Electronic Colloquium on Computational Complexity, Report No. 155 (2025)

Lower Bounds for Linear Operators

Young Kun Ko

Department of Computer Science and Engineering, Pennsylvania State University

Email: ykko@psu.edu

October 22, 2025

Abstract

We consider a static data structure problem of computing a linear operator under cell-probe model. Given a linear operator $M \in \mathbb{F}_2^{m \times n}$, the goal is to pre-process a vector $X \in \mathbb{F}_2^n$ into a data structure of size s to answer any query $\langle M_i, X \rangle$ in time t. We prove that for a random operator M, any such data structure requires:

$$t \geq \Omega(\min\{\log(m/s), n/\log s\}).$$

This result overcomes the well-known logarithmic barrier in static data structures [MNSW98, Sie04, PD06, PTW08, Pă11, DGW19] by using a random linear operator. Furthermore, it provides the first significant progress toward confirming a decades-old folklore conjecture: that non-linear pre-processing does not substantially help in computing most linear operators.

A straightforward modification of our proof also yields a wire lower bound of $\Omega(n \cdot \log^{1/d}(n))$ for depth-d circuits with arbitrary gates that compute a specific linear operator $M \in \mathbb{F}_2^{O(n) \times n}$, even against some small constant advantage over random guessing. This bound holds even for circuits with only a small constant advantage over random guessing, improving upon longstanding results [RS03, Che08a, Che08b, GHK$^+$13] for a random operator.

Finally, our work partially resolves the communication form of the Multiphase Conjecture [Pat10] and makes progress on Jukna-Schnitger's Conjecture [JS11, Juk12]. We address the former by considering the Inner Product (mod 2) problem (instead of Set Disjointness) when the number of queries m is super-polynomial (e.g., $2^{n^{1/3}}$), and the total update time is $m^{0.99}$. Our result for the latter also applies to cases with super-polynomial m.

ISSN 1433-8092

1 Introduction

A fundamental, long-standing question in the theory of computation is:

Can non-linear computation provide an advantage in computing a linear operator?

The intuition that non-linear computation offers no significant help has persisted for over 70 years since the seminal work in circuit design by Shannon [Sha49]. For instance, Lupanov [Lup56] and later Valiant [Val77, Val92] drew on this idea to study the complexity of linear operators. This belief underpins major open problems like matrix rigidity–the search for a linear operator that cannot be computed by small, shallow circuits with linear gates (See the surveys [Lok09, Ram20] and references therein). Despite its widespread acceptance, this intuition has remained a folklore belief. Only recently did Jukna and Schnitger [JS11, Juk12] formalize it as an informal conjecture: **non-linear gates do not help in computing a linear operator**.

This conjecture extends naturally to the **cell-probe model** [Yao81], the most powerful model of computation for data structures. In this model, we only count memory accesses (probes), while all computation is free. An input is pre-processed into a data structure of s cells (here, we assume 1-bit cells, as we work over $\mathbb{F}_2$). To answer a query, an algorithm can probe up to t cells and perform arbitrary computation on their contents. Because of its strength, a lower bound in this model applies to any reasonable data structure.

This raises an analogous question for data structures: for a cell-probe data structure designed to compute $\langle M_i, X\rangle$ for a linear operator M, can non-linear pre-processing of the input X reduce the query time? This can be seen as a cell-probe analogue of the Jukna-Schnitger conjecture (See Section 2 of [DGW19] for detailed discussion).

Assuming this intuition is true has led to fruitful research on restricted models where pre-processing is linear or from related classes (See [Fre81, Cha90, Lar14, Aga17, DGW19, AFKL19, GPRW22] and references therein). These models are also deeply connected to matrix rigidity [DGW19, NRR20]. However, challenging this intuition, there are cases where highly efficient non-linear data structures exist for linear problems with no known linear equivalents [KU11].

Our primary goal is to rigorously establish that for **most** linear operators, **non-linear pre-processing provides no significant advantage**. This, in turn, implies an analogous result for circuit complexity.

Previous Results Progress on this front has been stalled by technical barriers in the cell-probe model. The primary tool for static cell-probe lower bounds is essentially a counting argument. A seminal result by Miltersen [Mil93] shows that for almost all non-linear problems, if the space s is slightly smaller than the number of possible outputs m (e.g., $s = m^{0.99}$), then the query time must be large ($t \geq \Omega(n^{0.99})$). If we *assume* linear pre-processing is optimal for linear problems, a similar counting argument yields strong lower bounds.

However, this argument breaks down completely when arbitrary non-linear pre-processing is allowed. The number of possible functions to compute even a single pre-processed bit is astronomical (2^{2^n}), rendering a simple counting argument over all possible data structures useless.

Prior to our work, the best lower bound for an arbitrary linear problem was derived from techniques used for explicit data structure problems [MNSW98, Sie04, PD06, PTW08, Pă11]. These techniques hit a ceiling known as the logarithmic barrier, yielding bounds of the form:

$$t \geq \Omega\left(\frac{\log \frac{|\mathcal{Q}|}{n}}{\log \frac{s}{n}}\right). \tag{1}$$

where $|\mathcal{Q}| = m$ is the number of distinct queries. Breaking this barrier for an explicit data structure is a holy-grail problem, with connections to major open questions in branching programs and circuit complexity [MNSW98, DGW19, Vio18].

In summary, it was previously unknown if any linear problem was truly hard against non-linear data structures. For decades, it was not ruled out that highly efficient non-linear data structures (e.g., with $t = O(1)$ and $s = |\mathcal{Q}|^{0.1}$) might exist for all linear problems.

1.1 Further Connections

1.1.1 Multiphase Program: Communication versus Cell-Probe

A more recent motivation for our work is its connection to the communication version of the Multiphase Conjecture [Pat10, Tho13, Bra22], a problem in 3-player Number-on-Forehead (NOF) communication.

Definition 1.1 (Multiphase Communication Game [Pat10])**.** *The 3-player Number-On-Forehead (NOF) communication game is defined as follows*

- *The inputs are distributed as follows: Merlin receives the linear operator $M \in \{0,1\}^{m \times n}$ and the vector $X \in \{0,1\}^n$. For a given query index $i \in [m]$, Alice receives M and i, while Bob receives T and i.*
- *Merlin sends a single-shot message of length $m^{0.99}$ to Bob.*
- *Alice and Bob proceed in standard two-party communication to output $\langle M_i, X \rangle$ [1] for any given $i \in [m]$.*

The Multiphase Conjecture (Communication Version) then states that for some $m = \text{poly}(n)$, Alice and Bob then must communicate n^{ε} for some constant $\varepsilon > 0$. The original motivation in [Pat10] behind the conjecture is its consequence in dynamic cell-probe lower bound.

Definition 1.2 (Multiphase Problem)**.** *Consider the following explicit dynamic data structure problem*

- *M is given as a pre-processing input. Pre-process the data structure using $|M| \cdot t_u$ time.*
- *Then X as a sequence is given as updates, updating the data structure using $|X| \cdot t_u$ total time.*
- *For any $i \in [m]$, the data structure must be able to output $\langle M_i, X \rangle$ in t_q time.*

The key observation in [Pat10] is that a polynomial lower bound for the Multiphase Game would imply a polynomial lower bound on $\max\{t_u, t_q\}$ for the Multiphase Problem, which is also known as the Multiphase Conjecture (Cell-Probe Version). The word "Multiphase Conjecture" has been used interchangeably to denote either the communication version of the conjecture or the cell-probe version of the conjecture.

To avoid any confusion, we will denote the communication conjecture as Multiphase Conjecture (Communication), the dynamic data structure conjecture as Multiphase Conjecture (Cell-Probe), and making progress on either one of these conjectures as the Multiphase Program. The main goal for the rest of the section is the following: (i) group the previous results depending on which conjecture they make a progress on; then (ii) illustrate why the Multiphase Conjecture (Communication) is a **different ball game** compared to the Multiphase Conjecture (Cell-Probe).

[1] Originally, the conjecture was phrased in Merlin sending a message of length $o(m)$, then Alice and Bob proceeding to output $\mathsf{DISJ}(M_i, X)$. However, analogous implications hold from this slightly weaker conjecture

Previous Results on the Multiphase Program Despite its significance, the Multiphase Program had limited progress for the first 10 years after its inception namely [CEEP12, CGL15, BL15]. [CGL15, BL15] tackle the Multiphase Conjecture (Cell-Probe), giving a lower bound of $\max\{t_u, t_q\} \geq \Omega(\log n)$. But it should be noted that their bounds do not carry over to the Multiphase Conjecture (Communication) due to the limitation of the technique involved.

There are results on the Multiphase Conjecture (Communication) as well. [CEEP12] first tried to tackle the Multiphase Conjecture (Communication), but their argument only works for a very restricted model of communication. Only recently [KW20] developed a tool to tackle the Multiphase Conjecture (Communication) for two rounds of communication between Alice and Bob. This was extended to a more general class of functions by [DL20].

Separation between Communication and Cell-Probe An intuitive high-level explanation on why tackling the Multiphase Game is a harder task is the following. Recall that in the reduction due to [Pat10], Merlin assumes the role of updates, and Alice assumes the role of query algorithm. But in the Multiphase Game, (a) Merlin only needs to send cells written by the update algorithm, as Merlin knows both M and X. We do not charge for reading the cells during the update; (b) Alice is given the linear operator M. Thus, communication is needed for Alice to learn about X to compute $\langle M_i, X \rangle$. On the other hand, if we were to directly approach the Multiphase Problem and give a cell-probe lower bound, the counting arguments in [CGL15, BL15] crucially rely on the fact that (a) the update algorithm has no prior knowledge on M, thus must probe into pre-processed cells to learn M; (b) the query algorithm has no prior knowledge on both M and X. The query algorithm must learn about both M and X by probing into the pre-processed and updated cells.

An explicit example of the separation between the two models exists as well. There exists an explicit function that is easy in the communication model while hard in dynamic cell-probe model, separating the two models. Consider the setting where each M_i has a single non-zero entry. This is so-called indexing problem, as each M_i only asks which coordinate of X to output. [BL15] (Theorem 1) shows that if $t_u \leq o(m)$, $t_q \geq \Omega(n)$ for so-called non-adaptive query. This easily translates to $t_q \geq \Omega(\log n)$ for general query, by simulating all possible adaptive queries in a single non-adaptive query.

We will now show why we cannot expect to show such statement for the indexing problem in the communication model. First of all, Merlin's message is too long. We cannot show an analogous statement in the communication model, as $n \cdot t_u$, which would be the length of Merlin's message, is already larger than m. If Merlin is allowed to send a message of length m, Merlin can simply write the answers for all possible queries. Bob then only needs to a single bit to announce the answer. Notice that this stems from the difference listed as (a).

Even if we relax the condition on the update to say $n \cdot t_u \leq o(m)$, thereby Merlin sending a message of length $o(m)$, we still run into separation, as we cannot give $t_q \geq \Omega(\log n)$ for indexing problem. Such a result would contradict [Dru12] which combined with the argument from [Vio18] gives a communication protocol with $O(k^3)$ communication between Alice and Bob while Merlin only gives an advice of length $O(n/k)$ when $m = O(n)$.[2] This shows that even if we give (a) for free, (b) is still crucial, separating the two model.

Now that we have established the fact that the Communication model is provably stronger than its dynamic cell-probe analogue, we would like to discuss evidences on how hard can it be, which has direct connection to our problem. Ko and Weinstein [KW20] observed that the Multiphase Communication Game can simulate a static cell-probe data structure for some random linear operator M over the input X – the main problem of our interest.

[2]One could also obtain $O(k^2)$ communication with advice of length $O\left(\frac{n \log n}{k \log \log n}\right)$

Merlin sends all the contents of the pre-processing. Alice then simulates the query using the two-party communication between Bob. This is possible as Alice knows M, the linear operator in question. Therefore, resolving the Multiphase Conjecture (Communication) would imply Jukna-Schnitger Conjecture in its cell-probe analogue. In particular, the Multiphase Conjecture (Communication) implies a static cell-probe lower bound of the form

$$t \geq \left(\frac{|\mathcal{Q}|}{s}\right)^c \tag{2}$$

for some small constant $c > 0$ against arbitrary linear operators with $|\mathcal{Q}|$-many outputs. Note that this is a super-exponential improvement over (1). Therefore, Ko and Weinstein (see Section 5 of [KW20]) phrased the connection as a **major setback** for the full resolution of Multiphase Conjecture (Communication), rather than adding the stakes for the Multiphase Conjecture (Communication). We emphasize that no such connection arises for the Multiphase Conjecture for the Multiphase Conjecture (Cell-Probe).

1.1.2 Other Connections

Our work also connects to index coding with side information [BYBJK11], the complexity of error-correcting codes [GHK+13], and function inversion in cryptography [CGK19]. These diverse connections all stem from the same high-level question: can non-linear computation help compute linear functions?

1.2 Our Result

Our main contribution is a query time lower bound that breaks the logarithmic barrier for a random linear operator:

$$t \geq \Omega\left(\log \frac{|\mathcal{Q}|}{s}\right) \tag{3}$$

This provides the first formal evidence supporting the half-century-old intuition that non-linear pre-processing offers little advantage for most linear operators. For example, when space $s = |\mathcal{Q}|^{0.99}$ (storing nearly all answers), previous bounds gave only a trivial $t \geq \Omega(1)$. Our bound gives a much stronger $t \geq \Omega(\log |\mathcal{Q}|)$. This affirms the cell-probe version of Jukna-Schnitger's Conjecture when the number of queries is super-polynomial.

Formally, we prove the following statement:

Theorem 1.3 (Main)**.** *For every $t, s, m \geq \omega(1)$ and a collection $\mathcal{S} \subset \mathbb{F}_2^n$ such that $m \geq \omega(s \cdot 2^{5t})$, and $\log |\mathcal{S}| \geq 10^4 \cdot t \log s$, there must exist some m linear functions from $\mathcal{S}$ say $\mathcal{Q}$ that does not admit a data structure using s-space and t-probes per query.*

When applied to the set of all linear functions $\mathcal{S} = \mathbb{F}_2^n$, this theorem yields our main trade-off.

1.2.1 Implications

Our result has direct implications in three key areas:

Static Data Structures We establish the existence of "hard" linear operators, providing the first non-trivial lower bounds against arbitrary non-linear pre-processing in the high-space regime when s is $|\mathcal{Q}|^{0.99}$. This is a significant improvement over any existing arguments, which fail in this setting.

Circuit Complexity Our data structure bounds translate to new circuit lower bounds. Consider Valiant's depth-2 circuit with **arbitrary** gates [Val77] computing a linear operator $x \mapsto Mx$ with $M \in \mathbb{F}_2^{m \times n}$, $m^{0.99}$-gates in the middle layer, and at most τ-wires per output gate. Our result directly implies that if $t \leq o(\sqrt{n})$ and $m \geq 2^{O(t)}$, there must exist a linear operator M that requires $\tau \geq \Omega(t)$. This partially resolves an open problem of Jukna-Schnitger [JS11, Juk12] on whether arbitrary gates can help in computing an arbitrary linear operator for a super-polynomial number of outputs.

Furthermore, if we turn to a general depth d circuit with arbitrary gates (unbounded fan-in), we can also replace the inverse Ackermann-type lower bounds due to [GHK$^+$13] on linear operators with poly-logarithmic decay in depth d. Our bound also supersedes the best-known explicit bound, which are not linear operators, due to [RS03, Che08a, Che08b] when $d \geq 4$. See Section 4 for details.

Multiphase Communication Game A general way to phrase our result is a progress in the Multiphase Conjecture (Communication). While the notation of our proof is used specifically to handle the static data structure lower bound, a careful examination of our proof shows that our result indeed resolves the Multiphase Conjecture (Communication) [Pat10] when the number of possible queries m is super-polynomial.

Since any progress on the Multiphase Conjecture (Communication) implies a corollary for the Multiphase Conjecture (Cell-Probe), our lower bound then implies the following corollary for the Multiphase Problem.

Corollary 1.4. *For the Multiphase Problem, if the total update time $n \cdot t_u \leq O(m^{0.99})$ (i.e. slightly less than writing down answers to all possible queries) then the query time must be $t_q \geq \Omega(n^{1/2})$*

We suspect further reductions to various explicit dynamic data structure problems from Corollary 1.4. Note that a trivial data structure exists for the above explicit dynamic data structure instance, achieving $t_q = O(n)$ by simply reading M_i and X to compute $\langle M_i, X \rangle$. This is important because one could have otherwise devised a dynamic data structure problem which selects arbitrary functions $f_1, \ldots, f_m : \{0,1\}^n \to \{0,1\}$, and outputs $f_i(X)$ at the query phase.

Using Miltersen's counting argument [Mil93], one can readily demonstrate the existence of functions $f_1, \ldots, f_m$ such that, if the total update time $n \cdot t_u = O(m^{0.99})$, then $t_q \geq \Omega(n^{0.99})$. However, mainly due to requiring an arbitrary function f_i, which requires 2^n-bits to describe per query, this prohibits any trivial query algorithm of $t_q = O(n)$, as one needs to read about f_i, nor reduces to any interesting problem in dynamic data structure. [3]

1.2.2 Technical Contribution

The main technical ingredient of our work is the extension of the analysis of Multiphase Game [KW20, DL20] to exponentially small advantage regimes using average min-entropy introduced in [DORS08]. We suspect that our technique to be a critical component in fully resolving the Multiphase Program [Pat10, Tho13, Bra22]. Analyzing the small advantage regime is crucial for data structure lower bounds, as demonstrated in recent breakthroughs in dynamic data structure lower bounds [LWY20, LY25].

The main technical contribution in [KW20] shows that after Merlin's message, if Alice and then Bob each speak once and correctly output $\mathsf{DISJ}(M_i, X)$ with $s = o(m)$, then Alice and Bob must communicate $\Omega(\sqrt{n})$ bits. [DL20] extended the result to $\langle M_i, X \rangle$ with $\Omega(n)$ bits of communication.

[3] Another way to view this phenomenon is that our instance requires small linear bits (n-bits) per query, while Miltersen's counting argument requires exponential bits per query.

The key open problem posed in [KW20] is to extend the previous result to the setting where Alice, Bob then Alice speaks (i.e. one extra round compared to [KW20, DL20]). This has connections to the fundamental question of linear vs. non-linear circuits as remarked in [KW20]. We reduce the general setting with t-rounds of communication to a setting where Alice and then Bob speak with reduced correctness, via following observation.

Alice does the following: she picks a single random t-probe decision tree path. Then Alice sends the path to Bob. Bob checks if the path is consistent with U, Merlin's message. If yes, Bob outputs the same output as the end of the path. Otherwise Bob simply makes a random guess about the output. The main observation is that Bob would say yes with 2^{-t} probability. If Bob says Yes, then the protocol is correct. Otherwise, we get 0 advantage. Therefore, this would result in $\Omega(2^{-t})$ advantage over random guessing.

We then focus on the setting in which Alice and then Bob each speak once, with Bob guessing the value of $\langle M_i, X\rangle$ with a small advantage over a random guessing. That is, we consider **the small advantage regime**. Unfortunately, previous arguments [KW20, DL20] incurred a constant blowup in error, making them insufficient for any meaningful bound for small advantage regimes. In fact, as observed in two-party Disjointness [BM13], analyzing small advantage regime often requires different tools from that of small error regime. Our technical novelty here is to leverage the average min-entropy framework introduced in [DORS08] instead of standard information theoretic arguments using KL-divergence, and carefully choose random variables that fit the framework of [KW20]. Average min-entropy allows us to naturally bound the ℓ_2-norm of the probability distribution, which then can be then used to argue about the underlying discrepancy.

2 Preliminary

2.1 Information Theory

In this section, we provide the necessary background on information theory and information complexity that are used in this paper. For further reference, we refer the reader to [CT06].

Definition 2.1 (Entropy)**.** *The entropy of a random variable X is defined as*

$$H(X) := \sum_x \Pr[X = x] \log \frac{1}{\Pr[X = x]}.$$

Similarly, the conditional entropy is defined as

$$H(X|Y) := \mathbb{E}_Y \left[\sum_x \Pr[X = x|Y = y] \log \frac{1}{\Pr[X = x|Y = y]} \right].$$

Fact 2.2 (Conditioning Decreases Entropy)**.** *For any random variable X and Y*

$$H(X) \geq H(X|Y)$$

With entropy defined, we can also quantify the correlation between two random variables, or how much information one random variable conveys about the other.

Definition 2.3 (Mutual Information)**.** *Mutual information between X and Y (conditioned on Z) is defined as*

$$I(X;Y|Z) := H(X|Z) - H(X|YZ).$$

Similarly, we can also define how much one distribution conveys information about the other distribution.

Definition 2.4 (KL-Divergence)**.** *KL-Divergence between two distributions μ and ν is defined as*

$$\mathsf{D}_{KL}(\mu||\nu) := \sum_x \mu(x) \log \frac{\mu(x)}{\nu(x)}.$$

To bound mutual information, it suffices to bound KL-divergence, due to the following fact.

Fact 2.5 (KL-Divergence and Mutual Information)**.** *The following equality between mutual information and KL-Divergence holds*

$$I(A; B|C) = \mathbb{E}_{B,C}\left[\mathsf{D}_{KL}(A|_{B=b,C=c}||A|_{C=c})\right].$$

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

$$\|P - Q\|_{TV} = \frac{1}{2}\|P - Q\|_1 \leq \sqrt{\frac{1}{2\log e} D(P||Q)}$$

We also make use of the following facts on Mutual Information throughout the paper.

Fact 2.7 (Chain Rule)**.** *For any random variable A, B, C and D*

$$I(AD; B|C) = I(D; B|C) + I(A; B|CD).$$

Fact 2.8. *For any random variable A, B, C and D, if $I(B; D|C) = 0$*

$$I(A; B|C) \leq I(A; B|CD).$$

Proof. By the chain rule and non-negativity of mutual information,

$$I(A; B|C) \leq I(AD; B|C) = I(B; D|C) + I(A; B|CD) = I(A; B|CD).$$

□

Fact 2.9. *For any random variable A, B, C and D, if $I(B; D|AC) = 0$*

$$I(A; B|C) \geq I(A; B|CD).$$

Proof. By the chain rule and non-negativity of mutual information,

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

□

2.2 ℓ_2-norm of a distribution

Another good measure of the randomness of a distribution is its ℓ_2-norm. The more "spread out" a distribution is, the smaller its ℓ_2-norm. We introduce the following definitions to argue about the ℓ_2-norm of a distribution.

Definition 2.10. *We define the renyi entropy $H_2(A)$ and min-entropy $H_\infty(A)$ as*

$$H_2(A) := -\log\left(\sum_a \Pr[A=a]^2\right)$$
$$H_\infty(A) := -\log\left(\max_a \Pr[A=a]\right)$$

Fact 2.11 (Renyi Entropy)**.** *Let A be a random variable. Then*

$$H(A) \geq H_2(A) \geq H_\infty(A)$$

In particular, for any fixed b we have

$$H_2(A|B=b) \geq H_\infty(A|B=b)$$

We use the following simple claim to argue about ℓ_∞-norm of a distribution when its KL-divergence with the uniform distribution is small.

Claim 2.12. *Let $\mathcal{D}$ be a distribution and $\mathcal{U}$ a uniform distribution over some $\mathcal{S} \subset \{0,1\}^n$. Then if $D(\mathcal{D}||\mathcal{U}) < t$ with $t > 1$, then for every $\alpha > 2$ there exists an event E such that*

$$\mathcal{D}(E) \geq 1 - \frac{1}{\alpha}$$
$$H_\infty(\mathcal{D}|_E) \geq \log|\mathcal{S}| - 2\alpha t$$

Proof. We partition X depending on $\mathcal{D}(X)$. Let E denote the set of X such that

$$\log\frac{\mathcal{D}(X)}{\mathcal{U}(X)} < \alpha t$$

Since $D(\mathcal{D}||\mathcal{U}) < t$, by Markov's inequality $1 - \mathcal{D}(E) < 1/\alpha$. With $\alpha > 2$, for any X that is in the support of $\mathcal{D}|_E$, we have

$$\mathcal{D}|_E(X) \leq 2 \cdot \mathcal{D}(X) \leq 2^{\alpha t+1} \cdot \mathcal{U}(X) = 2^{-\log|\mathcal{S}|+\alpha t+1} \leq 2^{-\log|\mathcal{S}|+2\alpha t}$$

which then gives $H_\infty(\mathcal{D}|_E) \geq \log|\mathcal{S}| - 2\alpha t$. □

We prove the following lemma to bound the ℓ_2-norm of a distribution.

Lemma 2.13. *Let A, B be random variables where B has at most 2^λ possible values. Then*

$$\mathbb{E}_{b\sim B}\left[\sum_a \Pr[A=a|B=b]^2\right] \leq 2^{-H_\infty(A)+\lambda}$$

We use the following lemma on "average" min-entropy.

Definition 2.14 (Average Min-Entropy)**.**

$$\tilde{H}_\infty(A|B) = -\log\left(\mathbb{E}_{b\sim B}\left[\max_a \Pr[A=a|B=b]\right]\right) = -\log\left(\mathbb{E}_{b\sim B}\left[2^{-H_\infty(A|B=b)}\right]\right)$$

This leads to the following proposition

Proposition 2.15.

$$\mathbb{E}_{b\sim B}\left[\sum_a \Pr[A=a|B=b]^2\right] \leq 2^{-\tilde{H}_\infty(A|B)}$$

Proof.

$$\sum_a \Pr[A=a|B=b]^2 = 2^{-H_2(A|B=b)} \leq 2^{-H_\infty(A|B=b)}$$

Therefore,

$$\mathbb{E}_{b\sim B}\left[\sum_a \Pr[A=a|B=b]^2\right] \leq \mathbb{E}_{b\sim B}\left[2^{-H_\infty(A|B=b)}\right] = 2^{-\tilde{H}_\infty(A|B)}$$

where the last equality holds from the definition of average min-entropy. □

Lemma 2.16 (Lemma 2.2 of [DORS08])**.** *Let A, B be random variables. Then if B has at most 2^λ possible values, then*

$$\tilde{H}_\infty(A|B) \geq \tilde{H}_\infty(A,B) - \lambda \geq H_\infty(A) - \lambda.$$

We are now ready to prove Lemma 2.13

Proof of Lemma 2.13. First, observe that Lemma 2.16 implies

$$-\log\left(\mathbb{E}_{b\sim B}\left[2^{-H_\infty(A|B=b)}\right]\right) \geq H_\infty(A) - \lambda$$

which is equivalent to (by switching sides and taking as exponents)

$$2^{-H_\infty(A)+\lambda} \geq \mathbb{E}_{b\sim B}\left[2^{-H_\infty(A|B=b)}\right]$$

Now using Fact 2.11, you get

$$\mathbb{E}_{b\sim B}\left[2^{-H_\infty(A|B=b)}\right] \geq \mathbb{E}_{b\sim B}\left[2^{-H_2(A|B=b)}\right] = \mathbb{E}_{b\sim B}\left[\sum_a \Pr[A=a|B=b]^2\right].$$

which is the desired inequality. □

Using the ℓ_2 norm, the following claim (so-called Lindsey's Lemma) bounds the discrepancy of the Hadamard matrix under a product distribution.

Claim 2.17 (Lindsey's Lemma)**.** *Let H be a Hadamard matrix. Let P and Q be distributions. Then*

$$P^T H Q \leq \|P\|_2 \|Q\|_2 \cdot 2^{n/2}$$

Proof. Recall that the operator norm of the Hadamard matrix is exactly

$$\|H\|_2 = 2^{n/2}$$

Then by the definition of the operator norm,

$$\frac{P^T H Q}{\|P\|_2\|Q\|_2} \leq \|H\|_2 = 2^{n/2}$$

Rearranging the inequality, we obtain the desired claim. □

3 Main Proof

In this section, we prove the following main theorem.

Theorem 3.1 (Main)**.** *For every $t, s, m \geq \omega(1)$ and a collection $\mathcal{S} \subset \mathbb{F}_2^n$ such that $m \geq \omega(s \cdot 2^{5t})$, and $\log |\mathcal{S}| \geq 10^4 \cdot t \log s$, there must exist some m linear functions from $\mathcal{S}$ say $\mathcal{Q}$ that does not admit a data structure using s-space and t-probes per query.*

3.1 Compression and Random Process

We prove this by contradiction. We will start with the following assumption. Suppose for any $M_1, \ldots, M_m \in \mathcal{S} \subseteq \mathbb{F}_2^n$ (these are fixed and publicly known), there exists a data structure under the cell-probe model which pre-processes any given $X \in \mathbb{F}_2^n$ using s-space and answers $\chi_{M_i}(X) := (-1)^{\langle M_i, X \rangle}$ for any given $i \in [|\mathcal{Q}|] = [m]$ using t-probe. This implies the following procedure exists in the cell-probe model.

1. The querier is given $\vec{M} = M_1, \ldots, M_m$. Given $i \in [m]$, the querier makes t queries to U, using a decision tree $\mathcal{T}$ of depth t.
2. At the leaf node of the deicision tree $\mathcal{T}$, the querier outputs the final guess in $\{\pm 1\}$ which is guaranteed to be $\chi_{M_i}(X)$.

Protocol 1: Static Data structure with adaptive probes

We now argue that an effective static data structure implies an too-good-to-be true communication process in the so-called Multiphase Communication Game. While Pătraşcu [Pat10] considered computing the Disjointness of M_i and X, here we consider computing the inner product over mod 2 between M_i and X.

The above data structure implies the existence of the following $(t+1)$-round communication process (between 3 players) under the independent uniform distribution for all $M_1, \ldots, M_m \in \mathcal{S} \subseteq \mathbb{F}_2^n$ and uniform distribution for $X \in \mathbb{F}_2^n$ where Merlin sends a single-shot message of length s, $U(M_1, \ldots, M_m, X)$ to Bob. Then Alice and Bob proceed in t-round communication to compute $\chi_{M_i}(X)$. By the correctness of the underlying static data structure, this process correctly outputs the inner product of M_i and X for any values of $i \in [m]$.

1. Alice holds $M_1, \ldots, M_m$ and $i \in [m]$. Bob holds X and $i \in [m]$. Merlin holds $M_1, \ldots, M_m, X$.
2. Merlin sends $U(M_1, \ldots, M_m, X)$ to Bob using s bits.
3. Alice sends a location $q \in [s]$ (given in her decision tree $\mathcal{T}$) using $\log s$-bits to Bob. Bob replies with the queried bit U_q.
4. Repeat the Step 3 for t rounds.
5. Alice announces $\chi_{M_i}(X)$.

Protocol 2: $t+1$-round Communication Process

We denote this as a process, rather than a protocol, to contrast with the usual two-party communication protocol. Due to the side information U that Bob has on Alice's input, the usual

techniques for two-party communication complexity (namely cut-and-paste or rectangular property) do not hold. Therefore, we call this a communication process instead of a communication protocol.

Compression Overview The hope is to show that if s is sufficiently small then t must be large for such a $(t+1)$-round communication process. Unfortunately, current techniques are insufficient to provide a complete lower bound for such a general process. Instead, we would like to "compress" the communication between Alice and Bob with a loss in parameters, resulting in a 3-round communication process between Merlin, Alice, and Bob. We can manage this setting from the technique developed in [KW20].

The high-level intuition is for Alice to randomly sample a path in her decision tree, along with the final answer. Bob simply checks whether or not queried bits indeed matches with U. If it matches, Bob announces the answer as given in Alice's message. Otherwise, Bob simply makes an independent random guess. We formally state the compression as following.

Formal Compression Here is the formal description of our compressed 3-round communication process.

1. Merlin sends $U(M_1, \ldots, M_m, X)$ to Bob using s bits.
2. Let $P_i \in ([s], \{0,1\})^t \times \{\pm 1\}$, namely query i's (i) the sequence of the addresses probed by the probing algorithm; (ii) the probed result and; (iii) the final answer (i.e. the contents of a path in the decision tree used by Alice) Note that there are at most 2^t many possible P_i. Alice picks a $P_i = p_i$ such that
$$p_i := \arg\max \Pr_X[P_i = p_i | \vec{M} = \vec{m}]$$
then sends to Bob using $t \log s + t + 1$ bits.
3. Bob checks if P_i agrees with U. If yes, set B_i as 1. Output Z_i, which is the notation for Bob's guess of $\chi_{M_i}(X)$, as the final output value in P_i. Otherwise, set B_i as 0. Output Z_i as ± 1 each with probability $1/2$ independently at random.

Protocol 3: Communication Process between Alice and Bob for Adaptive Probe

Remark 3.2. *Observe that our argument works against even a general three party communication, where after Merlin's message, Alice and Bob proceed to communicate t bits using t-rounds of communication (i.e. a single bit per round). As long as Alice's bits are a function of previous transcript, and her input $(M_1, \ldots, M_m, i)$; and Bob's bits are a function of previous transcript, his input (X, i) and Merlin's message, the above compression scheme works. Bob simply needs to check if the response chosen by the path is consistent with his actual response.*

Due to Remark 3.2, the rest of proof also works against arbitrary Number-On-Forehead three party communication, thereby attacking the Multiphase Conjecture (Communication). However, for conciseness, we stick to above notations against static data structure lower bound.

Random Process from Protocol 3 Here is the plan for the remainder of the proof. Given a too-good-to-be true compressed communication process (Protocol 3), we would like to "extract" a random process Z from our compressed communication process which would violate some combinatorial property. We will show that some choice of Z (after some conditioning E which is introduced

due to technical reasons) satisfies three properties simultaneously: conditioned on Z, large average Min-Entropy on M_i and X, low correlation between M_i and X, while achieving a good advantage over randomly guessing $\chi_{M_i}(X)$. Then in Section 3.2, we will then show that such a choice of Z cannot exist, resulting in a contradiction.

Given the formal description of our compression, we formally state our random process Z. Pick a sequence of random distinct numbers $\mathcal{P}$ of length $\ell \leq m/100$ in $[m]$, an instance of which we will denote as $(\rho_1, \ldots, \rho_\ell)$, and random $J \in [\ell]$, an instance of which we will denote as j. Then we write

$$Z := \mathcal{P}, J, P_{\mathcal{P}_{<J}}, M_{\mathcal{P}_{<J}}, P_{\mathcal{P}_J}, B_{\mathcal{P}_J}, Z_{\mathcal{P}_J}$$

which is the set of random "path" of coordinates along with Alice's compressed messages and S_i's along the path. We also include Bob's output (whether or not $P_{\mathcal{P}_J}$ agrees with U) for the particular J of our choosing. For brevity, we write Z_{Alice} for

$$Z_{Alice} := \mathcal{P}, J, P_{\mathcal{P}_{<J}}, M_{\mathcal{P}_{<J}}, P_{\mathcal{P}_J}$$

that is Z without $B_{\mathcal{P}_J}$ (i.e. Bob's message) and the final guess $Z_{\mathcal{P}_J}$ due to Bob's message. We emphasize here that Z_{Alice} is completely determined by $M_1, \ldots, M_m = \vec{M}$ and independent randomness $\mathcal{P}, J$, and has no dependence on X (Bob's input) which is a crucial property of Z_{Alice} and why the variable is named so.

3.1.1 Large Average Min-Entropy

In this section, using the communication process Z, first, we argue on large average Min-Entropy for $M_{\mathcal{P}_J}$ and X conditioned on Z. As a comparison, we remark that [KW20, DL20] extract $Z = z$ with a small KL-divergence which is directly implied by small mutual information between Z and respective $M_{\mathcal{P}_J}$ and X.

We start with the following lemma which directly follows from the technique used in [KW20].

Lemma 3.3. *If $|\mathcal{P}| = \ell = |\mathcal{Q}|/100$ then*

$$I(M_{\mathcal{P}_J}; Z_{Alice}) = \mathbb{E}_{z_{Alice}} \left[D(M_{\mathcal{P}_J}|_{Z_{Alice}=z_{Alice}} || M_{\mathcal{P}_J}) \right] \leq 3t \log s$$

We attach the proof of Lemma 3.3 in the appendix for completeness, as it is exactly the same as in [KW20]. What Lemma 3.3 guarantees is Z_{Alice} such that conditioning on which has low divergence with original uniform distribution on $M_{\mathcal{P}_J}$.

The following is the main lemma of this section, which combined with Lemma 3.3, we can extract large average Min-Entropy part for $M_{\mathcal{P}_J}$ and X.

Lemma 3.4 (Large Average Min-Entropy)**.** *For any setting of fixed $Z_{Alice} = z_{Alice}$ such that*

$$D(M_{\mathcal{P}_J}|_{Z_{Alice}=z_{Alice}} || M_{\mathcal{P}_J}) \leq \beta$$

with $\beta > 2$, there exists an associated event E with

$$\begin{aligned}
&\Pr\left[E | Z_{Alice} = z_{Alice}\right] \geq 3/4 \\
&H(E | Z_{Alice}, M_{\mathcal{P}_J}) = 0 \\
&\tilde{H}_\infty(M_{\rho_j} | Z_{\rho_j} B_{\rho_j}, E = 1, Z_{Alice} = z_{Alice}) + \tilde{H}_\infty(X | Z_{\rho_j} B_{\rho_j}, E = 1, Z_{Alice} = z_{Alice}) \\
&\geq n + \log|\mathcal{S}| - 10\beta
\end{aligned}$$

Towards proving the lemma, we first prove the following two propositions.

Proposition 3.5. *Suppose*

$$D(M_{\mathcal{P}_J}|_{Z_{Alice}=z_{Alice}}||M_{\mathcal{P}_J}) \leq \beta.$$

Then there exists an event E with $\Pr[E|Z_{Alice} = z_{Alice}] \geq 3/4$ such that

$$H(E|Z_{Alice}, M_{\mathcal{P}_J}) = 0$$
$$H_\infty(M_{\rho_j}|Z_{Alice} = z_{Alice}, E = 1) \geq \log|\mathcal{S}| - 8\beta$$

Proof. Recall that Claim 2.12 gives an event E where if the underlying distribution has low KL-divergence with the uniform distribution over $\mathcal{S}$, conditioned on E, the distribution has large H_∞. As

$$D(M_{\rho_j}|_{Z_{Alice}=z_{Alice}}||M_{\rho_j}) \leq \beta.$$

with $M_{\mathcal{P}_J}$ being the uniform distribution over $\mathcal{S}$, we can apply Claim 2.12 to give an event E which is completely determined by Z_{Alice} and $M_{\mathcal{P}_J}$ (as E is determined by $M_{\mathcal{P}_J}$ under fixed Z_{Alice}) such that

$$H_\infty(M_{\rho_j}|Z_{Alice} = z_{Alice}, E = 1) \geq \log|\mathcal{S}| - 8\beta$$

with

$$\Pr[E|Z_{Alice} = z_{Alice}] \geq 3/4$$

by setting the appropriate parameter ($\alpha = 4$ in Claim 2.12). □

The next proposition is H_∞ bound for X.

Proposition 3.6. *For any setting of $Z_{Alice} = z_{Alice}$, and $E = 1$, we have*

$$H_\infty(X|Z_{Alice} = z_{Alice}, E = 1) = n.$$

Proof. First, by our setting of Z_{Alice} and E, which depends only on $\vec{M} = M_1, \ldots, M_m$ (Alice's input),

$$I(Z_{Alice}, E; X) \leq I(\vec{M}; X) = 0.$$

Therefore, for any setting of Z_{Alice} and $E = 1$, we get

$$H_\infty(X|Z_{Alice} = z_{Alice}, E = 1) = n.$$

□

We are now ready for the proof of Lemma 3.4

Proof of Lemma 3.4. Due to Proposition 3.5, there exists an event E, that is completely determined by corresponding S_{ρ_j} such that

$$H(E|Z_{Alice}, M_{\mathcal{P}_J}) = 0$$
$$\Pr[E|Z_{Alice} = z_{Alice}] \geq 3/4$$
$$H_\infty(M_{\rho_j}|Z_{Alice} = z_{Alice}, E = 1) \geq \log|\mathcal{S}| - 8\beta$$

which further implies that, due to Lemma 2.16, as $Z_{\rho_j}B_{\rho_j}$ is 2-bit,

$$\tilde{H}_\infty(M_{\rho_j}|Z_{\rho_j}B_{\rho_j}, Z_{Alice} = z_{Alice}, E = 1) \geq \log|\mathcal{S}| - 8\beta - 2. \quad (4)$$

Due to Proposition 3.6, and Lemma 2.16, for any setting of Z_{Alice} we have

$$\tilde{H}_\infty\left(X|Z_{\rho_j}B_{\rho_j}, Z_{Alice} = z_{Alice}, E = 1\right) \geq n - 2 \quad (5)$$

Then adding (4) and (5), we obtain our desired lemma as

$$\begin{aligned}&\tilde{H}_\infty\left(X|B_{\rho_j}, Z_{Alice}=z_{Alice}, E=1\right)+\tilde{H}_\infty(M_{\rho_j}|B_{\rho_j}, Z_{Alice}=z_{Alice}, E=1)\\&\geq n+\log|\mathcal{S}|-8\beta-4\geq n+\log|\mathcal{S}|-10\beta.\end{aligned}$$

where the last inequality holds from our assumption $\beta > 2$. □

3.1.2 Low Correlation

Finally, we show that the correlation between $M_{\mathcal{P}_J}$ and X is small in expectation over random Z and conditioned on E, which essentially follows the analogous argument in [KW20] using the Chain Rule in Mutual Information (Fact 2.9).

Lemma 3.7 (Low Correlation)**.**

$$I(M_{\mathcal{P}_J}; X|Z, E=1)\leq\frac{2|U|}{|\mathcal{P}|}=\frac{2s}{\ell}$$

Proof. We use an analogous technique from [KW20]. First, note that

$$I(M_{\mathcal{P}_J}; X|Z, E=1)\leq I(M_{\mathcal{P}_J}; UX|Z, E=1)$$

Now we plug in the definition of Z as the concatenation of $Z_{Alice}, Z_{\mathcal{P}_j}B_{\mathcal{P}_j}$ along with $E=1$. Then, for any fixed $Z_{Alice}=z_{Alice}$, we get

$$I(M_{\mathcal{P}_J}; UX|Z_{Alice}=z_{Alice}, E=1, Z_{\rho_j}B_{\rho_j})\leq I(M_{\mathcal{P}_J}; UX|Z_{Alice}=z_{Alice}, E=1, B_{\rho_j})$$

As $I(Z_{\rho_j}; UX|B_{\rho_j}, Z_{Alice}=z_{Alice}, E=1, M_{\mathcal{P}_J})=0$. If $B_{\rho_j}=1$, Z_{ρ_j} is completely determined from z_{Alice}. Otherwise it is an independent random variable. Then

$$\begin{aligned}&I(M_{\mathcal{P}_J}; UX|Z_{Alice}=z_{Alice}, E=1, B_{\rho_j})\leq I(M_{\mathcal{P}_J}; UX|Z_{Alice}=z_{Alice}, E=1)\\&\leq\frac{1}{\Pr[E=1|Z_{Alice}=z_{Alice}]}I(M_{\mathcal{P}_J}; UX|Z_{Alice}=z_{Alice}, E)\leq 2\cdot I(M_{\mathcal{P}_J}; UX|Z_{Alice}=z_{Alice}, E)\\&\leq 2\cdot I(M_{\mathcal{P}_J}; UX|Z_{Alice}=z_{Alice})\end{aligned}$$

where the inequalities hold by Fact 2.9 along with

$$I(M_{\mathcal{P}_J}; B_{\rho_j}|Z_{Alice}=z_{Alice}, E=1, UX)\leq H(B_{\rho_j}|Z_{Alice}=z_{Alice}, E=1, UX)=0$$

and the properties of E guaranteed by Lemma 3.4, namely $\Pr[E=1|Z_{Alice}=z_{Alice}]\geq 3/4$ and

$$I(E; UX|Z_{Alice}=z_{Alice}, M_{\mathcal{P}_J})\leq H(E|Z_{Alice}=z_{Alice}, M_{\mathcal{P}_J})=0.$$

Next, we bound $I(M_{\mathcal{P}_J}; UT|Z_{Alice})$ term. Note that $I(M_{\mathcal{P}_J}; X|Z_{Alice})=0$ as

$$I(M_{\mathcal{P}_J}; X|Z_{Alice})\leq I(M_{\mathcal{P}_J}Z_{Alice}; X)\leq I(\vec{M}; X)=0$$

Next, we consider $I(M_{\mathcal{P}_J}; U|Z_{Alice}, X)$ then plugging in the definition of Z_{Alice},

$$\begin{aligned}&I(M_{\mathcal{P}_J}; U|Z_{Alice}, X)=I(M_{\mathcal{P}_J}; U|\mathcal{P}, J, P_{\mathcal{P}_J}, P_{\mathcal{P}_{<J}}, M_{\mathcal{P}_{<J}}, X)\\&\leq I(P_{\mathcal{P}_J}, M_{\mathcal{P}_J}; U|\mathcal{P}, J, P_{\mathcal{P}_{<J}}, M_{\mathcal{P}_{<J}}, X).\end{aligned}$$

Then taking the expectation over the random coordinate J (by standard direct-sum technique), as all other variables are chosen independently of J, we get

$$\begin{aligned}
&I(P_{\mathcal{P}_J}, M_{\mathcal{P}_J}; U|\mathcal{P}, J, P_{\mathcal{P}_{<J}}, M_{\mathcal{P}_{<J}}, X) \\
&= \frac{1}{\ell}\sum_{j=1}^{\ell} I(P_{\mathcal{P}_j}, M_{\mathcal{P}_J}; U|\mathcal{P}, J=j, P_{\mathcal{P}_{<j}}, M_{\mathcal{P}_{<J}}, T) \\
&= \frac{1}{\ell}\sum_{j=1}^{\ell} I(P_{\mathcal{P}_j}, M_{\mathcal{P}_J}; U|\mathcal{P}, P_{\mathcal{P}_{<j}}, M_{\mathcal{P}_{<J}}, T) \\
&= \frac{I(P_{\mathcal{P}}, M_{\mathcal{P}}; U|\mathcal{P}, T)}{\ell} \leq \frac{H(U)}{\ell} \leq \frac{s}{\ell}.
\end{aligned}$$

This completes the proof of the lemma. □

3.1.3 Good Advantage

We also have the following simple observation on the advantage of Protocol 3 over a random guessing.

Lemma 3.8 (Advantage)**.** *Fix any $\vec{M}$. For any $i \in [m]$, the probability of Protocol 3 outputting $\chi_{M_i}(X)$ correctly over X is at least $\frac{1+2^{-t}}{2}$*

Proof. If we fix $\vec{M}$, first observe that X is independently at random. As $\vec{M}$ completely determines Alice's message, P_i, the only remaining part of the protocol is B_i and Z_i. There are two possible scenarios for B_i. If U agrees with P_i, or not, that is if $B_i = 1$ or not. If $B_i = 1$, this implies that the last bit in P_i is indeed $\chi_{M_i}(X)$ due to the correctness of the original $t+1$-round Communication Process Protocol 2. If $B_i = 0$, the process takes a random guess. Therefore, the probability of being correct can be written exactly as

$$\Pr_X[B_i = 1] + \Pr_X[B_i = 0] \cdot \frac{1}{2} = \Pr_X[B_i = 1] + (1 - \Pr_X[B_i = 1]) \cdot \frac{1}{2} = \frac{1 + \Pr_X[B_i = 1]}{2}$$

Therefore it suffices to bound $\Pr_X[B_i = 1]$. Observe that we choose P_i as the maximum likelihood path. Therefore $\Pr_X[P_i] \geq \frac{1}{2^t}$, as there are at most 2^t many possible P_i. Denote P_i' as the true decision tree path given by fixed $\vec{M}$ and T. Then $\Pr_X[B_i = 1] = \Pr_X[P_i = P_i'] = \Pr_X[P_i]$, that is the probability of path P_i over possible T's. Then, we get

$$\Pr_X[B_i = 1] \geq \frac{1}{2^t}$$

Therefore, the probability of outputting $\chi_{M_i}(X)$ correctly is at least

$$\frac{1 + \Pr_X[B_i = 1]}{2} \geq \frac{1 + 2^{-t}}{2}$$

□

This immediately implies the following corollary, which we will use towards our final contradiction.

Corollary 3.9. *For any setting of $Z_{Alice} = z_{Alice}, M_{\rho_j}$, and $E = 1$ (which are all completely determined by $\vec{M}, \mathcal{P}, J$)*

$$\Pr[Z_{\rho_j} = \chi_{M_{\rho_j}}(X) | M_{\rho_j}, Z_{Alice} = z_{Alice}, E = 1] \geq \frac{1 + 2^{-t}}{2}$$

3.2 Combinatorial Lemma

The main corresponding combinatorial lemma to the [KW20, DL20] is the following. This shows that no random process that achieves even the slightest advantage over random guessing with low correlation and large average min-entropy can exist.

Lemma 3.10. *No setting of a random process $Z = z$ and C, which contains z_{out} can simultaneously satisfy all three of the following inequalities for any $\gamma \geq 3$*

$$\mathbb{E}_{C|Z=z}\left[\left|\mathbb{E}_{M_i,X|_{C=c,Z=z}}\left[\chi_{M_i}(X)\cdot z_{out}|C=c,Z=z\right]\right|\right] \geq 2^{-\gamma+2} \tag{6}$$

$$\tilde{H}_\infty(M_i|C,Z=z)+\tilde{H}_\infty(X|C,Z=z)\geq n+2\gamma \tag{7}$$

$$I(M_i;X|C,Z=z)\leq 2^{-2\gamma} \tag{8}$$

Proof. For the sake of contradiction, suppose such z and C exists. First as z_{out} is ± 1, we can write

$$\left|\mathbb{E}_{M_i,X|_{C=c,Z=z}}\left[\chi_{M_i}(X)\cdot z_{out}|C=c,Z=z\right]\right| = \left|\mathbb{E}_{M_i,X|_{C=c,Z=z}}\left[\chi_{M_i}(X)|C=c,Z=z\right]\right| \tag{9}$$

Then we can use the ℓ_1 bound to have

$$\left|\mathbb{E}_{M_i,X|_{C=c,Z=z}}\left[\chi_{M_i}(X)|C=c,Z=z\right]\right| \leq |M_i|_{C=c,Z=z}\cdot H\cdot X|_{C=c,Z=z}| \tag{10}$$

$$+\|M_i|_{C=c,Z=z}\times X|_{C=c,Z=z}-(M_i,X)|_{C=c,Z=z}\|_1 \tag{11}$$

We bound the expectation of (11). Our KL-divergence term is then equal to the mutual information between M_i and X conditioned on $Z=z$. Namely, using the chain rule for the KL divergence,

$$\begin{aligned} D(M_i,X|_{Z=z}||M_i|_{Z=z}\times X|_{Z=z}) &= \underbrace{D(X|_{Z=z}||X|_{Z=z})}_{=0}+\mathbb{E}_{x\sim X|_{Z=z}}\left[D(M_i|_{X=x,Z=z}||M_i|_{Z=z})\right] \\ &= I(M_i;X|Z=z) \end{aligned}$$

Then, due to Pinsker's inequality (Fact 2.6), we have

$$\|M_i|_{C=c,Z=z}\times X|_{C=c,Z=z}-(M_i,X)|_{C=c,Z=z}\|_1 \leq 2\sqrt{I(M_i;X|C=c,Z=z)}$$

Then taking expectation over C and applying Jensen's inequality,

$$\mathbb{E}_{C|_{Z=z}}\left[\|M_i|_{C=c,Z=z}\times X|_{C=c,Z=z}-(M_i,X)|_{C=c,Z=z}\|_1\right] \leq 2\sqrt{I(M_i;X|C,Z=z)} \leq 2^{-\gamma+1} \tag{12}$$

where the last bound holds from (8).

Next, we bound (10). Due to Claim 2.17 and Cauchy-Schwarz Inequality,

$$\begin{aligned} &\mathbb{E}_{C|_{Z=z}}\left[|M_i|_{C=c,Z=z}\cdot H\cdot X|_{C=c,Z=z}|\right] \leq \mathbb{E}_{C|_{Z=z}}\left[2^{n/2}\cdot\|M_i|_{C=c,Z=z}\|_2\cdot\|X|_{C=c,Z=z}\|_2\right] \\ &\leq 2^{n/2}\cdot\sqrt{\mathbb{E}_{C|_{Z=z}}\left[\|M_i|_{C=c,Z=z}\|_2^2\right]\cdot\mathbb{E}_{C|_{Z=z}}\left[\|X|_{C=c,Z=z}\|_2^2\right]} \end{aligned}$$

Proposition 2.15 implies

$$\begin{aligned} \mathbb{E}_{C|_{Z=z}}\left[\|M_i|_{C=c,Z=z}\|_2^2\right] &\leq 2^{-\tilde{H}_\infty(M_i|C,Z=z)} \\ \mathbb{E}_{C|_{Z=z}}\left[\|X|_{C=c,Z=z}\|_2^2\right] &\leq 2^{-\tilde{H}_\infty(X|C,Z=z)} \end{aligned}$$

which would in turn imply

$$\sqrt{\mathbb{E}_{C|_{Z=z}}\left[\|M_i|_{C=c,Z=z}\|_2^2\right]\cdot\mathbb{E}_{C|_{Z=z}}\left[\|X|_{C=c,Z=z}\|_2^2\right]} \leq 2^{-(\tilde{H}_\infty(M_i|C,Z=z)+\tilde{H}_\infty(X|C,Z=z))/2}$$
$$\leq 2^{-\frac{n+2\gamma}{2}} = 2^{-n/2}\cdot 2^{-\gamma}$$

which then implies (10) is upper bounded by

$$(10) \leq 2^{n/2}\cdot 2^{-n/2}\cdot 2^{-\gamma} = 2^{-\gamma} \tag{13}$$

Therefore, we get

$$\mathbb{E}_{C|Z=z}\left[\left|\mathbb{E}_{M_i,X|_{C=c,Z=z}}\left[\chi_{M_i}(X)\cdot z_{out}|C=c,Z=z\right]\right|\right] \leq \frac{3}{2^\gamma} \tag{14}$$

which contradicts (6). □

3.3 Combining the lemmas

Finally, we combine the lemmas from the previous two sections to prove the main theorem via contradiction. Recall the statement of our main theorem.

Theorem 3.1 (Main)**.** *Let $t, s, m \geq 100$ be parameters such that $m = |\mathcal{Q}| \geq \omega(s\cdot 2^{3t}), \log|\mathcal{S}| \geq 40\cdot t\log s$. Consider any data structure which answers m many linear functions from a collection $\mathcal{S}\subseteq\mathbb{F}_2^n$ say $\mathcal{Q}$. There must exist some $\mathcal{Q}$ such that there is no data structure for $\mathcal{Q}$ using s-space, t probes per query.*

Proof. We will prove via contradiction. Suppose otherwise. Then we know that Protocol 3 must exist as well with the provided definitions of Z_{Alice}, E, and Z. Our goal is to show that there exists a setting of $Z_{Alice} = z_{Alice}, E = 1$ and $Z_{\rho_j}B_{\rho_j}$ that would violate Lemma 3.10, thus leading to a contradiction.

First, we will fix $Z_{Alice} = z_{Alice}, E = 1$ which satisfies large average Min-Entropy and low correlation simultaneously. That is, we will fix $Z_{Alice} = z_{Alice}, E = 1$ such that

$$I(M_{\mathcal{P}_J}; X|Z_{Alice} = z_{Alice}, E = 1, Z_{\rho_j}B_{\rho_j}) \leq \frac{6s}{\ell} \tag{15}$$
$$\tilde{H}_\infty(M_{\rho_j}|Z_{\rho_j}B_{\rho_j}, E = 1, Z_{Alice} = z_{Alice}) + \tilde{H}_\infty(X|Z_{\rho_j}B_{\rho_j}, E = 1, Z_{Alice} = z_{Alice})$$
$$\geq n + \log|\mathcal{S}| - 300t\log s \tag{16}$$

Over the random choice of $Z_{Alice} = z_{Alice}$, Lemma 3.3 and Lemma 3.4 implies that

$$\tilde{H}_\infty(M_{\rho_j}|Z_{\rho_j}B_{\rho_j}, E = 1, Z_{Alice} = z_{Alice}) + \tilde{H}_\infty(X|Z_{\rho_j}B_{\rho_j}, E = 1, Z_{Alice} = z_{Alice}) \geq n + \log|\mathcal{S}| - 30t\log s$$

Lemma 3.7 implies that over the random choice of $Z_{Alice} = z_{Alice}$,

$$I(M_{\mathcal{P}_J}; X|Z_{Alice} = z_{Alice}, E = 1, Z_{\rho_j}B_{\rho_j}) \leq \frac{2s}{\ell}$$

Due to Markov argument, there exists $Z_{Alice} = z_{Alice}$ which simultaneously satisfy both (15) and (16).

On the other hand, Corollary 3.9 implies that for any setting of $Z_{Alice} = z_{Alice}, E = 1$ and S_{ρ_j}

$$\Pr[Z_{\rho_j} = \chi_{M_{\rho_j}}(X)|M_{\rho_j}, Z_{Alice} = z_{Alice}, E = 1] \geq \frac{1+2^{-t}}{2}$$

or equivalently

$$\mathbb{E}_{M_{\rho_j},X|_{Z_{\rho_j}B_{\rho_j},Z_{Alice}=z_{Alice},E=1}}\left[\left|\chi_{M_{\rho_j}}(X)\cdot Z_{\rho_j}\right|\right]\geq 2^{-t} \tag{17}$$

Then consider (15), (16) and (17). To obtain a desired contradiction with Lemma 3.10, setting $\gamma = t+2$ in the lemma, it suffices to have

$$\begin{aligned} &n+\log|\mathcal{S}|-30t\log s\geq n+2(t+2)\\ &\frac{6s}{\ell}\leq 2^{-2(t+2)} \end{aligned}$$

which is implied by

$$\begin{aligned} &40t\log s\leq \log|\mathcal{S}|\\ &6s2^{2(t+2)}\leq s\cdot 2^{3t}\leq \ell = |\mathcal{Q}|/100 \end{aligned}$$

These conditions are implied by the choice of parameters given in the statement of the theorem. This would contradict Lemma 3.10, completing the proof of the theorem. □

4 Wire Lower Bound for Circuits with Arbitrary Gates

In this section, we show that a random linear operator satisfies one of the conditions laid out by [Vio18] to obtain a breakthrough for lower bounds in circuits with arbitrary gates. This shows that a random linear operator does beat the state-of-the-art lower bounds given in [Che08a, Che08b, GHK+13].

4.1 Circuit with Arbitrary Gates

In this section, we formally define a circuit with arbitrary gates. (See Chapter 13 of [Juk12] and [Dru12] for further references) We would like to compute a Boolean operator $f:\{0,1\}^n\to\{0,1\}^m$ using a circuit where a gate can compute any function with unbounded fan-in. Note that this is the strongest possible model of circuit as computations are given for free. Therefore, its lower bound should apply to all possible models of circuit. Furthermore, it is meaningless to count the number of gates, as any f can be computed with m gates. What we measure instead is information transfer, quantified by the number of wires. Then a trivial upper-bound is $n\cdot m$, while a trivial lower bound is $\max\{n,m\}$. Assuming $m>n$, a non-trivial lower bound on the number of wires would be of the form $\omega(m)$.

In this unrealistically powerful model, we want to study the trade-off between the number of wires required for f as a function of the depth of the circuit d, output size m and input size n. Here we consider the setting where $m=O(n)$.

4.1.1 Previous Results

In order to describe the previous results, we need the following definition.

Definition 4.1. $\lambda_1(n)=\lfloor\sqrt{n}\rfloor$, $\lambda_2(n)=\lceil\log n\rceil$. *For* $d\geq 3$, $\lambda_d(n):=\lambda^*_{d-2}(n)$ *where* * *denotes the number of times the function must be applied to* n *to reach a value* ≤ 1.

Due to a simple counting argument, most arbitrary operators require $\tilde{\Omega}(n^2)$-wires [JS10]. However, if we turn to finding an explicit (or even semi-explicit) operator with $\omega(n)$-wires, the best explicit bound known (an improvement over [RS03]) is due to Cherukhin's Bound $\Omega_d(n\cdot\lambda_{d-1}(n))$

[Che08b, Che08a, Dru12]. However, the operator in consideration is not (and cannot be) a linear operator, though explicit. On the other hand, if we turn to finding some hard linear operator, [GHK+13] shows that computing any "good" linear error correcting code[4] requires $\Omega(n \cdot \lambda_{2\cdot\lfloor d/2 \rfloor}(n))$-many wires.

Depth	[Che08b, Che08a]	[GHK+13]	Our Result
$d = 2$	$\Omega(n \cdot \sqrt{n})$	$\Omega(n \cdot (\log(n)/\log\log(n))^2)$	$\Omega(n \cdot \log^{1/2}(n))$
$d = 3$	$\Omega(n \cdot \log(n))$	$\Omega(n \cdot \log\log(n))$	$\Omega(n \cdot \log^{1/3}(n))$
$d = 4$	$\Omega(n \cdot \log\log(n))$	$\Omega(n \cdot \log^*(n))$	$\Omega(n \cdot \log^{1/4}(n))$
d	$\Omega_d(n \cdot \lambda_{d-1}(n))$	$\Omega(n \cdot \lambda_{2\cdot\lfloor d/2 \rfloor}(n))$	$\Omega(n \cdot \log^{1/d}(n))$

Table 1: Wire lower bounds for Circuits with Arbitrary Gates

The primary technical reason for the rapid decay of wire lower bounds with increasing depth is the reliance on the so-called *superconcentrator* technique [Val75, Pip77, DDPW83, Pud94, PR94, RTS00, GHK+13] or more refined *Strong Multiscale Entropy* (SME) approach [RS03, Che08b, Che08a, Juk12]. These are the only previously known methods–aside from naive counting–for establishing lower bounds on circuits with arbitrary gates. Moreover, the limitations of these techniques are well-documented. For instance, [DDPW83, Pud94, RTS00, GHK+13] exhibited superconcentrators with small circuits. [Dru12] demonstrated that neither the SME approach nor a generalization of [GHK+13] can yield further improvements.

4.1.2 Our Result

We show a wire lower bound which only suffers a polynomial decay per depth d compared to the inverse Ackermann function type bound in previous works at the cost of using a random linear operator. In particular, we show the following theorem.

Theorem 4.2. *For most linear operators $M : \{0,1\}^n \to \{0,1\}^{O(n)}$, when computed by a circuit of depth d requires*

$$w \geq \Omega(n \cdot \log^{1/d}(n))$$

For depth-2 circuit, our bound is weaker than that of [GHK+13]. But we get a super-exponential improvement for $d \geq 3$. We also beat Cherukhin's Bound [Che08b, Che08a] for $d \geq 4$. This also implies a superlinear wire lower bound as long as $d = o(\log\log n)$. The result is summarized in Table 1.

We remark that such lower bound is only possible with an approach that is drastically different from superconcentrator or SME due to known limitations of these techniques [DDPW83, Pud94, RTS00, Dru12].

Separation between Representing a Linear Operator This also gives an exponentially stronger separation between "representing" a linear operator [Juk10, Dru12, Juk12] and "computing" a linear operator under circuits with arbitrary gates. A circuit C *represents* a linear operator f if there exists some basis $B \subset \mathbb{F}_2^n$ such that for every $b \in B$, $f(b) = C(b)$. Note that if a circuit is a **linear** circuit, representing and computing are equivalent tasks. But for **non-linear** circuits, they are not necessarily equivalent. Drucker [Dru12] showed that most linear operators can be represented by a circuit with $O(n)$ wires in depth-3, while [GHK+13] shows $\Omega(n \log\log n)$-wire lower

[4] most linear operators are "good" linear error correcting codes due to Gilbert-Varshamov bound holding for random linear operators [GRS22]. And there are explicit constructions of "good" linear error correcting codes.

bound for computing a linear operator, giving a separation for two tasks for non-linear circuits. Our result implies $\Omega(n \log^{1/3} n)$ lower bound.

4.2 Proof of Theorem 4.2

We use the following theorem to translate the cell-probe lower bound to circuit wire lower bound.

Theorem 4.3 (Theorem 2.3 of [Vio18])**.** *Suppose the operator $f : \{0,1\}^n \to \{0,1\}^m$ has a circuit of depth d with w wires, consisting of unbounded fan-in, arbitrary gates. Then f has a data structure with space $s = n + r$ and time $(w/r)^d$ for any r.*

We show the following lemma which follows from modifying the proof of Theorem 3.1. Note that this is stronger than what is necessary as we show a lower bound against small constant advantage over random guessing.

Lemma 4.4. *For random linear operator $M : \{0,1\}^n \to \{0,1\}^m$ with $m = 10^9 \cdot n$, any cell-probe data structure which correctly outputs $f_i(x)$ with probability at least $2/3$ for all $i \in [m]$ using $s = 1.01n$-space must have $t \geq \Omega(\log n)$.*

Note that Lemma 4.4 and Theorem 4.3 directly imply Theorem 4.2.

Proof Sketch. Here we sketch and highlight the required modification to the main proof. Suppose there exists a cell-probe data structure for most random linear operator $f : \{0,1\}^n \to \{0,1\}^m$ using $s = 1.01n$-space with $t \leq 0.1 \log n$.

We replace Protocol 3. We redefine P_i to represent the entire decision tree rather than a single path within it. Instead of sending a single path in the decision tree, as $t \leq 0.1 \log n$, we can afford to send the whole decision tree using $2^t \cdot \log s = o(n^{0.2})$-bits. Then Bob can compute the whole outcome of the decision tree using U. Furthermore, the correctness of the cell-probe data structure guarantees Alice and Bob correctly output the answer with probability $> 2/3$. Therefore, the random variable B_i is unnecessary and we can have Z_i as the output.

With this change, we leave the reader to verify that we can replace (15) and (16) in Theorem 3.1 as

$$I(M_{\mathcal{P}_J}; T | Z_{Alice} = z_{Alice}, E = 1, Z_{\rho_j}) \leq \frac{6s}{\ell} \leq \frac{1}{10^5} \tag{18}$$

$$\tilde{H}_\infty(M_{\rho_j} | Z_{\rho_j}, E = 1, Z_{Alice} = z_{Alice}) + \tilde{H}_\infty(X | Z_{\rho_j}, E = 1, Z_{Alice} = z_{Alice}) \geq 2n - n^{0.2} \tag{19}$$

while $\left| Z_{\rho_j} \cdot \chi_{M_{\rho_j}}(X) \right|$ is $\geq 1/3$ in expectation. This then violates Lemma 3.10 by taking γ as the appropriate constant. □

Remark 4.5. *The simple modification can also be adapted to a carefully manipulated version of the original argument in [KW20, DL20], thereby giving the same $t \geq \Omega(\log n)$ bound when the data structure is guaranteed to output the correct answer all the time. However, as noted from the sketch of the proof, the new argument works against circuits obtaining some $\Omega(1)$ advantage over random guessing. That is our proof works against the circuit C's such that for every $i \in [m]$*

$$\Pr_x [C_i(x) = f_i(x)] \geq \frac{1 + \Omega(1)}{2}$$

where C_i denotes the i-th output of the circuit C. The argument from [KW20, DL20] cannot be used to handle such small advantage regime.

5 Acknowledgment

We thank Sasha Golovnev for his feedback in the earlier draft of the paper. A part of this work was done while attending DIMACS Workshop on Lower Bounds and Frontiers in Data Structures 2022.

References

[AFKL19] Peyman Afshani, Casper Benjamin Freksen, Lior Kamma, and Kasper Green Larsen. Lower Bounds for Multiplication via Network Coding. In Christel Baier, Ioannis Chatzigiannakis, Paola Flocchini, and Stefano Leonardi, editors, *46th International Colloquium on Automata, Languages, and Programming (ICALP 2019)*, volume 132 of *Leibniz International Proceedings in Informatics (LIPIcs)*, pages 10:1–10:12, Dagstuhl, Germany, 2019. Schloss Dagstuhl–Leibniz-Zentrum fuer Informatik.

[Aga17] Pankaj K. Agarwal. Simplex Range Searching and Its Variants: A Review. In Martin Loebl, Jaroslav Nešetřil, and Robin Thomas, editors, *A Journey Through Discrete Mathematics*, pages 1–30. Springer International Publishing, Cham, 2017.

[BL15] Joshua Brody and Kasper Gren Larsen. Adapt or Die: Polynomial Lower Bounds for Non-Adaptive Dynamic Data Structures. *THEORY OF COMPUTING*, 11:19, 2015.

[BM13] Mark Braverman and Ankur Moitra. An information complexity approach to extended formulations. In *Proceedings of the forty-fifth annual ACM symposium on Theory of Computing*, pages 161–170, Palo Alto California USA, June 2013. ACM.

[Bra22] Mark Braverman. Communication and information complexity. *Proc. Int. Cong. Math. 2022*, 2022.

[BYBJK11] Ziv Bar-Yossef, Yitzhak Birk, T. S. Jayram, and Tomer Kol. Index Coding With Side Information. *IEEE Transactions on Information Theory*, 57(3):1479–1494, March 2011.

[CEEP12] Arkadev Chattopadhyay, Jeff Edmonds, Faith Ellen, and Toniann Pitassi. A Little Advice Can Be Very Helpful. In *Proceedings of the Twenty-Third Annual ACM-SIAM Symposium on Discrete Algorithms*, pages 615–625. Society for Industrial and Applied Mathematics, January 2012.

[CGK19] Henry Corrigan-Gibbs and Dmitry Kogan. The Function-Inversion Problem: Barriers and Opportunities. In Dennis Hofheinz and Alon Rosen, editors, *Theory of Cryptography*, volume 11891, pages 393–421. Springer International Publishing, Cham, 2019. Series Title: Lecture Notes in Computer Science.

[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.

[Cha90] Bernard Chazelle. Lower Bounds for Orthogonal Range Searching: Part II. The Arithmetic Model. *J. ACM*, 37(3):439–463, 1990.

[Che08a] Dmitriy Yu Cherukhin. Lower Bounds for Boolean Circuits with Finite Depth and Arbitrary Gates. *Electronic Colloquium on Computational Complexity (ECCC)*, TR08-032, 2008.

[Che08b] Dmitriy Yu. Cherukhin. Lower Bounds for Depth-2 and Depth-3 Boolean Circuits with Arbitrary Gates. In Edward A. Hirsch, Alexander A. Razborov, Alexei Semenov, and Anatol Slissenko, editors, *Computer Science – Theory and Applications*, volume 5010, pages 122–133. Springer Berlin Heidelberg, Berlin, Heidelberg, 2008. Series Title: Lecture Notes in Computer Science.

[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.

[DDPW83] Danny Dolev, Cynthia Dwork, Nicholas Pippenger, and Avi Wigderson. Superconcentrators, generalizers and generalized connectors with limited depth. In *Proceedings of the fifteenth annual ACM symposium on Theory of computing - STOC '83*, pages 42–51, Not Known, 1983. ACM Press.

[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.

[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. Artwork Size: 15 pages, 547135 bytes ISBN: 9783959771740 Medium: application/pdf Publisher: Schloss Dagstuhl – Leibniz-Zentrum für Informatik.

[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.

[Dru12] Andrew Drucker. Limitations of Lower-Bound Methods for the Wire Complexity of Boolean Operators. In *2012 IEEE 27th Conference on Computational Complexity*, pages 170–180, June 2012. ISSN: 1093-0159.

[Fre81] Michael L. Fredman. A Lower Bound on the Complexity of Orthogonal Range Queries. *J. ACM*, 28(4):696–705, 1981.

[GHK+13] Anna Gál, Kristoffer Arnsfelt Hansen, Michal Koucký, Pavel Pudlák, and Emanuele Viola. Tight Bounds on Computing Error-Correcting Codes by Bounded-Depth Circuits With Arbitrary Gates. *IEEE Transactions on Information Theory*, 59(10):6611–6627, October 2013.

[GPRW22] Alexander Golovnev, Gleb Posobin, Oded Regev, and Omri Weinstein. Polynomial Data Structure Lower Bounds in the Group Model. *SIAM Journal on Computing*, pages FOCS20–74–FOCS20–101, March 2022.

[GRS22] Venkatesan Guruswami, Atri Rudra, and Madhu Sudan. *Essential Coding Theory*. 2022.

[JS10] S. Jukna and G. Schnitger. Circuits with arbitrary gates for random operators, 2010. Version Number: 1.

[JS11] Stasys Jukna and Georg Schnitger. Min-rank conjecture for log-depth circuits. *J. Comput. Syst. Sci.*, 77(6):1023–1038, 2011.

[Juk10] Stasys Jukna. Representing (0, 1)-matrices by boolean circuits. *Discrete Mathematics*, 310(1):184–187, January 2010.

[Juk12] Stasys Jukna. *Boolean function complexity: advances and frontiers*, volume 27. Springer Science & Business Media, 2012.

[KU11] Kiran S. Kedlaya and Christopher Umans. Fast Polynomial Factorization and Modular Composition. *SIAM J. Comput.*, 40(6):1767–1802, 2011.

[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.

[Lar14] Kasper Green Larsen. On Range Searching in the Group Model and Combinatorial Discrepancy. *SIAM J. Comput.*, 43(2):673–686, 2014.

[Lok09] Satyanarayana V. Lokam. Complexity Lower Bounds using Linear Algebra. *Found. Trends Theor. Comput. Sci.*, 4(1-2):1–155, 2009.

[Lup56] Oleg B Lupanov. On rectifier and switching-and-rectifier schemes. In *Dokl. Akad. Nauk SSSR*, volume 111, pages 1171–1174, 1956. Issue: 6.

[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.

[Mil93] Peter Bro Miltersen. The Bit Probe Complexity Measure Revisited. In *STACS 1993*, pages 662–671, 1993.

[MNSW98] Peter Bro Miltersen, Noam Nisan, Shmuel Safra, and Avi Wigderson. On Data Structures and Asymmetric Communication Complexity. *J. Comput. Syst. Sci.*, 57(1):37–49, 1998.

[NRR20] Sivaramakrishnan Natarajan Ramamoorthy and Cyrus Rashtchian. Equivalence of Systematic Linear Data Structures and Matrix Rigidity. *LIPIcs, Volume 151, ITCS 2020*, 151:35:1–35:20, 2020. Artwork Size: 20 pages, 605863 bytes ISBN: 9783959771344 Medium: application/pdf Publisher: Schloss Dagstuhl – Leibniz-Zentrum für Informatik Version Number: 1.0.

[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.

[Pip77] Nicholas Pippenger. Superconcentrators. *SIAM Journal on Computing*, 6(2):298–304, June 1977.

[PR94] P. Pudlák and V. Rödl. Some combinatorial-algebraic problems from complexity theory. *Discrete Mathematics*, 136(1):253–279, December 1994.

[PTW08] Rina Panigrahy, Kunal Talwar, and Udi Wieder. A Geometric Approach to Lower Bounds for Approximate Near-Neighbor Search and Partial Match. In *2008 49th Annual IEEE Symposium on Foundations of Computer Science*, pages 414–423, Philadelphia, PA, USA, October 2008. IEEE.

[Pud94] P. Pudlak. Communication in bounded depth circuits. *Combinatorica*, 14(2):203–216, June 1994.

[Pă11] Mihai Pătraşcu. Unifying the Landscape of Cell-Probe Lower Bounds. *SIAM J. Comput.*, 40(3):827–847, 2011.

[Ram20] C. Ramya. Recent Progress on Matrix Rigidity – A Survey, September 2020. arXiv:2009.09460 [cs].

[RS03] Ran Raz and Amir Shpilka. Lower Bounds for Matrix Product in Bounded Depth Circuits with Arbitrary Gates. *SIAM Journal on Computing*, 32(2):488–513, January 2003.

[RTS00] Jaikumar Radhakrishnan and Amnon Ta-Shma. Bounds for Dispersers, Extractors, and Depth-Two Superconcentrators. *SIAM Journal on Discrete Mathematics*, 13(1):2–24, January 2000.

[Sha49] Claude. E. Shannon. The synthesis of two-terminal switching circuits. *The Bell System Technical Journal*, 28(1):59–98, January 1949. Conference Name: The Bell System Technical Journal.

[Sie04] Alan Siegel. On Universal Classes of Extremely Random Constant-Time Hash Functions. *SIAM J. Comput.*, 33(3):505–543, 2004.

[Tho13] Mikkel Thorup. Mihai PăTraşCu: Obituary and Open Problems. *SIGACT News*, 44(1):110–114, March 2013.

[Val75] Leslie G. Valiant. On non-linear lower bounds in computational complexity. In *Proceedings of seventh annual ACM symposium on Theory of computing - STOC '75*, pages 45–53, Albuquerque, New Mexico, United States, 1975. ACM Press.

[Val77] Leslie G. Valiant. Graph-theoretic arguments in low-level complexity. In G. Goos, J. Hartmanis, P. Brinch Hansen, D. Gries, C. Moler, G. Seegmüller, J. Stoer, N. Wirth, and Jozef Gruska, editors, *Mathematical Foundations of Computer Science 1977*, volume 53, pages 162–176. Springer Berlin Heidelberg, Berlin, Heidelberg, 1977.

[Val92] Leslie G. Valiant. Why is Boolean Complexity Theory Difficult? In *Poceedings of the London Mathematical Society Symposium on Boolean Function Complexity*, pages 84–94, New York, NY, USA, 1992. Cambridge University Press. event-place: London, United Kingdom.

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

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

A Omitted Proof

We restate the lemma.

Lemma 3.3. *If* $|\mathcal{P}| = \ell = \frac{|\mathcal{Q}|}{100} = \frac{m}{100}$ *then*

$$I(M_{\mathcal{P}_J}; Z_{Alice}) = E_{z_{Alice}}\left[D(M_{\mathcal{P}_J}|_{Z_{Alice}=z_{Alice}}||M_{\mathcal{P}_J})\right] \leq 3 \cdot t \log s$$

Proof. We plug in the definition of Z_{Alice}.

$$\begin{aligned} &I(M_{\mathcal{P}_J}; Z_{Alice}) = I(M_{\mathcal{P}_J}; \mathcal{P}, J, P_{\mathcal{P}_{<J}}, M_{\mathcal{P}_{<J}}, P_{\mathcal{P}_J},) \\ &= I(M_{\mathcal{P}_J}; \mathcal{P}, J, P_{\mathcal{P}_{<J}}, M_{\mathcal{P}_{<J}}) + \underbrace{I(M_{\mathcal{P}_J}; P_{\mathcal{P}_J}|\mathcal{P}, J, P_{\mathcal{P}_{<J}}, M_{\mathcal{P}_{<J}})}_{\leq t\log s + t + 1} \\ &\leq I(M_{\mathcal{P}_J}; \mathcal{P}, J, P_{\mathcal{P}_{<J}}, M_{\mathcal{P}_{<J}}) + t\log s + t + 1 \end{aligned}$$

Now we upper bound $I(M_{\mathcal{P}_J}; \mathcal{P}, J, P_{\mathcal{P}_{<J}}, M_{\mathcal{P}_{<J}})$ term. We first know that due to J and $\mathcal{P}$ being chosen independently at random

$$I(M_{\mathcal{P}_J}; \mathcal{P}J, P_{\mathcal{P}_{<J}}, M_{\mathcal{P}_{<J}}) = I(M_{\mathcal{P}_J}; \mathcal{P}, P_{\mathcal{P}_{<J}}, M_{\mathcal{P}_{<J}}|\mathcal{P}, J)$$

Now consider a fixed $J = j$.

$$I(M_{\mathcal{P}_J}; \mathcal{P}, P_{\mathcal{P}_{<j}}, M_{\mathcal{P}_{<J}}|\mathcal{P}, J = j) = I(M_{\mathcal{P}_J}; P_{\mathcal{P}_{<j}}, M_{\mathcal{P}_{<J}}|\mathcal{P}) = I(M_{\mathcal{P}_J}; P_{\mathcal{P}_{<j}}|M_{\mathcal{P}_{<J}}, \mathcal{P})$$

Now consider fixed $\mathcal{P}_{<j} = \rho_{<j}$. Then the above term becomes

$$\begin{aligned} &I(M_{\mathcal{P}_J}; P_{\rho_{<j}}|M_{\rho_{<j}}, \mathcal{P}_{<j} = \rho_{<j}, \mathcal{P}_j) = \frac{1}{m-(j-1)} \sum_{i \notin \rho_{<j}} I(M_i; P_{\rho_{<j}}|M_{\rho_{<j}}) \\ &\leq \frac{1}{m-(j-1)} I(S_{[m]-\rho_{<j}}; P_{\rho_{<j}}|M_{\rho_{<j}}) \leq \frac{j}{m-\ell}(t\log s + t + 1) \leq 0.1t\log s \end{aligned}$$

due to Fact 2.8 and the last inequality holds due to our choice of ℓ

As the inequality holds for any fixed j and $\mathcal{P}_{<j} = \rho_{<j}$,

$$I(M_{\mathcal{P}_J}; Z_{Alice}) \leq t\log s + t + 1 + 0.1t\log s \leq 3 \cdot t \log s$$

completing the proof of our lemma.

□

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