIL, he has verified Alice's transcript against U, and his output is correct. Here we do not need Bob's output on the remaining runs at all: Steps 2–4 use only the event $B_Q = 1$, on which the answer is "disjoint" with certainty.

We now choose the parameters so that the protocol is too good: Bob accepts a correct "disjoint" answer with probability $n^{-o(1)}$.

Corollary 3.5 (Choice of parameters)**.** *If $\tau_i = o\big(\log n / \log(t_u w)\big)$, then $\rho_1 \ge n^{-o(1)}$.*

Proof. Since $t_u w \ge 64/\varepsilon$, $\log(1/p) = \log(32\,t_u w/\varepsilon) \le 2\log(t_u w)$, so $p^{8\tau_i} \ge 2^{-16\tau_i \log(t_u w)} = n^{-o(1)}$. By Theorem 3.3(d), $\rho_1 \ge \frac{1}{8} n^{-o(1)} - e^{-n^{1/4}/3} \ge n^{-o(1)}$. □

3.3 Communication lower bound

The goal of this section is to show that such a protocol is impossible, by showing that $\rho_1 < n^{-1/6}$ for every $i \in \mathcal{I}$ (Lemma 3.13). As in [Ko26], and going back to [KW20], the key idea is to extract from the protocol a random variable Z that contains Bob's output B_Q, such that for some value z of Z with $B_Q = 1$,

(i) $H_\infty\big(S_Q^{(i)} \mid Z = z\big)$ and $H_\infty\big(X^{(i)} \mid Z = z\big)$ are both large, and

(ii) $I\big(S_Q^{(i)}; X^{(i)} \mid Z = z\big)$ is small.

In [Ko26], a discrepancy bound for Inner Product then shows that Bob's output is nearly uncorrelated with the answer. Here, by Theorem 3.3(a), the distribution of $(S_Q^{(i)}, X^{(i)})$ given $Z = z$ is supported on disjoint pairs. Together with (i) and (ii), it is a monochromatic near-rectangle (Definition 2.13), and the corruption bound shows that no such near-rectangle exists.

Fix $i \in \mathcal{I}$ with $\rho_1 > 0$. Following [Ko26], we take

$$Z \;:=\; \big(Z_{\text{alice}}, B_Q\big), \qquad Z_{\text{alice}} \;:=\; \big(\Phi, A_{\text{all}}, Q, S_{<Q}\big),$$

where $S_{<Q} := (S_1, \dots, S_{Q-1})$, and $A_{\text{all}} := (A_1, \dots, A_m)$ with A_q the transcript that Alice sends when the query is q. As in [Ko26], Z contains Alice's messages for all queries, which lets us apply direct-sum arguments, but not Merlin's message U. By Theorem 3.3(e), A_{all} is a function of $(\vec{S}, \Phi)$, and B_Q is a function of (Z_{alice}, U), since Ψ is part of Φ and A_Q is determined by A_{all} and Q. We only consider values of Z with $B_Q = 1$, and write them as $(z, 1)$, where z is a value of Z_{alice}.

We use two facts throughout. First, Merlin's message is short compared to the number $m = n$ of query sets: by Theorem 3.3(b) and Fact 2.2(a),

$$H(U) \;\le\; 5\, t_u w\, s_i \;\le\; n^{3/5}, \tag{1}$$

since $s_i \le s_\ell \le \sqrt{n}$ (Claim 3.2(c)) and $t_u w \le n^{1/95}$. Second, the following standard direct-sum identity holds.

Claim 3.6 (Direct sum)**.** *For any random variables V and W that are functions of $(\vec{S}, X, R_{\text{Bob}})$,*

$$I\big(S_Q; W \;\big|\; V, Q, S_{<Q}\big) \;=\; \frac{1}{m}\, I\big(\vec{S}; W \;\big|\; V\big).$$

Proof. Since Q is uniform and independent of $(\vec{S}, X, R_{\text{Bob}})$, Fact 2.2(c) and the chain rule (Fact 2.3(a)) give

$$I\big(S_Q; W \;\big|\; V, Q, S_{<Q}\big) \;=\; \frac{1}{m} \sum_{q \in [m]} I\big(S_q; W \;\big|\; V, S_{<q}\big) \;=\; \frac{1}{m}\, I\big(\vec{S}; W \;\big|\; V\big). \qquad \square$$

3.3.1 Small information on S_Q

First, we would like to show that the min-entropy of $S_Q^{(i)}$ conditioned on $Z = (z, 1)$ is large. Here Z_{alice} contains Alice's messages for all m queries, which are long, but they carry only $O(t_q w)$ bits about a random S_Q on average. This is a bound on divergence rather than on min-entropy, so, as in [Ko26], we pass to min-entropy by smoothing, which requires conditioning $S_Q^{(i)}$ on a likely event E.

Proposition 3.7 (The query side)**.** *If $\rho_1 \ge n^{-1/6}$, then*

$$\Pr_{z \sim Z_{\text{alice}} | B_Q = 1} \left[\exists E \subseteq \binom{B_i}{s_i} : \begin{array}{l} \Pr\big[S_Q^{(i)} \in E \;\big|\; Z = (z,1)\big] \;\ge\; \dfrac{3}{4} \quad \textit{and} \\[2mm] H_\infty\big(S_Q^{(i)} \;\big|\; Z = (z,1),\, S_Q^{(i)} \in E\big) \;\ge\; \log \dbinom{n_i}{s_i} - 32\, n^{1/5} \end{array} \right] \;\ge\; \frac{3}{4}.$$

Proof. First we upper bound $I\big(S_Q^{(i)}; Z_{\text{alice}}, B_Q\big)$. Since $S_Q^{(i)}$ is part of S_Q, and S_Q is independent of $(Q, S_{<Q})$, the chain rule (Fact 2.3) gives

$$I\big(S_Q^{(i)}; Z_{\text{alice}}, B_Q\big) \;\le\; I\big(S_Q; \Phi, A_{\text{all}}, Q, S_{<Q}, B_Q\big) \;=\; I\big(S_Q; \Phi, A_{\text{all}}, B_Q \;\big|\; Q, S_{<Q}\big).$$

Since B_Q is a function of $(\Phi, A_{\text{all}}, Q, U)$, the chain rule and Claim 3.6 (with V constant) give

$$I\big(S_Q; \Phi, A_{\text{all}}, B_Q \bigm| Q, S_{<Q}\big) \;\le\; I\big(S_Q; \Phi, A_{\text{all}}, U \bigm| Q, S_{<Q}\big) \;=\; \frac{1}{m} I\big(\vec{S}; \Phi, A_{\text{all}}, U\big).$$

By the chain rule and Fact 2.2,

$$I\big(\vec{S}; \Phi, A_{\text{all}}, U\big) \;=\; I\big(\vec{S}; \Phi, U\big) + I\big(\vec{S}; A_{\text{all}} \bigm| \Phi, U\big) \;\le\; H(U) + H(A_{\text{all}}),$$

where $I(\vec{S}; \Phi, U) \le I(\vec{S}; X^{(\neq i)}, R_{\text{Bob}}, U) = I(\vec{S}; U | X^{(\neq i)}, R_{\text{Bob}}) \le H(U)$ by Fact 2.2(e) and the chain rule, since Ψ is a function of (U, R_{Bob}) (Theorem 3.3(e)) and $\vec{S}$ is independent of $(X^{(\neq i)}, R_{\text{Bob}})$. By (1) and $m = n$, $H(U)/m \le n^{-2/5}$, and by Theorem 3.3(c) and Fact 2.2(a), $H(A_{\text{all}})/m \le 17 t_q w$. Since $t_q \le \log^2 n$ and $w = \Theta(\log n)$,

$$I\big(S_Q^{(i)}; Z_{\text{alice}}, B_Q\big) \;\le\; n^{-2/5} + 17 t_q w \;=\; O(\log^3 n).$$

Next we condition on $B_Q = 1$, an event of probability ρ_1. Let $\mathcal{U}$ be the uniform distribution on $\binom{B_i}{s_i}$, which is the distribution of $S_Q^{(i)}$. By Fact 2.5, and since divergence is nonnegative (Fact 2.2(d)),

$$\begin{aligned} I\big(S_Q^{(i)}; Z_{\text{alice}}, B_Q\big) \;&=\; \mathop{\mathbb{E}}_{(z,b)\sim(Z_{\text{alice}}, B_Q)} \Big[D\big(S_Q^{(i)}|_{Z=(z,b)} \,\big\|\, \mathcal{U}\big)\Big] \\ &\ge\; \rho_1 \cdot \mathop{\mathbb{E}}_{z\sim Z_{\text{alice}}|_{B_Q=1}} \Big[D\big(S_Q^{(i)}|_{Z=(z,1)} \,\big\|\, \mathcal{U}\big)\Big]. \end{aligned}$$

Hence, since $\rho_1 \ge n^{-1/6}$,

$$\mathop{\mathbb{E}}_{z\sim Z_{\text{alice}}|_{B_Q=1}} \Big[D\big(S_Q^{(i)}|_{Z=(z,1)} \,\big\|\, \mathcal{U}\big)\Big] \;\le\; \frac{O(\log^3 n)}{\rho_1} \;\le\; n^{1/5}.$$

Finally, by Markov's inequality, $D\big(S_Q^{(i)}|_{Z=(z,1)} \,\|\, \mathcal{U}\big) \le 4n^{1/5}$ with probability at least $3/4$ over $z \sim Z_{\text{alice}}|_{B_Q=1}$. For such z, Fact 2.11, applied to the distribution of $S_Q^{(i)}$ given $Z = (z,1)$ with $\mathcal{S} = \binom{B_i}{s_i}$, $t = 4n^{1/5}$ and $\alpha = 4$, gives the desired set E. □

3.3.2 Small information on X

Next, we show that conditioning on $Z = (z,1)$ costs little min-entropy of $X^{(i)}$. The only information about $X^{(i)}$ in Z_{alice} is Bob's message Ψ, which is short, and conditioning on $B_Q = 1$ costs $\log(1/\rho_1)$ more bits.

Proposition 3.8 (The update side)**.**

$$\Pr_{z\sim Z_{\text{alice}}|_{B_Q=1}} \left[H_\infty\big(X^{(i)} \bigm| Z = (z,1)\big) \;\ge\; \log \binom{n_i}{s_i} - \frac{\varepsilon s_i}{4} - \log\frac{1}{\rho_1} - 2 \right] \;\ge\; \frac{3}{4}.$$

Proof. First we lower bound $\widetilde{H}_\infty\big(X^{(i)} \bigm| Z_{\text{alice}}\big)$. The set $X^{(i)}$ is uniform on $\binom{B_i}{s_i}$ and independent of $(\vec{S}, X^{(\neq i)}, Q)$, and Z_{alice} is a function of $(\Psi, \vec{S}, X^{(\neq i)}, Q)$ by Theorem 3.3(e). By Fact 2.7(c), and then (a) with $|\Psi| \le \varepsilon s_i/4$ (Theorem 3.3(b)),

$$\widetilde{H}_\infty\big(X^{(i)} \bigm| Z_{\text{alice}}\big) \;\ge\; \widetilde{H}_\infty\big(X^{(i)} \bigm| \Psi, \vec{S}, X^{(\neq i)}, Q\big) \;\ge\; \log \binom{n_i}{s_i} - \frac{\varepsilon s_i}{4}.$$

Next we condition on $B_Q = 1$, an event of probability ρ_1. By Fact 2.8,

$$\widetilde{H}_\infty\big(X^{(i)} \mid Z_{\text{alice}}, B_Q{=}1\big) \;\geq\; \widetilde{H}_\infty\big(X^{(i)} \mid Z_{\text{alice}}\big) - \log\frac{1}{\rho_1} \;\geq\; \log\binom{n_i}{s_i} - \frac{\varepsilon s_i}{4} - \log\frac{1}{\rho_1},$$

where the left side is the average min-entropy with respect to the distribution conditioned on $B_Q = 1$:

$$\widetilde{H}_\infty\big(X^{(i)} \mid Z_{\text{alice}}, B_Q{=}1\big) \;=\; -\log \mathop{\mathbb{E}}_{z\sim Z_{\text{alice}}|_{B_Q=1}} \Big[2^{-H_\infty(X^{(i)}\,|\,Z=(z,1))}\Big].$$

Finally, by Fact 2.7(b) with $r = 2$, applied to this distribution, $H_\infty(X^{(i)}|Z = (z, 1))$ is at least the left side minus 2 with probability at least $1 - 2^{-2} = 3/4$ over $z \sim Z_{\text{alice}}|_{B_Q=1}$. □

3.3.3 Small correlation between S_Q and X

Finally, we show that $S_Q^{(i)}$ and $X^{(i)}$ are weakly correlated after conditioning on $Z = (z, 1)$. Merlin's message U is long, but it is shared by all m query sets, so it reveals little about the random query set S_Q. Since B_Q is a function of (Z_{alice}, U), conditioning on it adds no correlation beyond what U already carries. This is again a standard technique from [KW20] and [Ko26].

Proposition 3.9 (Correlation on accepted runs)**.**

$$\mathop{\mathbb{E}}_{z\sim Z_{\text{alice}}|_{B_Q=1}} \Big[I\big(S_Q^{(i)}; X^{(i)} \mid Z = (z,1)\big)\Big] \;\leq\; \frac{n^{-2/5}}{\rho_1}.$$

Proof. We first bound the correlation before conditioning on B_Q. Since $S_Q^{(i)}$ is part of S_Q, the chain rule (Fact 2.3) and Claim 3.6 give

$$I\big(S_Q^{(i)}; X^{(i)}, U \mid Z_{\text{alice}}\big) \;\leq\; I\big(S_Q; X^{(i)}, U \mid \Phi, A_{\text{all}}, Q, S_{<Q}\big) \;=\; \frac{1}{m} I\big(\vec{S}; X^{(i)}, U \mid \Phi, A_{\text{all}}\big).$$

Since A_{all} is a function of $(\vec{S}, \Phi)$, Fact 2.4 lets us drop it from the conditioning:

$$I\big(\vec{S}; X^{(i)}, U \mid \Phi, A_{\text{all}}\big) \;\leq\; I\big(\vec{S}; X^{(i)}, U \mid \Phi\big).$$

Recall that $\Phi = (X^{(\neq i)}, \Psi)$, where Ψ is a function of (U, R_{Bob}) (Theorem 3.3(e)). Since $\vec{S}$ is independent of X, and R_{Bob} is independent of $(\vec{S}, X)$, the chain rule and Fact 2.2(e) and (f) give

$$I\big(\vec{S}; X^{(i)}, U \mid X^{(\neq i)}, \Psi\big) \;\leq\; I\big(\vec{S}; X^{(i)}, U, R_{\text{Bob}} \mid X^{(\neq i)}\big) \;=\; I\big(\vec{S}; U \mid X^{(i)}, X^{(\neq i)}\big) \;\leq\; H(U).$$

By (1), we conclude that $I(S_Q^{(i)}; X^{(i)}, U|Z_{\text{alice}}) \leq H(U)/m \leq n^{-2/5}$, since $m = n$. Next we condition on B_Q. Since B_Q is a function of (Z_{alice}, U), the chain rule gives

$$I\big(S_Q^{(i)}; X^{(i)} \mid Z_{\text{alice}}, B_Q\big) \;\leq\; I\big(S_Q^{(i)}; X^{(i)}, U, B_Q \mid Z_{\text{alice}}\big) \;=\; I\big(S_Q^{(i)}; X^{(i)}, U \mid Z_{\text{alice}}\big) \;\leq\; n^{-2/5}.$$

Finally, by Fact 2.2(b) and (c),

$$\begin{aligned} I\big(S_Q^{(i)}; X^{(i)} \mid Z_{\text{alice}}, B_Q\big) \;&=\; \mathop{\mathbb{E}}_{(z,b)\sim(Z_{\text{alice}},B_Q)} \Big[I\big(S_Q^{(i)}; X^{(i)} \mid Z = (z,b)\big)\Big] \\ &\geq\; \rho_1 \cdot \mathop{\mathbb{E}}_{z\sim Z_{\text{alice}}|_{B_Q=1}} \Big[I\big(S_Q^{(i)}; X^{(i)} \mid Z = (z,1)\big)\Big]. \end{aligned}$$

□

3.3.4 The main contradiction

We now show that these bounds lead to a contradiction. First, they give a monochromatic near-rectangle, after conditioning $S_Q^{(i)}$ on a likely event.

Lemma 3.10 (A monochromatic near-rectangle)**.** *If $\rho_1 \geq n^{-1/6}$, then there are a value $(z,1)$ of Z and a set $E \subseteq \binom{B_i}{s_i}$ with $\Pr\left[S_Q^{(i)} \in E \;\middle|\; Z=(z,1)\right] \geq 3/4$ such that, writing $\mathcal{E}$ for the event that $Z=(z,1)$ and $S_Q^{(i)} \in E$,*

(i) $H_\infty\big(S_Q^{(i)} \mid \mathcal{E}\big) \;\geq\; \log\binom{n_i}{s_i} - \dfrac{\varepsilon s_i}{2}$ *and* $H_\infty\big(X^{(i)} \mid \mathcal{E}\big) \;\geq\; \log\binom{n_i}{s_i} - \dfrac{\varepsilon s_i}{2}$, *and*

(ii) $I\big(S_Q^{(i)}; X^{(i)} \mid \mathcal{E}\big) \;\leq\; n^{-1/5}$.

In particular, the distribution of $(S_Q^{(i)}, X^{(i)})$ given $\mathcal{E}$ is a monochromatic $(\varepsilon s_i/2,\, n^{-1/5})$-near-rectangle on $\binom{B_i}{s_i} \times \binom{B_i}{s_i}$.

Proof. Draw $z \sim Z_{\text{alice}}|_{B_Q=1}$. Each of the following fails with probability at most $1/4$:

(a) there is a set E as in Proposition 3.7;

(b) $H_\infty(X^{(i)}|Z=(z,1)) \geq \log\binom{n_i}{s_i} - \varepsilon s_i/4 - \log(1/\rho_1) - 2$, by Proposition 3.8;

(c) $I(S_Q^{(i)}; X^{(i)}|Z=(z,1)) \leq 4\,n^{-2/5}/\rho_1 \leq 4\,n^{-7/30}$, by Proposition 3.9 and Markov's inequality.

So some z satisfies all three. Fix such z and E, and let $\mathcal{E}$ be the event that $Z=(z,1)$ and $S_Q^{(i)} \in E$. Given $Z=(z,1)$, the event $S_Q^{(i)} \in E$ is a function of $S_Q^{(i)}$ and has probability at least $3/4$, so Fact 2.9 applies to it.

- *Min-entropy of $S_Q^{(i)}$.* By (a), $H_\infty(S_Q^{(i)}|\mathcal{E}) \geq \log\binom{n_i}{s_i} - 32\,n^{1/5} \geq \log\binom{n_i}{s_i} - \varepsilon s_i/2$, since $s_i \geq n^{1/3}$.
- *Min-entropy of $X^{(i)}$.* By (b) and Fact 2.9, $H_\infty(X^{(i)}|\mathcal{E}) \geq \log\binom{n_i}{s_i} - \varepsilon s_i/4 - \log(1/\rho_1) - 3 \geq \log\binom{n_i}{s_i} - \varepsilon s_i/2$, since $\log(1/\rho_1) \leq (\log n)/6$.
- *Correlation.* By (c) and Fact 2.9, $I(S_Q^{(i)}; X^{(i)}|\mathcal{E}) \leq \frac{4}{3}\cdot 4\,n^{-7/30} \leq n^{-1/5}$.

Finally, by Theorem 3.3(a), $S_Q \cap X = \emptyset$ whenever $B_Q = 1$, so the distribution of $(S_Q^{(i)}, X^{(i)})$ given $\mathcal{E}$ is supported on disjoint pairs. □

Next, as in [Ko26], we replace this distribution by the product of its two marginals, and apply Lemma 2.15.

Corollary 3.11 (A nearly monochromatic rectangle)**.** *If $\rho_1 \geq n^{-1/6}$, let $\mathcal{E}$ be the event of Lemma 3.10. Then*

$$R \;:=\; S_Q^{(i)}|_{\mathcal{E}} \otimes X^{(i)}|_{\mathcal{E}}$$

is an $(\varepsilon s_i/2)$-rectangle on $\binom{B_i}{s_i} \times \binom{B_i}{s_i}$, and

$$\Pr_{(S,X)\sim R}\left[S \cap X \neq \emptyset\right] \;\leq\; n^{-1/10}.$$

Proof. By Lemma 3.10, the distribution of $(S_Q^{(i)}, X^{(i)})$ given $\mathcal{E}$ is a monochromatic $(\varepsilon s_i/2,\, n^{-1/5})$-near-rectangle, and its marginals are $S_Q^{(i)}|_{\mathcal{E}}$ and $X^{(i)}|_{\mathcal{E}}$. By Lemma 2.15 with $\delta = 0$, R is an $(\varepsilon s_i/2)$-rectangle with

$$\Pr_{(S,X)\sim R}\left[S \cap X \neq \emptyset\right] \;\le\; \sqrt{\frac{\ln 2}{2}\, n^{-1/5}} \;\le\; n^{-1/10}. \qquad \square$$

Finally, a corruption bound shows that no such rectangle exists: two independent sparse sets with high min-entropy intersect with noticeable probability. For sets of size $\sqrt{N}$ in a universe of size N, this is the rectangle bound of Babai, Frankl and Simon [BFS86]. We need it for sparser sets, and prove it in Section 4.

Corollary 3.12 (High min-entropy rectangles are not monochromatic)**.** *Let $\varepsilon := 1/40$ and $c := 1/100$. Let $s \ge 200$, let N be a multiple of s^2, let $\mathcal{S} := \binom{[N]}{s}$, and let ν_A, ν_B be distributions on $\mathcal{S}$ with*

$$H_\infty(\nu_A),\, H_\infty(\nu_B) \;\ge\; \log|\mathcal{S}| - (\varepsilon s - 1).$$

Then $\Pr_{a\sim\nu_A,\, b\sim\nu_B}[a \cap b \neq \emptyset] \ge cs^2/N$.

The following lemma completes the proof of the communication lower bound.

Lemma 3.13 (Communication lower bound)**.** *For every $i \in \mathcal{I}$, the protocol G_i of Theorem 3.3 has $\rho_1 < n^{-1/6}$.*

Proof. Suppose, for contradiction, that $\rho_1 \ge n^{-1/6}$. By Corollary 3.11, R is an $(\varepsilon s_i/2)$-rectangle on $\binom{B_i}{s_i} \times \binom{B_i}{s_i}$ that puts mass at most $n^{-1/10}$ on intersecting pairs. On the other hand, apply Corollary 3.12 with $N := n_i$ and $s := s_i$, identifying B_i with $[n_i]$; this is allowed since $n_i = \ell s_i^2$, $s_i \ge n^{1/3} \ge 200$ and $\varepsilon s_i/2 \le \varepsilon s_i - 1$. So R puts mass at least $cs_i^2/n_i = c/\ell \ge c/\log n > n^{-1/10}$ on intersecting pairs, a contradiction. $\square$

To see how Lemma 3.13 completes the proof of Theorem 3.1, recall that $\ell = \Theta(\log n/\log(t_u w))$ by Claim 3.2(a), and suppose that $t_q = o(\ell^2)$. Since $\sum_{i\in\mathcal{I}} \tau_i \le \mathbb{E}_{\mathcal{D},Q}[T(Q)] \le t_q$ and $|\mathcal{I}| \ge \ell/3 - 1$, some $i \in \mathcal{I}$ has $\tau_i \le t_q/|\mathcal{I}| = o(\ell)$. For this i, Corollary 3.5 gives $\rho_1 \ge n^{-o(1)} \ge n^{-1/6}$, whereas Lemma 3.13 shows that $\rho_1 < n^{-1/6}$. This gives the claimed lower bound on t_q.

4 A corruption bound for sparse sets

This section is purely combinatorial and independent of the data structure. Its goal is to show that large rectangles of sparse sets, under the uniform distribution over them, are not nearly monochromatic for the answer "disjoint". For sets of size $\sqrt{N}$ in a universe of size N, this is exactly the rectangle bound in the seminal work of Babai, Frankl and Simon [BFS86]. We adapt their argument to sparser regimes, at the cost of a smaller error tolerance.

To state the bound in generality, we consider a universe of size N and sets of size s, where N is a multiple of s^2, and write $\mathcal{S} := \binom{[N]}{s}$ for the family of s-subsets of $[N]$. The bound of Babai, Frankl and Simon is exactly the case $N = s^2$. In this section, a and b denote sets of size s, and $\mathcal{A}$ and $\mathcal{B}$ families of such sets. We write $\mathcal{U}_{\mathcal{A}}$ and $\mathcal{U}_{\mathcal{B}}$ for the uniform distributions on $\mathcal{A}$ and $\mathcal{B}$, and $\mathcal{U}$ for the uniform distribution on $\mathcal{S}$. The key lemma of this section is the following corruption bound.

Lemma 4.1 (Corruption for sparse sets)**.** *Let $\varepsilon := 1/40$ and $c := 1/100$. Let $s \geq 200$, and let N be a multiple of s^2. If $\mathcal{A}, \mathcal{B} \subseteq \mathcal{S}$ satisfy $|\mathcal{A}|, |\mathcal{B}| \geq 2^{-\varepsilon s} |\mathcal{S}|$, then*

$$\Pr_{\substack{a \sim \mathcal{U}_{\mathcal{A}} \\ b \sim \mathcal{U}_{\mathcal{B}}}} [a \cap b \neq \emptyset] \;\geq\; c \cdot \frac{s^2}{N}$$

Before we delve into the proof of Lemma 4.1, we derive from it the corruption bound in terms of min-entropy that we used in Step 4 (Section 3.3.4), by a simple application of Corollary 2.14. We restate it for convenience.

Corollary 3.12. Let $s \geq 200$, let N be a multiple of s^2, and let ν_A, ν_B be distributions on $\mathcal{S}$ with

$$H_\infty(\nu_A),\, H_\infty(\nu_B) \;\geq\; \log |\mathcal{S}| - (\varepsilon s - 1).$$

Then $\Pr_{a \sim \nu_A,\, b \sim \nu_B}[a \cap b \neq \emptyset] \geq cs^2/N$.

Proof. The product $\nu_A \otimes \nu_B$ is an $(\varepsilon s - 1)$-rectangle. Since $|\mathcal{S}| \geq 2^{\varepsilon s}$, some integer K satisfies $2^{-\varepsilon s} |\mathcal{S}| \leq K \leq 2^{1-\varepsilon s} |\mathcal{S}|$. By Corollary 2.14 with this K, the product is a convex combination of uniform distributions on combinatorial rectangles $\mathcal{A} \times \mathcal{B}$ with $|\mathcal{A}| = |\mathcal{B}| = K \geq 2^{-\varepsilon s} |\mathcal{S}|$. By Lemma 4.1, each of them puts mass at least cs^2/N on intersecting pairs. The mass on a fixed set of pairs is linear in the distribution, so $\nu_A \otimes \nu_B$ does too. □

The bound is one-sided: the analogous bound for disjoint pairs is false, by the classic example showing that Disjointness has large discrepancy. If $\log(N/s) \leq \varepsilon s$, as in our application, then the family $\mathcal{A}$ of sets that contain a fixed element has $|\mathcal{A}| = (s/N) |\mathcal{S}| \geq 2^{-\varepsilon s} |\mathcal{S}|$, but all pairs in $\mathcal{A} \times \mathcal{A}$ intersect. This is why the main proof uses only runs whose answer is "disjoint".

4.1 Counting facts

Just like in the proof of [BFS86], we list some counting facts for the proof. Write $H_{\mathrm{b}}(\alpha) := -\alpha \log \alpha - (1-\alpha) \log(1-\alpha)$ for the binary entropy function.

Fact 4.2 (Binomial estimates)**.** *For all integers $u \geq 0$ and $k \geq 1$, $\binom{u}{k} \leq \left(\frac{eu}{k}\right)^k$. For every integer $g \geq 1$ and every $0 < \alpha \leq 1/2$,*

$$\sum_{0 \leq h \leq \alpha g} \binom{g}{h} \;\leq\; 2^{H_{\mathrm{b}}(\alpha) g}.$$

For $a \sim \mathcal{U}$, the probability that a contains a fixed set or avoids a fixed set can be read off binomial coefficients. For $T, V \subseteq [N]$ with $|T| = k \leq s$,

$$\Pr_{a \sim \mathcal{U}}[T \subseteq a] \;=\; \binom{N-k}{s-k} \Big/ \binom{N}{s} \;=\; \prod_{h=0}^{k-1} \frac{s-h}{N-h} \;\leq\; \left(\frac{s}{N}\right)^k,$$

$$\Pr_{a \sim \mathcal{U}}[a \cap V = \emptyset] \;=\; \binom{N-|V|}{s} \Big/ \binom{N}{s} \;=\; \prod_{h=0}^{s-1} \frac{N-|V|-h}{N-h} \;\leq\; \left(1 - \frac{|V|}{N}\right)^s,$$

and $\Pr[T \subseteq a] = 0$ if $|T| > s$. We use the following two consequences.

Claim 4.3 (Two counting facts)**.** *Let $1 \leq s \leq N$.*

(a) For every $P \subseteq [N]$ and every integer $k \geq 1$, $\Pr_{a \sim \mathcal{U}}\left[\,|a \cap P| \geq k\right] \leq \binom{|P|}{k} \left(\frac{s}{N}\right)^k \leq \left(\frac{e|P|s}{kN}\right)^k$.

(b) For every $V \subseteq [N]$, $\Pr_{a\sim\mathcal{U}}[a \cap V = \emptyset] \le \left(1 - \frac{|V|}{N}\right)^s \le e^{-s|V|/N}$.

Proof. (a) If $|a \cap P| \ge k$, then a contains some $T \subseteq P$ with $|T| = k$. By a union bound over the $\binom{|P|}{k}$ choices of T, the bound $\Pr[T \subseteq a] \le (s/N)^k$, and Fact 4.2,

$$\Pr_{a\sim\mathcal{U}}\left[\,|a \cap P| \ge k\right] \;\le\; \Pr_{a\sim\mathcal{U}}\left[\exists\, T \subseteq P \text{ with } |T| = k \text{ and } T \subseteq a\right] \;\le\; \sum_{\substack{T\subseteq P\\ |T|=k}} \Pr_{a\sim\mathcal{U}}[T \subseteq a]$$

$$\le \binom{|P|}{k}\left(\frac{s}{N}\right)^k \;\le\; \left(\frac{e\,|P|\,s}{kN}\right)^k.$$

(b) By the second display above, and since each factor satisfies $\frac{N-|V|-h}{N-h} \le 1 - \frac{|V|}{N}$ and $1 - x \le e^{-x}$,

$$\Pr_{a\sim\mathcal{U}}[a \cap V = \emptyset] \;=\; \binom{N-|V|}{s}\Big/\binom{N}{s} \;=\; \prod_{h=0}^{s-1}\frac{N-|V|-h}{N-h}$$

$$\le \left(1 - \frac{|V|}{N}\right)^s \;\le\; e^{-s|V|/N}.$$

□

4.2 Proof of the corruption bound

For $N = s^2$, Lemma 4.1 is the rectangle bound behind the $\Omega(\sqrt{N})$ lower bound for Disjointness of [BFS86] (see also [CP10, Section 3.2, proof of Theorem 6]). Our proof generalizes their argument, so we first summarize the high-level argument of [BFS86]. Suppose that $\mathcal{A} \times \mathcal{B}$ is large and nearly disjoint.

- *A disjoint chain in $\mathcal{A}$:* Since $\mathcal{A}$ is large, it contains rows $a_1, \ldots, a_t$ with $t = \Theta(s)$, each of which has $s/2$ elements outside the earlier rows. These pieces of new elements are pairwise disjoint and together cover a constant fraction of $[N]$.
- *Columns of $\mathcal{B}$ avoid the chain:* A typical row meets few columns of $\mathcal{B}$, so a typical column of $\mathcal{B}$ meets few of the pieces. That is, it lies in the complement of all but a few of the pieces.
- *Few sets avoid the chain:* A uniform s-set meets each piece with constant probability. A union bound over which pieces it meets shows that only a $2^{-\Omega(s)}$ fraction of all s-sets meet so few pieces. Hence $\mathcal{B}$ is small, a contradiction.

For sparser sets, i.e. when N is a multiple of s^2, we make two changes.

- *A longer chain:* We take $t = N/(6s)$ rows instead of $\Theta(s)$; the two agree up to a constant factor when $N = s^2$. The t rows still cover at most $N/6$ elements, so the next row with many new elements can still be found, and the pieces cover $N/12$ elements.
- *Grouping the pieces:* We group the t pieces into $g = s/6$ groups of N/s^2 pieces each, and apply the union bound to groups instead of pieces. A group has $N/(2s)$ elements, so a uniform s-set meets it with constant probability, just as it meets a single piece when $N = s^2$. For $N = s^2$, every group is a single piece.

Proof of Lemma 4.1. For brevity set $\delta := cs^2/N$. Recall that $|\mathcal{A}|, |\mathcal{B}| \geq 2^{-\varepsilon s}|\mathcal{S}|$. Suppose, for contradiction, that

$$\Pr_{a\sim\mathcal{U}_\mathcal{A}, b\sim\mathcal{U}_\mathcal{B}}[a \cap b \neq \emptyset] \;<\; \delta.$$

We show that $|\mathcal{B}| < 2^{-\varepsilon s}|\mathcal{S}|$.

Good rows. Let $\mathcal{A}' := \{a \in \mathcal{A} : \Pr_{b\sim\mathcal{U}_\mathcal{B}}[a \cap b \neq \emptyset] \leq 2\delta\}$. By Markov's inequality, $|\mathcal{A}'| \geq |\mathcal{A}|/2 \geq 2^{-\varepsilon s-1}|\mathcal{S}|$.

A chain of rows with new elements. Let $t := N/(6s)$. We select a chain of t rows in $\mathcal{A}'$, each of which brings $s/2$ new elements.

Claim 4.4 (A chain of rows with new elements)**.** *There are $a_1, \ldots, a_t \in \mathcal{A}'$ and sets $D_1, \ldots, D_t$ such that, for every $j \in [t]$,*

$$D_j \;\subseteq\; a_j \setminus (a_1 \cup \cdots \cup a_{j-1}) \qquad \text{and} \qquad |D_j| = s/2.$$

In particular, $D_1, \ldots, D_t$ are pairwise disjoint.

Proof. We choose $a_1, a_2, \ldots$ greedily, one at a time. Let $1 \leq j \leq t$, suppose that $a_1, \ldots, a_{j-1}$ have been chosen, and let $P_j := a_1 \cup \cdots \cup a_{j-1}$ (so $P_1 = \emptyset$). It suffices to find $a_j \in \mathcal{A}'$ with $|a_j \cap P_j| < s/2$: then $|a_j \setminus P_j| > s/2$, and we let D_j be any $s/2$ elements of $a_j \setminus P_j$.

Each earlier row has s elements, so $|P_j| \leq (j-1)s < ts = N/6$. Hence only few sets have $s/2$ or more elements in P_j: by Claim 4.3(a) with $k = s/2$,

$$\Pr_{a\sim\mathcal{U}}\big[\,|a \cap P_j| \geq s/2\big] \;\leq\; \Big(\frac{e\,|P_j|\;s}{(s/2)\,N}\Big)^{s/2} \;\leq\; \Big(\frac{e}{3}\Big)^{s/2} \;\leq\; 2^{-0.071s}.$$

On the other hand, $|\mathcal{A}'| \geq 2^{-\varepsilon s-1}|\mathcal{S}| > 2^{-0.071s}|\mathcal{S}|$. So $\mathcal{A}'$ is not contained in the family of sets with $s/2$ or more elements in P_j, and some $a_j \in \mathcal{A}'$ has $|a_j \cap P_j| < s/2$. Finally, $D_j \subseteq a_j \setminus P_j$ is disjoint from $D_1, \ldots, D_{j-1} \subseteq P_j$. □

Grouping the pieces. Let $g := s/6$, and partition $[t]$ into g sets $\Gamma_1, \ldots, \Gamma_g$, each of size $t/g = N/s^2$. For $u \in [g]$, let

$$E_u \;:=\; \bigcup_{j\in\Gamma_u} D_j.$$

The sets $E_1, \ldots, E_g$ are pairwise disjoint, and each has $\frac{t}{g} \cdot \frac{s}{2} = \frac{N}{2s}$ elements.

Most columns miss most of the groups. Since every a_j lies in $\mathcal{A}'$,

$$\mathop{\mathbb{E}}_{b\sim\mathcal{U}_\mathcal{B}}\big[\;|\{j \in [t] : a_j \cap b \neq \emptyset\}|\;\big] \;=\; \sum_{j=1}^{t} \Pr_{b\sim\mathcal{U}_\mathcal{B}}[a_j \cap b \neq \emptyset] \;\leq\; 2\delta t,$$

so by Markov's inequality the set $\mathcal{B}'$ of $b \in \mathcal{B}$ that meet at most $4\delta t$ of the rows a_j satisfies $|\mathcal{B}'| \geq |\mathcal{B}|/2$. If b meets E_u, then it meets $a_j \supseteq D_j$ for some $j \in \Gamma_u$, and the sets Γ_u are disjoint. So every $b \in \mathcal{B}'$ meets at most $4\delta t$ of the sets E_u, and by our choice of parameters

$$4\delta t \;=\; 4 \cdot \frac{cs^2}{N} \cdot \frac{N}{6s} \;=\; \frac{2cs}{3} \;=\; \frac{s}{150} \;=\; \frac{g}{25}.$$

Counting $\mathcal{B}'$. Every $b \in \mathcal{B}'$ is determined by the set of $h \leq g/25$ groups it meets, together with an s-set that avoids the union V of the remaining $g - h$ groups. This union has $|V| = (g-h)\frac{N}{2s}$

elements, so $s\,|V|\,/N = \frac{g-h}{2} \geq \frac{24}{25} \cdot \frac{g}{2} = \frac{2s}{25}$. Hence, by Claim 4.3(b) and Fact 4.2,

$$\begin{aligned} |\mathcal{B}'| \;&\leq\; \sum_{h \leq g/25} \binom{g}{h} \binom{N-(g-h)\frac{N}{2s}}{s} \;\leq\; \sum_{h \leq g/25} \binom{g}{h} e^{-2s/25}\,|\mathcal{S}| \\ &\leq\; 2^{H_{\mathrm{b}}(1/25)\,g} \cdot e^{-2s/25}\,|\mathcal{S}| \;\leq\; 2^{0.0404s} \cdot 2^{-0.1154s}\,|\mathcal{S}| \;=\; 2^{-0.075s}\,|\mathcal{S}|\,, \end{aligned}$$

where we used $H_{\mathrm{b}}(1/25) \leq 0.2423$ and $g = s/6$. Finally, $|\mathcal{B}| \leq 2\,|\mathcal{B}'| \leq 2^{-0.075s+1}\,|\mathcal{S}| < 2^{-\varepsilon s}\,|\mathcal{S}|$. This contradicts $|\mathcal{B}| \geq 2^{-\varepsilon s}\,|\mathcal{S}|$. □

AI methodology

The proof of the technical Lemma 4.1 (Section 4.2) was developed with the assistance of an AI model (Claude Opus 5.5), starting from the author's prompt to adapt the corruption argument of Babai, Frankl and Simon [BFS86] by changing its parameters. The underlying argument is essentially [BFS86]; the modifications are a longer chain of rows and a union bound over groups of pieces rather than individual pieces. The author has verified the proof in full and takes responsibility for its correctness.

References

[AW14] A. Abboud and V. V. Williams. Popular Conjectures Imply Strong Lower Bounds for Dynamic Problems. In *2014 IEEE 55th Annual Symposium on Foundations of Computer Science*, pages 434–443, October 2014.

[AW26] Josh Alman and Virginia Vassilevska Williams. Truly Subquadratic 3SUM and Truly Subcubic APSP via Triangles in Sparse Lopsided Graphs, October 2026. arXiv:2610.06783 [cs.DS].

[BFS86] László Babai, Peter Frankl, and Janos Simon. Complexity classes in communication complexity theory (preliminary version). In *27th Annual Symposium on Foundations of Computer Science*, pages 337–347, 1986.

[BL15] Joshua Brody and Kasper Green Larsen. Adapt or Die: Polynomial Lower Bounds for Non-Adaptive Dynamic Data Structures. *Theory of Computing*, 11(19), 2015.

[CEEP16] Arkadev Chattopadhyay, Jeff Edmonds, Faith Ellen, and Toniann Pitassi. Upper and Lower Bounds on the Power of Advice. *SIAM Journal on Computing*, 45(4):1412–1432, January 2016.

[CGL15] Raphael Clifford, A. Grønlund, and K. G. Larsen. New Unconditional Hardness Results for Dynamic and Online Problems. In *2015 IEEE 56th Annual Symposium on Foundations of Computer Science*, pages 1089–1107, October 2015.

[CP10] Arkadev Chattopadhyay and Toniann Pitassi. The story of set disjointness. *ACM SIGACT News*, 41(3):59–85, September 2010.

[CT06] Thomas M. Cover and Joy A. Thomas. *Elements of Information Theory (Wiley Series in Telecommunications and Signal Processing)*. Wiley-Interscience, New York, NY, USA, 2006.

[DGW19] Zeev Dvir, Alexander Golovnev, and Omri Weinstein. Static data structure lower bounds imply rigidity. In *Proceedings of the 51st Annual ACM SIGACT Symposium on Theory of Computing*, STOC 2019, pages 967–978, Phoenix, AZ, USA, June 2019. Association for Computing Machinery.

[DI04] Camil Demetrescu and Giuseppe F. Italiano. A new approach to dynamic all pairs shortest paths. *Journal of the ACM*, 51(6):968–992, November 2004.

[DL20] Pavel Dvořák and Bruno Loff. Lower Bounds for Semi-adaptive Data Structures via Corruption. *LIPIcs, Volume 182, FSTTCS 2020*, 182:20:1–20:15, 2020.

[DORS08] Yevgeniy Dodis, Rafail Ostrovsky, Leonid Reyzin, and Adam Smith. Fuzzy Extractors: How to Generate Strong Keys from Biometrics and Other Noisy Data. *SIAM Journal on Computing*, 38(1):97–139, January 2008.

[FS89] Michael L. Fredman and Michael E. Saks. The Cell Probe Complexity of Dynamic Data Structures. In *Proceedings of the 21st Annual ACM Symposium on Theory of Computing*, pages 345–354, 1989.

[HDLT01] Jacob Holm, Kristian De Lichtenberg, and Mikkel Thorup. Poly-logarithmic deterministic fully-dynamic algorithms for connectivity, minimum spanning tree, 2-edge, and biconnectivity. *Journal of the ACM*, 48(4):723–760, July 2001.

[HF98] M. R. Henzinger and M. L. Fredman. Lower Bounds for Fully Dynamic Connectivity Problems in Graphs. *Algorithmica*, 22(3):351–362, November 1998.

[HKNS15] Monika Henzinger, Sebastian Krinninger, Danupon Nanongkai, and Thatchaphol Saranurak. Unifying and Strengthening Hardness for Dynamic Problems via the Online Matrix-Vector Multiplication Conjecture. In *Proceedings of the Forty-seventh Annual ACM Symposium on Theory of Computing*, STOC '15, pages 21–30, New York, NY, USA, 2015. ACM.

[IPZ01] Russell Impagliazzo, Ramamohan Paturi, and Francis Zane. Which Problems Have Strongly Exponential Complexity? *Journal of Computer and System Sciences*, 63(4):512–530, December 2001.

[JMV15] Hamid Jahanjou, Eric Miles, and Emanuele Viola. Local Reductions. In Magnús M. Halldórsson, Kazuo Iwama, Naoki Kobayashi, and Bettina Speckmann, editors, *Automata, Languages, and Programming*, volume 9134 of *Lecture Notes in Computer Science*, pages 749–760. Springer Berlin Heidelberg, Berlin, Heidelberg, 2015.

[JRS09] Rahul Jain, Jaikumar Radhakrishnan, and Pranab Sen. A property of quantum relative entropy with an application to privacy in quantum communication. *Journal of the ACM*, 56(6):1–32, September 2009.

[Kin99] V. King. Fully dynamic algorithms for maintaining all-pairs shortest paths and transitive closure in digraphs. In *40th Annual Symposium on Foundations of Computer Science*, pages 81–89, New York City, NY, USA, 1999. IEEE Comput. Soc.

[KKM13] Bruce M. Kapron, Valerie King, and Ben Mountjoy. Dynamic graph connectivity in polylogarithmic worst case time. In *Proceedings of the Twenty-Fourth Annual ACM-SIAM Symposium on Discrete Algorithms*, pages 1131–1142. Society for Industrial and Applied Mathematics, January 2013.

[KLMS26] Michal Koucký, Bruno Loff, Tulasimohan Molli, and Michael E. Saks. The Natural Proofs Barrier against Data-Structure Lower-Bounds. In *Proceedings of the 58th Annual ACM Symposium on Theory of Computing*, pages 1338–1343, Salt Lake City UT USA, June 2026. ACM.

[Ko26] Young Kun Ko. An $\Omega((\log n / \log \log n)ˆ2)$ Cell-Probe Lower Bound for Dynamic Boolean Data Structures. In *Proceedings of the 67th IEEE Symposium on Foundations of Computer Science (FOCS) (to appear)*. IEEE, 2026.

[KPI25] Oliver Korten, Toniann Pitassi, and Russell Impagliazzo. Stronger Cell Probe Lower Bounds via Local PRGs. In *2025 IEEE 66th Annual Symposium on Foundations of Computer Science (FOCS)*, pages 2480–2504, Sydney, Australia, December 2025. IEEE.

[KW20] Young Kun Ko and Omri Weinstein. An Adaptive Step Toward the Multiphase Conjecture. In *2020 IEEE 61st Annual Symposium on Foundations of Computer Science (FOCS)*, pages 752–761, Durham, NC, USA, November 2020. IEEE.

[Lar12] Kasper Green Larsen. The Cell Probe Complexity of Dynamic Range Counting. In *Proceedings of the Forty-fourth Annual ACM Symposium on Theory of Computing*, STOC '12, pages 85–94, New York, NY, USA, 2012. ACM.

[LWY20] Kasper Green Larsen, Omri Weinstein, and Huacheng Yu. Crossing the Logarithmic Barrier for Dynamic Boolean Data Structure Lower Bounds. *SIAM Journal on Computing*, 49(5):STOC18–323–STOC18–367, January 2020.

[LY25] Kasper Green Larsen and Huacheng Yu. Super-Logarithmic Lower Bounds for Dynamic Graph Problems. *SIAM Journal on Computing*, pages FOCS23–42–FOCS23–69, February 2025.

[MSVT94] Peter Bro Miltersen, Sairam Subramanian, Jeffrey Scott Vitter, and Roberto Tamassia. Complexity models for incremental computation. *Theoretical Computer Science*, 130(1):203–236, August 1994.

[Pat10] Mihai Patrascu. Towards polynomial lower bounds for dynamic problems. In Leonard J. Schulman, editor, *Proceedings of the 42nd ACM Symposium on Theory of Computing, STOC 2010, Cambridge, Massachusetts, USA, 5-8 June 2010*, pages 603–610. ACM, 2010.

[PD06] Mihai Patrascu and Erik D. Demaine. Logarithmic Lower Bounds in the Cell-Probe Model. *SIAM J. Comput.*, 35(4):932–963, April 2006.

[San04] P. Sankowski. Dynamic Transitive Closure via Dynamic Matrix Inverse (Extended Abstract). In *45th Annual IEEE Symposium on Foundations of Computer Science*, pages 509–517, Rome, Italy, 2004. IEEE.

[Tho05] Mikkel Thorup. Worst-case update times for fully-dynamic all-pairs shortest paths. In *Proceedings of the thirty-seventh annual ACM symposium on Theory of computing*, pages 112–119, Baltimore MD USA, May 2005. ACM.

[Vad12] Salil P. Vadhan. Pseudorandomness. *Foundations and Trends® in Theoretical Computer Science*, 7(1–3):1–336, 2012.

[VDBNS19] Jan Van Den Brand, Danupon Nanongkai, and Thatchaphol Saranurak. Dynamic Matrix Inverse: Improved Algorithms and Matching Conditional Lower Bounds. In *2019 IEEE 60th Annual Symposium on Foundations of Computer Science (FOCS)*, pages 456–480, Baltimore, MD, USA, November 2019. IEEE.

[Vio18] Emanuele Viola. Lower bounds for data structures with space close to maximum imply circuit lower bounds. In *ECCC*, volume 25, 2018.

[Wil13] Ryan Williams. Improving Exhaustive Search Implies Superpolynomial Lower Bounds. *SIAM J. Comput.*, 42(3):1218–1244, 2013.

[Yao81] Andrew Chi-Chih Yao. Should Tables Be Sorted? *J. ACM*, 28(3):615–628, July 1981.

A Reductions to dynamic graph problems

In this appendix, we derive lower bounds for dynamic graph problems from Theorem 3.1, by a reduction of Pătraşcu [Pat10]. A single reduction works for all of the following problems. In each, a data structure maintains a graph on a fixed set of N vertices.

- *Incremental directed reachability.* The graph is directed and initially has no edges. An update inserts an edge, and a query asks whether b is reachable from a.
- *Incremental undirected shortest paths.* The graph is undirected and initially has no edges. An update inserts an edge, and a query asks for the distance between a and b, or an approximation of it within a factor smaller than 2.
- *Incremental subgraph connectivity.* An undirected graph is fixed in advance, and every vertex is on or off. An update turns a vertex on, and a query asks whether a and b are connected in the subgraph induced by the vertices that are on.

The reduction. Given sets $S_1, \ldots, S_n \subseteq [n]$ and $X \subseteq [n]$, let G be the graph with vertices $u_1, \ldots, u_n$, $v_1, \ldots, v_n$ and t, and with the edges $\{u_q, v_j\}$ for $j \in S_q$ and $\{v_j, t\}$ for $j \in X$; for directed reachability, the edges are directed from u_q to v_j and from v_j to t. The three phases of multiphase Disjointness become the following operations.

(i) *Phase I.* Insert the edges $\{u_q, v_j\}$ for $j \in S_q$. For subgraph connectivity, the fixed graph has these edges and the edges $\{v_j, t\}$ for all $j \in [n]$, and initially only $u_1, \ldots, u_n$ and t are on.

(ii) *Phase II.* When j is inserted into X, insert the edge $\{v_j, t\}$, or, for subgraph connectivity, turn v_j on.

(iii) *Phase III.* On query q, ask about the pair (u_q, t).

In all three problems, the answer determines whether $S_q \cap X = \emptyset$. Every neighbour of u_q is some v_j with $j \in S_q$, and v_j is adjacent to t (or on) exactly when $j \in X$. Hence t is reachable from u_q, and the distance between u_q and t is 2, exactly when $S_q \cap X \neq \emptyset$; otherwise the distance is at least 4, since G is bipartite with u_q and t on the same side. In the subgraph of the vertices that are on, every neighbour v_j of u_q has $j \in X$ and is adjacent to t, so u_q and t are connected exactly when $S_q \cap X \neq \emptyset$.

Phase I is preprocessing, which the multiphase model does not charge. Each insertion into X is a single update, and the query is a single query, whose writes can be kept locally since it is the last

operation. So a dynamic data structure with worst-case update time t_u and expected query time t_q yields a data structure for insert-only multiphase Disjointness with the same word size, worst-case update time and expected query time, and Theorem 3.1 gives the following.

Theorem A.1 (Dynamic graph problems)**.** *Every zero-error cell-probe data structure for incremental directed reachability, incremental undirected shortest paths or incremental subgraph connectivity on N vertices, with word size $w = \Theta(\log N)$, worst-case update time t_u and expected query time t_q, has*

$$t_q \;\geq\; \Omega\left(\left(\frac{\log N}{\log(t_u w)}\right)^2\right), \qquad \textit{and in particular} \qquad \max(t_u, t_q) \;\geq\; \Omega\left(\left(\frac{\log N}{\log\log N}\right)^2\right).$$

For directed reachability, this holds already on depth-2 DAGs whose queries ask whether a fixed sink is reachable.

Proof. Let $n := (N-1)/2$, so that G has N vertices, and $\log n = \Theta(\log N)$. The first bound follows from the reduction and Theorem 3.1. For the second, if $t_u \geq \log^2 N$ there is nothing to prove, and otherwise $\log(t_u w) = O(\log\log N)$. □

ECCC ISSN 1433-8092
https://eccc.weizmann.ac.il