Lower Bounds for Semi-adaptive Data Structures via Corruption

Pavel Dvořák
Charles University, Prague, Czech Republic
koblich@iuuk.mff.cuni.cz

Bruno Loff
INESC-Tec and University of Porto, Portugal
bruno.loff@gmail.com

─── **Abstract** ───

In a dynamic data structure problem we wish to maintain an encoding of some data in memory, in such a way that we may efficiently carry out a sequence of queries and updates to the data. A long-standing open problem in this area is to prove an unconditional polynomial lower bound of a trade-off between the update time and the query time of an adaptive dynamic data structure computing some explicit function. Ko and Weinstein provided such lower bound for a restricted class of *semi-adaptive* data structures, which compute the Disjointness function. There, the data are subsets $x_1, \ldots, x_k$ and y of $\{1, \ldots, n\}$, the updates can modify y (by inserting and removing elements), and the queries are an index $i \in \{1, \ldots, k\}$ (query i should answer whether x_i and y are disjoint, i.e., it should compute the Disjointness function applied to (x_i, y)). The semi-adaptiveness places a restriction in how the data structure can be accessed in order to answer a query. We generalize the lower bound of Ko and Weinstein to work not just for the Disjointness, but for any function having high complexity under the smooth corruption bound.

1 Introduction

In a dynamic data structure problem we wish to maintain an encoding of some data in memory, in such a way that we may efficiently carry out a sequence of queries and updates to the data. A suitable computational model to study dynamic data structures is the cell probe model of Yao [21]. Here we think of the memory divided into registers, or *cells*, where each cell can carry w bits, and we measure efficiency by counting the number of memory accesses, or *probes*, needed for each query and each update – these are respectively called the *query time* t_q and *update time* t_u. The main goal of this line of research is to understand the inherent trade-off between w, t_q and t_u, for various interesting problems. Specifically, one would like to show lower bounds on $t = \max\{t_q, t_u\}$ for reasonable choices of w (which is typically logarithmic in the size of the data).

The first lower bound for this setting was proven by Fredman and Saks [8], which proved $t = \Omega\big(\log n / \log \log n\big)$ for various problems. These lower bounds were successively improved [15, 17, 13, 14], and we are now able to show that certain problems with non-Boolean queries require $t = \Omega\big((\log n / \log \log n)^2\big)$, and certain problems with Boolean queries require $t = \Omega\big((\log n / \log \log n)^{3/2}\big)$.

The major unsolved question in this area is to prove a polynomial lower bound on t. For example, consider the dynamic reachability problem, where we wish to maintain a directed n-vertex graph in memory, under edge insertions and deletions, while being able to answer reachability queries ("*is vertex i connected to vertex j?*"). Is it true that any scheme for the dynamic reachability problem requires $t = \Omega(n^\delta)$, for some constant $\delta > 0$? Indeed, such a lower bound is known under various complexity-theoretic assumptions[1], the question is whether such a lower bound may be proven unconditionally.

In an influential paper [18], Mihai Pătraşcu proposed an approach to this unsolved question. He defines a data structure problem, called the *multiphase problem*. Let us represent partial functions $f : \{0,1\}^n \times \{0,1\}^n \to \{0,1\}$ as total functions $f' : \{0,1\}^n \times \{0,1\}^n \to \{0,1,*\}$ where $f'(x,y) = *$ if $f(x,y)$ is not defined. Then associated with a partial Boolean function $f : \{0,1\}^n \times \{0,1\}^n \to \{0,1,*\}$, and a natural number $k \geq 1$, we may define a corresponding *multiphase problem of f* as the following dynamic process:

Phase I - Initialization. We are given k inputs $x_1, \ldots, x_k \in \{0,1\}^n$, and are allowed to preprocess this input in time $nk \cdot t_{\mathrm{p}}$.

Phase II - Update. We are then given another input $y \in \{0,1\}^n$, and we have time $n \cdot t_{\mathrm{u}}$ to read and update the memory locations from the data structure constructed in Phase I.

Phase III - Query. Finally, we are given a query $i \in [k]$, we have time t_{q} to answer the question whether $f(x_i, y) = 1$. If $f(x_i, y)$ is not defined, the answer can be arbitrary.

Typically we will have $k = \mathrm{poly}(n)$. Let us be more precise, and consider randomized solutions to the above problem.

▶ **Definition 1** (Scheme for the multiphase problem of f). *Let $f : \{0,1\}^n \times \{0,1\}^n \to \{0,1,*\}$ be a partial Boolean function. A* scheme *for the multiphase problem of f with preprocessing time t_{p}, update time t_{u} and query time t_{q} is a triple $D = \big(E, \{U_y\}_{y \in \{0,1\}^n}, \{Q_i\}_{i \in [k]}\big)$, where:*

- $E : \big(\{0,1\}^n\big)^k \to \big(\{0,1\}^w\big)^s$ *maps the input x to the memory contents $E(x)$, where each of the s memory locations holds w bits. E must be computed in time $nk \cdot t_{\mathrm{p}}$ by a Random-Access Machine (RAM).*
- *For each $y \in \{0,1\}^n$, $U_y : \big(\{0,1\}^w\big)^s \to \big(\{0,1\}^w\big)^u$ is a decision-tree of depth $\leq n \cdot t_{\mathrm{u}}$, which reads $E(x)$ and produces a sequence $U_y\big(E(x)\big)$ of u updates.[2]*
- *For each $i \in [k]$, $Q_i : \big(\{0,1\}^w\big)^s \times \big(\{0,1\}^w\big)^u \to \{0,1\}$ is a decision-tree of depth $\leq t_{\mathrm{q}}$.[3]*
- *For all $x \in \big(\{0,1\}^n\big)^k$, $y \in \{0,1\}^n$, and $i \in [k]$,*

$$f(x_i, y) \neq * \implies Q_i\big(E(x), U_y(E(x))\big) = f(x_i, y).$$

---

[1] See [16, 1]. Strictly speaking, these conditional lower bounds only work if the preprocessing time, which is the time taken to encode the data into memory, is also bounded. But we will ignore this distinction.

[2] In the usual way of defining the update phase, we have a read/write decision-tree U_y which changes the very same cells that it reads. But when $w = \Omega(\log s)$, this can be seen to be equivalent, up to constant factors, to the definition we present here, where we have a decision-tree U_y that writes the updates on a separate location. In order to simulate a scheme that uses a read/write decision-tree, we may use a hash table with $O(1)$ worst-case lookup time, such as cuckoo hashing. Then we have a read-only decision-tree $U_y'(E(x))$ whose output is the hash table containing all the $i \in [s]$ which were updated by $U_y(E(x))$, associated with their final value in the execution of $U_y(E(x))$. Note that the hash table itself is static.

[3] All our results will hold even if Q_i is allowed to depend arbitrarily on x_i. This makes for a less natural model, however, so we omit this from the definitions.

In a randomized scheme *for the multiphase problem of f, each U_y and Q_i are distributions over decision trees, and it must hold that for all $x \in \left(\{0,1\}^n\right)^k$, $y \in \{0,1\}^n$, and $i \in [k]$,*

$$f(x_i, y) \neq * \implies \Pr_{Q_i, U_y}\left[Q_i\big(E(x), U_y(E(x))\big) = f(x_i, y)\right] \geq 1 - \varepsilon.$$

The value ε is called the error probability *of the scheme.*

Pătraşcu [18] considered this problem where $f = \mathsf{DISJ}$ is the Disjointness function:

$$\mathsf{DISJ}(x, y) = \begin{cases} 0 & \text{if there exists } i \in [n] \text{ such that } x_i = y_i = 1 \\ 1 & \text{otherwise} \end{cases}$$

He conjectured that any scheme for the multiphase problem of DISJ will necessarily have $\max\{t_\mathrm{p}, t_\mathrm{u}, t_\mathrm{q}\} \geq n^\delta$ for some constant $\delta > 0$.

Pătraşcu shows that such lower bounds on the multiphase problem for DISJ would imply polynomial lower bounds for various dynamic data structure problems. For example such lower bounds would imply that dynamic reachability requires $t = \Omega(n^\delta)$. He also shows that these lower bounds hold true under the assumption that 3SUM has no sub-quadratic algorithms.

Finally, Pătraşcu then defines a 3-player Number-On-Forehead (NOF) communication game, such that lower bounds on this game imply matching lower bounds for the multiphase problem. The game associated with a function $f : \{0,1\}^n \times \{0,1\}^n \to \{0,1\}$ is as follows:

1. Alice is given $x_1, \ldots, x_k \in \{0,1\}^n$ and $i \in [k]$, Bob gets $y \in \{0,1\}^n$ and $i \in [k]$ and Charlie gets $x_1, \ldots, x_k$ and y.
2. Charlie sends a private message of ℓ_1 bits to Bob and then he is silent.
3. Alice and Bob communicate ℓ_2 bits and want to compute $f(x_i, y)$.

Pătraşcu [18] conjectured that if ℓ_1 is $o(k)$, then ℓ_2 has to be bigger than the communication complexity of f. However, this conjecture turned out to be false. The randomized communication complexity of DISJ is $\Omega(n)$ [19, 10, 3], but Chattopadhyay et al. [6] construct a protocol for $f = \mathsf{DISJ}$ where both $\ell_1, \ell_2 = O\big(\sqrt{n} \cdot \log k\big)$. They further show that any randomized scheme in the above model can be derandomized.

So the above communication model is more powerful than it appears at first glance.[4] However, a recent paper by Ko and Weinstein [11] succeeds in proving lower bounds for a simpler version of the multiphase problem, which translate to lower bounds for a restricted class of dynamic data structure schemes. They manage to prove a lower bound of $\Omega(\sqrt{n})$ for the simpler version of the multiphase problem which is associated with the Disjointness function $f = \mathsf{DISJ}$. Our paper generalizes their result:

- We generalize their lower bound to any function f having large complexity according to the smooth corruption bound, under a product distribution. Disjointness is such a function [2], but so is Inner Product, Gap Orthogonality, and Gap Hamming Distance [20].
- The new lower-bounds we obtain (for Inner-product, Gap Orthogonality, and Gap Hamming Distance) are stronger – $\Omega(n)$ instead of the lower-bound $\Omega(\sqrt{n})$ for disjointness. As far as was known before our result, it could well have been that every function had a scheme for the simpler version of the multiphase problem using only $O(\sqrt{n})$ communication.

---

[4] The conjecture remains that if $\ell_1 = o(k)$, then ℓ_2 has to be larger than the maximum distributional communication complexity of f under a product distribution. This is $\tilde{\Theta}\big(\sqrt{n}\big)$ for Disjointness [2].

▬ Ko and Weinstein derive their lower-bound via a cut-and-paste lemma which works specifically for disjointness. This cut-and-paste lemma is a more robust version of the one appearing in [3], made to work not only for protocols, where the inputs $\mathbf{x}$ and $\mathbf{y}$ are independent given the transcript $\mathbf{z}$ of the protocol, but also for random-variables that are "protocol-like", namely any $(\mathbf{x}, \mathbf{y}, \mathbf{z})$ where $I(\mathbf{x} : \mathbf{y} \mid \mathbf{z})$ is close to 0. Instead, we directly derive the existence of a large nearly-monochromatic rectangle, from the existence of such protocol-like random-variables, which is what then allows us to use the smooth corruption bound. This result is our core technical contribution, and may be of independent interest.

All of the above lower bounds will be shown to hold also for randomized schemes, and not just for deterministic schemes.

1.1 Semi-adaptive Multiphase Problem

Let us provide rigorous definitions.

▶ **Definition 2** (Semi-adaptive random data structure [11])**.** *Let* $f : \{0,1\}^n \times \{0,1\}^n \to \{0,1,*\}$ *be a partial function. A scheme* $D = \left(E, \{U_y\}_{y \in \{0,1\}^n}, \{Q_i\}_{i \in [k]}\right)$ *for the multiphase problem of* f *is called* semi-adaptive *if any path on the decision-tree* $Q_i : \left(\{0,1\}^w\right)^s \times \left(\{0,1\}^w\right)^u \to \{0,1\}$ *first queries the first part of the input (the* $E(x)$ *part), and then queries the second part of the input (the* $U(E(x))$ *part). If* D *is randomized, then this property must hold for every randomized choice of* Q_i*.*

We point out that the reading of the cells in each part is completely adaptive. The restriction is only that the data structure cannot read cells of $E(x)$ if it already started to read cells of $U(E(x))$. Ko and Weinstein state their result for deterministic data structures, i.e., $\varepsilon = 0$ thus the data structure always returns the correct answer.

▶ **Theorem 3** (Theorem 4.9 of Ko and Weinstein [11])**.** *Let* $k \geq \omega(n)$*. Any semi-adaptive deterministic data structure that solves the multiphase problem of the* DISJ *function, must have either* $t_u \cdot n \geq \Omega(k/w)$ *or* $t_q \geq \Omega(\sqrt{n}/w)$*.*

To prove the lower bound they reduce the semi-adaptive data structure into a low correlation random process.

▶ **Theorem 4** (Reformulation of Lemma 4.1 of Ko and Weinstein [11])**.** *Let* $\mathbf{x}_1, \ldots, \mathbf{x}_k$ *be random variables over* $\{0,1\}^n$ *and each of them is independently distributed according to the same distribution* μ_1 *and let* $\mathbf{y}$ *be a random variable over* $\{0,1\}^n$ *distributed according to* μ_2 *(independently of* $\mathbf{x}_1, \ldots, \mathbf{x}_k$*). Let* D *be a randomized semi-adaptive scheme for the multiphase problem for a partial function* $f : \{0,1\}^n \times \{0,1\}^n \to \{0,1,*\}$ *with error probability bounded by* ε*. Then, for any* $p \leq o(k)$ *there is a random variable* $\mathbf{z} \in \{0,1\}^m$ *and* $i \in [k]$ *such that:*
1. $\Pr\left[f(\mathbf{x}_i, \mathbf{y}) \neq *, \ \mathbf{z}_m \neq f(\mathbf{x}_i, \mathbf{y})\right] \leq \varepsilon$.
2. $I\left(\mathbf{x}_i : \mathbf{y} \, \mathbf{z}\right) \leq t_q \cdot w + o(t_q \cdot w)$.
3. $I\left(\mathbf{y} : \mathbf{z}\right) \leq t_q \cdot w$.
4. $I\left(\mathbf{x}_i : \mathbf{y} \mid \mathbf{z}\right) \leq O\left(\frac{t_u \cdot n \cdot w}{p}\right)$.

Above, I stands for mutual information between random variables, see Section 2.2 for the definition. The random variable $\mathbf{z}$ consists of some $\mathbf{x}_j$'s and transcripts of query phases of D for some $j \in [k]$. The theorem can be interpreted as saying that the last bit of $\mathbf{z}$ predicts $f(\mathbf{x}_i, \mathbf{y})$, $\mathbf{z}$ has little information about $\mathbf{x}_i$ and $\mathbf{y}$, and the triple $(\mathbf{x}_i, \mathbf{y}, \mathbf{z})$ is "protocol-like", in the sense that $\mathbf{x}_i$ and $\mathbf{y}$ are close to being independent given $\mathbf{z}$. Ko and Weinstein [11]

proved Theorem 4 for the deterministic schemes for the DISJ function and in the case where $\mu_1 = \mu_2$. However, their proof actually works for any (partial) function f and for any two, possibly distinct distributions μ_1 and μ_2. Moreover, their proof also works for randomized schemes. The resulting statement for randomized schemes for any function f is what we have given above. To complete the proof of their lower bound, Ko and Weinstein proved that if we set p (and k) large enough so that $I(\mathbf{x}_i : \mathbf{y} \mid \mathbf{z}) \leq o(1)$ then such random variable $\mathbf{z}$ cannot exist when f is the DISJ function. It is this second step which we generalize.

Let $f : X \times Y \to \{0,1\}$ be a function and μ be a distribution over $X \times Y$. A set $R \subseteq X \times Y$ is a *rectangle* if there exist sets $A \subseteq X$ and $B \subseteq Y$ such that $R = A \times B$. For $b \in \{0,1\}$ and $0 \leq \rho \leq 1$, we say the rectangle R is *ρ-error b-monochromatic for f under μ* if $\mu\big(R \cap f^{-1}(1-b)\big) \leq \rho \cdot \mu\big(R\big)$. We say the distribution μ is a *product distribution* if there are two independent distribution μ_1 over X and μ_2 over Y such that $\mu(x,y) = \mu_1(x) \times \mu_2(y)$. For $0 \leq \alpha \leq \frac{1}{2}$, the distribution μ is *α-balanced according to f* if $\mu\big(f^{-1}(0)\big), \mu\big(f^{-1}(1)\big) \geq \alpha$. We will prove that the existence of a random variable $\mathbf{z}$ given by Theorem 4 implies that, for any $b \in \{0,1\}$, any balanced product distribution μ and any function g which is "close" to f, there is a large (according to μ) ρ-error b-monochromatic rectangle for g in terms of t_q. This technique is known as smooth corruption bound [4, 5] or smooth rectangle bound [9]. We denote the smooth corruption bound of f as $\mathsf{scb}_\mu^{\rho,\lambda}$. Informally, $\mathsf{scb}_\mu^{\rho,\lambda}(f) \geq s$ if there is $b \in \{0,1\}$ and a partial function $g : X \times Y \to \{0,1,*\}$ which is close[5] to f such that any ρ-error b-monochromatic rectangle $R \subseteq X \times Y$ for g has size (under μ) at most 2^{-s}. We will define smooth corruption bound formally in the next section. Thus, if we use Theorem 4 as a black box we generalize Theorem 3 for any function of large corruption bound.

▶ **Theorem 5** (Main Result). *Let $\lambda, \tilde{\varepsilon}, \tilde{\alpha} \geq 0$ such that $\alpha \geq 2\varepsilon$ for $\varepsilon = \tilde{\varepsilon} + \lambda, \alpha = \tilde{\alpha} - \lambda$. Let μ be a product distribution over $\{0,1\}^n \times \{0,1\}^n$ such that μ is $\tilde{\alpha}$-balanced according to a partial function $f : \{0,1\}^n \times \{0,1\}^n \to \{0,1,*\}$. Any semi-adaptive randomized scheme for the multiphase problem of f, with error probability bounded by $\tilde{\varepsilon}$, must have either $t_\mathrm{u} \cdot n \geq \Omega\big(k/w\big)$, or*

$$t_\mathrm{q} \cdot w \geq \Omega\left(\alpha \cdot \mathsf{scb}_\mu^{O(\varepsilon/\alpha),\lambda}(f)\right).$$

We point out that Ω and O in the bound given above hide absolute constants independent of α, ε and λ.

As a consequence of our main result, and of previously-known bounds on corruption, we are able to show new lower-bounds of $t_\mathrm{q} = \Omega(\frac{n}{w})$ against semi-adaptive schemes for the multiphase problem of the Inner Product, Gap Orthogonality and Gap Hamming Distance functions (where the gap is $\sqrt{n}$). These lower-bounds hold assuming that $t_\mathrm{u} = o(\frac{k}{wn})$. They follow from the small discrepancy of Inner Product, and from a bound shown by Sherstov on the corruption of Gap Orthogonality, followed by a reduction to Gap Hamming Distance [20]. This result also gives an alternative proof of the same lower-bound proven by Ko and Weinstein [11], for the Disjointness function, of $t_\mathrm{q} = \Omega(\frac{\sqrt{n}}{w})$. This follows from the bound on corruption of Disjointness under a product distribution, shown by Babai et al. [2].

The paper is organized as follows. In Section 2 we give important notation, and the basic definitions from information theory and communication complexity. The proof of Theorem 5 appears in Section 3. The various applications appear in Section 4.

---

[5] "Closeness" is measured by the parameter $\lambda \in \mathbb{R}$, see Section 2.1 for the formal definition.

2 Preliminaries

We use a notational scheme where sets are denoted by uppercase letters, such as X and Y, elements of the sets are denoted by the same lowercase letters, such as $x \in X$ and $y \in Y$, and random variables are denoted by the same lowercase boldface letters, such as $\mathbf{x}$ and $\mathbf{y}$. We will use lowercase greek letters, such as μ, to denote distributions. If μ is a distribution over a product set, such as $X \times Y \times Z$, and $(x, y, z) \in X \times Y \times Z$, then $\mu(x, y, z)$ is the probability of seeing (x, y, z) under μ. We will sometimes denote μ by $\mu(\mathrm{x}, \mathrm{y}, \mathrm{z})$, using non-italicized lowercase letters corresponding to $X \times Y \times Z$. This allows us to to use the notation $\mu(\mathrm{x})$ and $\mu(\mathrm{y})$ to denote the x and y-marginals of μ, for example; then if we use the same notation with italicized lowercase letters, we get the marginal probabilities, i.e., for each $x \in X$ and $y \in Y$

$$\mu(x) = \sum_{y,z} \mu(x, y, z) \qquad \mu(y) = \sum_{x,z} \mu(x, y, z).$$

If $y \in Y$, then we will also use the notation $\mu(\mathrm{x} \mid y)$ to denote the x-marginal of μ conditioned seeing the specific value y. Then for each $x \in X$ and $y \in Y$, we have

$$\mu(x \mid y) = \sum_z \mu(x, y, z).$$

We will also write $(\mathbf{x}, \mathbf{y}, \mathbf{z}) \sim \mu$ to mean that $(\mathbf{x}, \mathbf{y}, \mathbf{z})$ are random variables chosen according to the distribution $\mu(\mathrm{x}, \mathrm{y}, \mathrm{z})$, i.e., for all $(x, y, z) \in X \times Y \times Z$, $\Pr[\mathbf{x} = x, \mathbf{y} = y, \mathbf{z} = z] = \mu(x, y, z)$. Naturally if $A \subseteq X \times Y \times Z$, then $\mu(A) = \sum_{(x,y,z) \in A} \mu(x, y, z)$. We let $\mathsf{supp}(\mu)$ denote the *support* of μ, i.e., the set of (x, y, z) with $\mu(x, y, z) > 0$.

We now formally define the smooth corruption bound and related measures from communication complexity, and refer the book by Kushilevitz and Nisan [12] for more details. At the end of this section we provide necessary notions of information theory which are used in the paper, and for more details on these we refer to the book by Cover and Thomas [7].

2.1 Rectangle Measures

Let $f : X \times Y \to \{0, 1, *\}$ be a partial function, where $f(x, y) = *$ means f is not defined on (x, y). Let $\mu(\mathrm{x}, \mathrm{y})$ be a distribution over $X \times Y$. We say that f is λ-close to a partial function $g : X \times Y \to \{0, 1, *\}$ under μ if

$$\Pr_{(x,y) \sim \mu} \big[f(x, y) \neq g(x, y) \big] \leq \lambda.$$

For $b \in \{0, 1\}$, $\rho \in [0, 1]$, let

$$\mathcal{R}_\mu^{\rho,b}(f) = \big\{ R \subseteq X \times Y \text{ rectangle} \mid \mu\big(R \cap f^{-1}(1 - b)\big) \leq \rho \cdot \mu(R) \big\}$$

be the set of ρ-error b-monochromatic rectangles for f under μ. The complexity measure mono quantifies how large almost b-monochromatic rectangles can be for both $b \in \{0, 1\}$:

$$\mathsf{mono}_\mu^\rho(f) = \min_{b \in \{0,1\}} \max_{R \in \mathcal{R}_\mu^{\rho,b}(f)} \mu(R)$$

Using mono we can define the *corruption bound* of a function as $\mathsf{cb}_\mu^\rho(f) = \log \frac{1}{\mathsf{mono}_\mu^\rho(f)}$ and the *smooth corruption bound* as

$$\mathsf{scb}_\mu^{\rho,\lambda}(f) = \max_{g: \ \lambda\text{-close to } f \text{ under } \mu} \mathsf{cb}_\mu^\rho(g).$$

Thus, if $\mathsf{scb}_\mu^{\rho,\lambda}(f) \geq s$ then there is a $b \in \{0, 1\}$ and a function g which λ-close to f under μ such that for any ρ-error b-monochromatic rectangle for g under μ it holds that $\mu(R) \leq 2^{-s}$.

▶ **Remark.** In Razborov's paper where an $\Omega(n)$ lower-bound for disjointness is first proven [19], the (implicitly given) definition of a ρ-error b-monochromatic rectangle is $\mu(R \cap f^{-1}(1 - b)) \leq \rho \cdot \mu(R \cap f^{-1}(b))$. Later, a strong direct product theorem for corruption (under product distributions) was proven by Beame et al. [4], which uses instead the condition that $\mu(R \setminus f^{-1}(b)) \leq \rho \cdot \mu(R)$. The definition we present above comes from [20], where the condition is (we repeat it here) that $\mu(R \cap f^{-1}(1 - b)) \leq \rho \cdot \mu(R)$. So we have three different definitions of ρ-error b-monochromatic rectangle, and thus three different corruption bounds. Now, if the distribution μ is supported on the domain of f, all these three definitions result in (roughly) equivalent complexity measures. But if μ attributes some mass to inputs where f is undefined (which is sometimes useful if μ is a product distribution, as in our case), then the definitions are no longer equivalent. Our lower-bound will hold for any of the definitions, but the proof is somewhat simpler for the definition used in Sherstov's paper [20], which is the only corruption-based lower-bound we use, where μ attributes mass to undefined inputs.

The notion mono_μ^ρ is related to the *discrepancy* of a function:

$$\mathsf{disc}_\mu(f) = \max_{R \,:\, \text{rectangle of } X \times Y} \left| \mu\big(R \cap f^{-1}(0)\big) - \mu\big(R \cap f^{-1}(1)\big) \right|.$$

It is easy to see that for a total function f holds that $\mathsf{disc}_\mu(f) \geq (1 - 2\rho) \cdot \mathsf{mono}_\mu^\rho(f)$ for any ρ. Thus, Theorem 5 will give us lower bounds also for functions of small discrepancy.

2.2 Information Theory

We define several measures from information theory. If $\mu'(\mathrm{z}), \mu(\mathrm{z})$ are two distributions such that $\mathsf{supp}(\mu') \subseteq \mathsf{supp}(\mu)$, then the *Kullback-Leibler divergence* of μ' from μ is

$$\mathsf{D}_{\mathsf{KL}}\big(\mu' \parallel \mu\big) = \sum_z \mu'(z) \log \frac{\mu'(z)}{\mu(z)}.$$

With Kullback-Leibler divergence we can define the mutual information, which measures how close (according to KL divergence) is a joint distribution to the product of its marginals. If we have two random variables $(\mathbf{x}, \mathbf{y}) \sim \mu(\mathrm{x}, \mathrm{y})$, then we define their *mutual information* to be

$$I\big(\mathbf{x} : \mathbf{y}\big) = \mathsf{D}_{\mathsf{KL}}\big(\mu(\mathrm{x}, \mathrm{y}) \parallel \mu(\mathrm{x}) \times \mu(\mathrm{y})\big) = \underset{y \sim \mu(\mathrm{y})}{\mathbb{E}} \left[\mathsf{D}_{\mathsf{KL}}\big(\mu(\mathrm{x} \mid y) \parallel \mu(\mathrm{x})\big) \right].$$

If we have three random variables $(\mathbf{x}, \mathbf{y}, \mathbf{z}) \sim \mu(\mathrm{x}, \mathrm{y}, \mathrm{z})$, then the *mutual information of $\mathbf{x}$ and $\mathbf{y}$ conditioned by $\mathbf{z}$* is

$$I\big(\mathbf{x} : \mathbf{y} \mid \mathbf{z}\big) = \underset{z \sim \mu(\mathrm{z})}{\mathbb{E}} \left[I\big(\mathbf{x} : \mathbf{y} \mid \mathbf{z} = z\big) \right] = \underset{z \sim \mu(\mathrm{z})}{\mathbb{E}} \left[\mathsf{D}_{\mathsf{KL}}\big(\mu(\mathrm{x}, \mathrm{y} \mid z) \parallel \mu(\mathrm{x} \mid z) \times \mu(\mathrm{y} \mid z)\big) \right]$$

We present several facts about mutual information, the proofs can be found in the book of Cover and Thomas [7].

▶ **Fact 6** (Chain Rule). *For any random variables $\mathbf{x}_1, \mathbf{x}_2, \mathbf{y}$ and $\mathbf{z}$ holds that*

$$I\big(\mathbf{x}_1\mathbf{x}_2 : \mathbf{y} \mid \mathbf{z}\big) = I\big(\mathbf{x}_1 : \mathbf{y} \mid \mathbf{z}\big) + I\big(\mathbf{x}_2 : \mathbf{y} \mid \mathbf{z}, \mathbf{x}_1\big).$$

Since mutual information is never negative, we have the following corollary.

▶ **Corollary 7.** *For any random variables $\mathbf{x}, \mathbf{y}$ and $\mathbf{z}$ holds that $I\big(\mathbf{x} : \mathbf{y}\big) \leq I\big(\mathbf{x} : \mathbf{y}\,\mathbf{z}\big)$.*

The ℓ_1-distance between two distributions is defined as

$$\left\| \mu'(\mathrm{z}) - \mu(\mathrm{z}) \right\|_1 = \sum_z \left| \mu'(z) - \mu(z) \right|.$$

There is a relation between ℓ_1-distance and Kullback-Leibler divergence.

▶ **Fact 8** (Pinsker's Inequality). *For any two distributions $\mu'(\mathrm{z})$ and $\mu(\mathrm{z})$, we have*

$$\left\| \mu'(\mathrm{z}) - \mu(\mathrm{z}) \right\|_1 \leq \sqrt{2 \cdot \mathsf{D}_{\mathsf{KL}}\big(\mu'(\mathrm{z}) \parallel \mu(\mathrm{z})\big)}$$

3   The Proof of Theorem 5

Let $f : \{0,1\}^n \times \{0,1\}^n \to \{0,1,*\}$ be a partial function. Suppose there is a semi-adaptive random scheme D for the multiphase problem of f with error probability bounded by $\tilde{\varepsilon}$ such that $t_{\mathsf{u}} \cdot n \leq o\big(k/w\big)$. Let $\mu(\mathrm{x}, \mathrm{y}) = \mu_1(\mathrm{x}) \times \mu_2(\mathrm{y})$ be a product distribution over $\{0,1\}^n \times \{0,1\}^n$, such that $\mu(\mathrm{x}, \mathrm{y})$ is $\tilde{\alpha}$-balanced according to f. Let $b \in \{0,1\}$ and $g : \{0,1\}^n \times \{0,1\}^n \to \{0,1,*\}$ be a partial function which is λ-close to f under μ. We will prove there is a large almost b-monochromatic rectangle for g.

Let $\mathbf{x}_1, \dots, \mathbf{x}_k$ be independent random variables each of them distributed according to μ_1 and $\mathbf{y}$ be an independent random variable distributed according to μ_2. Let the random variable $\mathbf{z} \in \{0,1\}^m$ and the index $i \in [k]$ be given by Theorem 4 applied to the random variables $\mathbf{x}_1, \dots, \mathbf{x}_k, \mathbf{y}$ and the function f. For simplicity we denote $\mathbf{x} = \mathbf{x}_i$.

We will denote the joint distribution of $(\mathbf{x}_1, \dots, \mathbf{x}_k, \mathbf{y}, \mathbf{z})$ by $\mu(\mathrm{x}_1, \dots, \mathrm{x}_k, \mathrm{y}, \mathrm{z})$. Note that here the notation is consistent, in the sense that $\mu(x_i, y) = \mu_1(x_i) \times \mu_2(y)$ for all $i \in [k], x, y \in \{0,1\}^n$. We will then need to keep in mind that $\mu(\mathrm{z})$ is the z-marginal of the joint distribution of $(\mathbf{x}_1, \dots, \mathbf{x}_k, \mathbf{y}, \mathbf{z})$.

By $f(\mathbf{x}, \mathbf{y}) \neq^* \mathbf{z}_m$ we denote the event that the random variable $\mathbf{z}_m$ gives us the wrong answer on an input from the support of f, i.e. $f(\mathbf{x}, \mathbf{y}) \neq *$ and $f(\mathbf{x}, \mathbf{y}) \neq \mathbf{z}_m$ hold simultaneously. By Theorem 4 we know that $\Pr\big[f(\mathbf{x}, \mathbf{y}) \neq^* \mathbf{z}_m\big] \leq \tilde{\varepsilon}$. Since f and g are λ-close under μ, we have that μ is still balanced according to g and $g(\mathbf{x}, \mathbf{y}) \neq^* \mathbf{z}_m$ with small probability, as stated in the next observation.

▶ **Observation 9.** *Let $\alpha = \tilde{\alpha} - \lambda$ and $\varepsilon = \tilde{\varepsilon} + \lambda$. For the function g it holds that*
1. *The distribution $\mu(\mathrm{x}, \mathrm{y})$ is α-balanced according to g.*
2. $\Pr\big[g(\mathbf{x}, \mathbf{y}) \neq^* \mathbf{z}_m\big] \leq \varepsilon$.

Proof. Let $b' \in \{0,1\}$. We will bound $\mu\big(g^{-1}(b')\big)$.

$$
\begin{aligned}
\tilde{\alpha} \leq \Pr\big[f(\mathbf{x}, \mathbf{y}) = b'\big] &= \Pr\big[f(\mathbf{x}, \mathbf{y}) = b', f(\mathbf{x}, \mathbf{y}) = g(\mathbf{x}, \mathbf{y})\big] \\
&\quad + \Pr\big[f(\mathbf{x}, \mathbf{y}) = b', f(\mathbf{x}, \mathbf{y}) \neq g(\mathbf{x}, \mathbf{y})\big] \\
&\leq \Pr\big[g(\mathbf{x}, \mathbf{y}) = b'\big] + \lambda.
\end{aligned}
$$

Thus, by rearranging we get $\mu\big(g^{-1}(b')\big) \geq \tilde{\alpha} - \lambda = \alpha$. The proof of the second bound is similar:

$$
\begin{aligned}
\Pr\big[g(\mathbf{x}, \mathbf{y}) \neq^* \mathbf{z}_m\big] &= \Pr\big[f(\mathbf{x}, \mathbf{y}) \neq^* \mathbf{z}_m, f(\mathbf{x}, \mathbf{y}) = g(\mathbf{x}, \mathbf{y})\big] \\
&\quad + \Pr\big[g(\mathbf{x}, \mathbf{y}) \neq^* \mathbf{z}_m, f(\mathbf{x}, \mathbf{y}) \neq g(\mathbf{x}, \mathbf{y})\big] \leq \tilde{\varepsilon} + \lambda = \varepsilon. \qquad \blacktriangleleft
\end{aligned}
$$

Let c be the bound on $I\big(\mathbf{x} : \mathbf{y}\,\mathbf{z}\big)$ and $I\big(\mathbf{y} : \mathbf{z}\big)$ given by Theorem 4. Since $I\big(\mathbf{x} : \mathbf{z}\big) \leq I\big(\mathbf{x} : \mathbf{y}\,\mathbf{z}\big)$, we have $I\big(\mathbf{x} : \mathbf{z}\big), I\big(\mathbf{y} : \mathbf{z}\big) \leq t_{\mathsf{q}} \cdot w + o(t_{\mathsf{q}} \cdot w) = c$. We will prove that if we assume that $t_{\mathsf{u}} \cdot n < o\big(k/w\big)$ and we choose p large enough (p of Theorem 4) then we can find a

rectangle $R \subseteq X \times Y$ such that R is $O(\varepsilon/\alpha)$-error b-monochromatic for g and $\mu(R) \geq \frac{1}{2^{c'}}$ for $c' = O\left(\frac{t_{\mathsf{q}} \cdot w}{\alpha}\right)$. Thus, we have $\mathsf{mono}_{\mu}^{O(\varepsilon/\alpha)}(g) \geq 2^{-c'}$ and consequently

$$\mathsf{scb}_{\mu}^{O(\varepsilon/\alpha),\lambda}(f) \leq O\left(\frac{t_{\mathsf{q}} \cdot w}{\alpha}\right).$$

By rearranging, we get the bound of Theorem 5.

Let us sketch the proof of how we can find such a rectangle R. We will first fix the random variable $\mathbf{z}$ to z such that $\mathbf{x}$ and $\mathbf{y}$ are not very correlated conditioned on $\mathbf{z} = z$, i.e., the joint distribution $\mu(\mathrm{x}, \mathrm{y} \mid z)$ is very similar to the product distribution of the marginals $\mu(\mathrm{x} \mid z) \times \mu(\mathrm{y} \mid z)$. Moreover, we will pick z in such a way the probability of error $\Pr\left[g(\mathbf{x}, \mathbf{y}) \neq^* z_m \mid \mathbf{z} = z\right]$ is still small. Then, since $\mu(\mathrm{x}, \mathrm{y} \mid z)$ is close to $\mu(\mathrm{x} \mid z) \times \mu(\mathrm{y} \mid z)$, the probability of error under the latter distribution will be small as well, i.e., if $(\mathbf{x}', \mathbf{y}') \sim \mu(\mathrm{x} \mid z) \times \mu(\mathrm{y} \mid z)$, then $\Pr\left[g(\mathbf{x}', \mathbf{y}') \neq^* z_m\right]$ will also be small. Finally, we will find subsets $A \subseteq \mathsf{supp}\left(\mu(\mathrm{x} \mid z)\right), B \subseteq \mathsf{supp}\left(\mu(\mathrm{y} \mid z)\right)$ of large mass (under the original distributions μ_1 and μ_2), while keeping the probability of error on the rectangle $R = A \times B$ sufficiently small.

Let us then proceed to implement this plan. Let $\beta = \alpha - \varepsilon$. We will show that β is a lower bound for the probability that $\mathbf{z}_m$ is equal to b. Let γ be the bound on $I\left(\mathbf{x} : \mathbf{y} \mid \mathbf{z}\right)$ given by Theorem 4, i.e., $I\left(\mathbf{x}_i : \mathbf{y} \mid \mathbf{z}\right) \leq \gamma = O\left(\frac{t_{\mathsf{u}} \cdot n \cdot w}{p}\right)$.

▶ **Lemma 10.** *There exists $z \in Z$ such that*
1. $z_m = b$.
2. $I\left(\mathbf{x} : \mathbf{y} \mid \mathbf{z} = z\right) \leq \frac{5}{\beta} \cdot \gamma$.
3. $\mathsf{D}_{\mathsf{KL}}\left(\mu(\mathrm{x} \mid z) \parallel \mu(\mathrm{x})\right), \mathsf{D}_{\mathsf{KL}}\left(\mu(\mathrm{y} \mid z) \parallel \mu(\mathrm{y})\right) \leq \frac{5}{\beta} \cdot c$.
4. $\Pr\left[g(\mathbf{x}, \mathbf{y}) \neq^* z_m \mid \mathbf{z} = z\right] \leq \frac{5}{\beta} \cdot \varepsilon$.

Proof. Since μ is α-balanced according to g, we find that

$$\alpha \leq \Pr\left[g(\mathbf{x}, \mathbf{y}) = b\right]$$
$$= \Pr\left[g(\mathbf{x}, \mathbf{y}) = b, \mathbf{z}_m = b\right] + \Pr\left[g(\mathbf{x}, \mathbf{y}) = b, \mathbf{z}_m \neq b\right] \leq \Pr\left[\mathbf{z}_m = b\right] + \varepsilon.$$

Thus, by rearranging we get $\Pr\left[\mathbf{z}_m = b\right] \geq \alpha - \varepsilon = \beta$. By expanding the information $I\left(\mathbf{x} : \mathbf{y} \mid \mathbf{z}\right)$ we find

$$\gamma \geq I\left(\mathbf{x} : \mathbf{y} \mid \mathbf{z}\right) = \mathop{\mathbb{E}}_{z \sim \mu(\mathrm{z})}\left[I\left(\mathbf{x} : \mathbf{y} \mid \mathbf{z} = z\right)\right]$$

and by the Markov inequality we get that

$$\mathop{\Pr}_{z \sim \mu(\mathrm{z})}\left[I\left(\mathbf{x} : \mathbf{y} \mid \mathbf{z} = z\right) \geq \frac{5}{\beta} \cdot \gamma\right] \leq \frac{\beta}{5}.$$

Similarly, for the information $I\left(\mathbf{x} : \mathbf{z}\right)$:

$$c \geq I\left(\mathbf{x}\,\mathbf{y} : \mathbf{z}\right) \geq I\left(\mathbf{x} : \mathbf{z}\right) = \mathop{\mathbb{E}}_{z \sim \mu(\mathrm{z})}\left[\mathsf{D}_{\mathsf{KL}}\left(\mu(\mathrm{x} \mid z) \parallel \mu(\mathrm{x})\right)\right]$$

and so

$$\mathop{\Pr}_{z \sim \mu(\mathrm{z})}\left[\mathsf{D}_{\mathsf{KL}}\left(\mu(\mathrm{x} \mid z) \parallel \mu(\mathrm{x})\right) \geq \frac{5}{\beta} \cdot c\right] \leq \frac{\beta}{5}.$$

The bound for $I(\mathbf{y} : \mathbf{z})$ is analogous. Let $e_z = \Pr_\mu\big[g(\mathbf{x}, \mathbf{y}) \neq^* z_m | \mathbf{z} = z\big]$. Then,

$$\varepsilon \geq \Pr\big[g(\mathbf{x}, \mathbf{y}) \neq^* \mathbf{z}_m\big] = \sum_{z \in Z} \mu(z) \cdot e_z = \mathop{\mathbb{E}}_{z \sim \mu(\mathbf{z})} \big[e_z\big]$$

$$\Pr_{z \sim \mu(\mathbf{z})} \left[e_z \geq \frac{5}{\beta} \cdot \varepsilon\right] \leq \frac{\beta}{5}.$$

Thus, by a union bound we may infer the existence of the sought $z \in Z$. ◀

Let us now fix $z \in Z$ from the previous lemma. Let $\mu_z(\mathrm{x}, \mathrm{y}) = \mu(\mathrm{x}, \mathrm{y} \mid z)$ be the distribution $\mu(\mathrm{x}, \mathrm{y})$ conditioned on $\mathbf{z} = z$, and let $\mu'_z(\mathrm{x}, \mathrm{y}) = \mu(\mathrm{x} \mid z) \times \mu(\mathrm{y} \mid z)$ be the product of its marginals. Let S be the support of $\mu_z(\mathrm{x}, \mathrm{y})$, and let S_x and S_y be the supports of $\mu'_z(\mathrm{x})$ and $\mu'_z(\mathrm{y})$, respectively, i.e., S_x and S_y are the projections of S into X and Y.

Then Pinsker's inequality will give us that μ_z and μ'_z are very close. Let $\delta = \sqrt{\frac{10}{\beta} \cdot \gamma}$.

▶ **Lemma 11.** $\left\|\mu_z(\mathrm{x}, \mathrm{y}) - \mu'_z(\mathrm{x}, \mathrm{y})\right\|_1 \leq \delta$

Proof. Indeed, by Pinsker's inequality,

$$\left\|\mu_z(\mathrm{x}, \mathrm{y}) - \mu'_z(\mathrm{x}, \mathrm{y})\right\|_1 \leq \sqrt{2 \cdot \mathsf{D}_{\mathsf{KL}}\big(\mu_z(\mathrm{x}, \mathrm{y}) \,\|\, \mu'_z(\mathrm{x}, \mathrm{y})\big)}.$$

The right-hand side is $\sqrt{2 \cdot \mathsf{D}_{\mathsf{KL}}\big(\mu(\mathrm{x}, \mathrm{y} \mid z) \,\|\, \mu(\mathrm{x} \mid z) \times \mu(\mathrm{y} \mid z)\big)}$, which by definition of mutual information equals $\sqrt{2 \cdot I\big(\mathbf{x} : \mathbf{y} \mid \mathbf{z} = z\big)}$, and by Lemma 10 this is $\leq \sqrt{\frac{10}{\beta} \cdot \gamma} = \delta$. ◀

For the sake of reasoning, let $(\mathbf{x}', \mathbf{y}') \sim \mu'_z(\mathrm{x}, \mathrm{y})$ be random variables chosen according to to μ'_z. Let $\varepsilon' = \frac{5}{\beta} \cdot \varepsilon + \delta$. It then follows from Lemma 10 and Lemma 11 that:

▶ **Lemma 12.** $\Pr\big[g(\mathbf{x}', \mathbf{y}') \neq^* z_m\big] \leq \varepsilon'$.

Proof. We prove that

$$\left|\Pr\big[g(\mathbf{x}, \mathbf{y}) \neq^* z_m \mid \mathbf{z} = z\big] - \Pr\big[g(\mathbf{x}', \mathbf{y}') \neq^* z_m\big]\right| \leq \delta.$$

Since $\Pr\big[g(\mathbf{x}, \mathbf{y}) \neq^* z_m \mid \mathbf{z} = z\big] \leq \frac{5}{\beta} \cdot \varepsilon$ by Lemma 10, the lemma follows. Let

$$B = \big\{(x, y) \in S_\mathrm{x} \times S_\mathrm{y} : g(x, y) \neq z_m, g(x, y) \neq *\big\}.$$

$$\left|\Pr\big[g(\mathbf{x}, \mathbf{y}) \neq^* z_m \mid \mathbf{z} = z\big] - \Pr\big[g(\mathbf{x}', \mathbf{y}') \neq^* z_m\big]\right|$$
$$= \bigg|\sum_{(x,y) \in B} \mu_z(x, y) - \mu'_z(x, y)\bigg|$$
$$\leq \sum_{(x,y) \in B} \big|\mu_z(x, y) - \mu'_z(x, y)\big| \leq \delta \qquad \text{by the triangle inequality and Lemma 11}$$

◀

Let $c' = \frac{5}{\beta} \cdot c$. We will prove the ratio between $\mu'_z(\mathbf{x}')$ and $\mu(\mathbf{x}')$ is larger than $2^{O(c')}$ with only small probability (when $\mathbf{x}' \sim \mu'_z(\mathrm{x})$). The same holds for $\mu'_z(\mathbf{y}')$ and $\mu(\mathbf{y}')$.

▶ **Lemma 13.** $\Pr\left[\mu'_z(\mathbf{x}') \geq 2^{6c'} \cdot \mu(\mathbf{x}')\right], \Pr\left[\mu'_z(\mathbf{y}') \geq 2^{6c'} \cdot \mu(\mathbf{y}')\right] \leq \frac{1}{6}$.

Proof. We prove the lemma for $\mu'_z(\mathbf{x}')$, the proof for $\mu'_z(\mathbf{y}')$ is analogous. By Lemma 10 we know that $\mathsf{D}_{\mathsf{KL}}\big(\mu(\mathrm{x} \mid z) \parallel \mu(\mathrm{x})\big) = \mathsf{D}_{\mathsf{KL}}\big(\mu_z(\mathrm{x}) \parallel \mu(\mathrm{x})\big) = \mathsf{D}_{\mathsf{KL}}\big(\mu'_z(\mathrm{x}) \parallel \mu(\mathrm{x})\big) \leq c'$. We expand the Kullback-Leibler divergence:

$$c' \geq \mathsf{D}_{\mathsf{KL}}\big(\mu'_z(\mathrm{x}) \parallel \mu(\mathrm{x})\big) = \sum_{x \in S_{\mathrm{x}}} \mu'_z(x) \log \frac{\mu'_z(x)}{\mu(x)} = \mathbb{E}\left[\log \frac{\mu'_z(\mathbf{x}')}{\mu(\mathbf{x}')}\right],$$

and then use the Markov inequality:

$$\Pr\left[\mu'_z(\mathbf{x}') \geq 2^{6c'} \cdot \mu(\mathbf{x}')\right] = \Pr\left[\log \frac{\mu'_z(\mathbf{x}')}{\mu(\mathbf{x}')} \geq 6c'\right] \leq \frac{1}{6}. \qquad \blacktriangleleft$$

We now split S_{x} and S_{y} into buckets C^{x}_ℓ and C^{y}_ℓ (for $\ell \geq 1$), where the ℓ-th buckets are

$$C^{\mathrm{x}}_\ell = \left\{ x \in S_{\mathrm{x}} \ \Big| \ \frac{(\ell-1)}{2^{c'}} < \frac{\mu'_z(x)}{\mu(x)} \leq \frac{\ell}{2^{c'}} \right\},$$
$$C^{\mathrm{y}}_\ell = \left\{ y \in S_{\mathrm{y}} \ \Big| \ \frac{(\ell-1)}{2^{c'}} < \frac{\mu'_z(y)}{\mu(y)} \leq \frac{\ell}{2^{c'}} \right\}.$$

In a bucket C^{x}_ℓ there are elements of S_{x} such that their probability under $\mu'_z(\mathrm{x})$ is approximately $\frac{\ell}{2^{c'}}$-times bigger than their probability under $\mu(\mathrm{x})$. By Lemma 13, it holds that with high probability the elements $x \in S_{\mathrm{x}}, y \in S_{\mathrm{y}}$ are in the buckets $C^{\mathrm{x}}_{\ell_1}$ and $C^{\mathrm{y}}_{\ell_2}$ for $\ell_1, \ell_2 \leq 2^{7c'}$. Thus, if we find a bucket $C^{\mathrm{x}}_{\ell_1}$ for $\ell_1 \leq 2^{7c'}$ which has probability at least $\frac{1}{2^{O(c')}}$ under $\mu'_z(\mathrm{x})$, then it has also probability at least $\frac{1}{2^{O(c')}}$ under $\mu(\mathrm{x})$. The same holds also for buckets C^{y}_ℓ. In the next lemma we will show that there are buckets $C^{\mathrm{x}}_{\ell_1}$ and $C^{\mathrm{y}}_{\ell_2}$ of large probability under μ'_z such that the probability of error on $C^{\mathrm{x}}_{\ell_1} \times C^{\mathrm{y}}_{\ell_2}$ is still small.

▶ **Lemma 14.** *There exist buckets $C^{\mathrm{x}}_{\ell_1}$ and $C^{\mathrm{y}}_{\ell_2}$ such that*
1. $1 < \ell_1, \ell_2 \leq 2^{7c'}$.
2. $\Pr\big[\mathbf{x}' \in C^{\mathrm{x}}_{\ell_1}\big], \Pr\big[\mathbf{y}' \in C^{\mathrm{y}}_{\ell_2}\big] \geq \frac{1}{6 \cdot 2^{7c'}}$.
3. $\Pr\big[g(\mathbf{x}', \mathbf{y}') \neq^* z_m, (\mathbf{x}', \mathbf{y}') \in C^{\mathrm{x}}_{\ell_1} \times C^{\mathrm{y}}_{\ell_2}\big] \leq 6\varepsilon' \cdot \Pr\big[(\mathbf{x}', \mathbf{y}') \in C^{\mathrm{x}}_{\ell_1} \times C^{\mathrm{y}}_{\ell_2}\big]$.

Proof. We prove that ℓ_1, ℓ_2 exist via the probabilistic method. Let $\boldsymbol{\ell}_1$ and $\boldsymbol{\ell}_2$ be the buckets of $\mathbf{x}'$ and $\mathbf{y}'$, respectively. Thus $\Pr\big[\boldsymbol{\ell}_1 = \ell\big] = \Pr\big[\mathbf{x}' \in C^{\mathrm{x}}_\ell\big]$ and $\Pr\big[\boldsymbol{\ell}_2 = \ell\big] = \Pr\big[\mathbf{y}' \in C^{\mathrm{y}}_\ell\big]$.

Let $B_1, B_2 \subseteq L' = \{1, \ldots, 2^{7c'}\}$ be sets of indices of small probability, i.e., for $i \in \{1, 2\}$

$$B_i = \left\{ \ell \in L' \ \Big| \ \Pr[\boldsymbol{\ell}_i = \ell] \leq \frac{1}{6 \cdot 2^{7c'}} \right\}.$$

We will prove that with high probability we have $2^{7c'} \geq \boldsymbol{\ell}_1 > 1$ and $\boldsymbol{\ell}_1 \notin B_1$. The proof for $\boldsymbol{\ell}_2$ is analogous.

$$\Pr\big[\boldsymbol{\ell}_1 = 1\big] = \Pr\big[\mathbf{x}' \in C^{\mathrm{x}}_1\big] = \sum_{x \in C^{\mathrm{x}}_1} \mu'_z(x) \leq \frac{\sum_{x \in C^{\mathrm{x}}_1} \mu(x)}{2^{c'}} \leq \frac{1}{2^{c'}}$$

By Lemma 13, we get $\Pr\big[\boldsymbol{\ell}_1 > 2^{7c'}\big] = \Pr\big[\mu'_z(\mathbf{x}') \geq 2^{6c'} \cdot \mu(\mathbf{x}')\big] \leq \frac{1}{6}$. There is only small probability that $\boldsymbol{\ell}_1$ is in B_1.

$$\Pr\big[\boldsymbol{\ell}_1 \in B_1\big] = \sum_{\ell \in B_1} \Pr[\boldsymbol{\ell}_1 = \ell] \leq \frac{|L'|}{6 \cdot 2^{7c'}} = \frac{1}{6}$$

Thus, we have that $\boldsymbol{\ell}_i \in B_i$ or $\boldsymbol{\ell}_i = 1$ or $\boldsymbol{\ell}_i > 2^{7c'}$ with probability at most $\frac{2}{3} + \frac{2}{2^{c'}}$.

By Lemma 12, we have that $\Pr\big[g(\mathbf{x}', \mathbf{y}') \neq^* z_m\big] \leq \varepsilon'$. By expanding the probability and by Markov inequality we will now get the last inequality for $C_{\ell_1}^{\mathrm{x}}$ and $C_{\ell_2}^{\mathrm{y}}$. Let

$$e(\ell_1, \ell_2) = \Pr\big[g(\mathbf{x}', \mathbf{y}') \neq^* z_m \mid \mathbf{x}' \in C_{\ell_1}^{\mathrm{x}}, \mathbf{y}' \in C_{\ell_2}^{\mathrm{y}}\big].$$

We will prove there is ℓ_1 and ℓ_2 such that $e(\ell_1, \ell_2) \leq 6\varepsilon'$. This is equivalent to the third bound of the lemma. We have: $\varepsilon' \geq \Pr\big[g(\mathbf{x}', \mathbf{y}') \neq^* z_m\big] = \mathbb{E}\big[e(\boldsymbol{\ell}_1, \boldsymbol{\ell}_2)\big]$ and thus, by Markov, $\Pr\big[e(\boldsymbol{\ell}_1, \boldsymbol{\ell}_2) > 6\varepsilon'\big] \leq \frac{1}{6}$. By a union bound we conclude that there must exist $1 < \ell_1, \ell_2 \leq 2^{7c'}$ such that $\Pr[\boldsymbol{\ell}_1 = \ell_1], \Pr[\boldsymbol{\ell}_2 = \ell_2] \geq \frac{1}{6 \cdot 2^{7c'}}$ and $e(\ell_1, \ell_2) \leq 6\varepsilon'$.   ◀

As a corollary we will prove that the rectangle $C_{\ell_1}^{\mathrm{x}} \times C_{\ell_2}^{\mathrm{y}}$ (given by the previous lemma) is a good rectangle under the original distribution μ. We remark that the proof of the following corollary is the only place where we use the fact that $\mathbf{x}$ and $\mathbf{y}$ are independent.

▶ **Corollary 15.** *There exists a rectangle $R \subseteq S_{\mathrm{x}} \times S_{\mathrm{y}}$ such that*
1. $\Pr\big[(\mathbf{x}, \mathbf{y}) \in R\big] \geq \frac{1}{36 \cdot 2^{26c'}}$.
2. $\Pr\big[g(\mathbf{x}, \mathbf{y}) \neq^* z_m, (\mathbf{x}, \mathbf{y}) \in R\big] \leq 24\varepsilon' \cdot \Pr\big[(\mathbf{x}, \mathbf{y}) \in R\big]$.

Proof. Let $R = C_{\ell_1}^{\mathrm{x}} \times C_{\ell_2}^{\mathrm{y}}$ where $C_{\ell_1}^{\mathrm{x}}$ and $C_{\ell_2}^{\mathrm{y}}$ are buckets given by Lemma 14. By Lemma 14, we get

$$\frac{1}{6 \cdot 2^{7c'}} \leq \Pr\big[\mathbf{x}' \in C_{\ell_1}^{\mathrm{x}}\big] = \sum_{x \in C_{\ell_1}^{\mathrm{x}}} \mu_z'(x) \leq \sum_{x \in C_{\ell_1}^{\mathrm{x}}} \frac{\ell_1 \cdot \mu(x)}{2^{c'}} = \Pr\big[\mathbf{x} \in C_{\ell_1}^{\mathrm{x}}\big] \cdot \frac{\ell_1}{2^{c'}}.$$

By rearranging we get

$$\Pr\big[\mathbf{x} \in C_{\ell_1}^{\mathrm{x}}\big] \geq \frac{2^{c'}}{6\ell_1 \cdot 2^{7c'}} \geq \frac{1}{6 \cdot 2^{13c'}}$$

The bound for $\Pr\big[\mathbf{y} \in C_{\ell_2}^{\mathrm{y}}\big]$ is analogous, thus we have $\Pr\big[(\mathbf{x}, \mathbf{y}) \in R\big] \geq \frac{1}{36 \cdot 2^{26c'}}$. (Here and below, we crucially use the fact that $\mathbf{x}, \mathbf{y}$ are given by a product distribution.) Now we prove the second bound for R. Let $B = \big\{(x, y) \in R : g(x, y) \neq z_m, g(x, y) \neq *\big\}$.

$$\begin{aligned}
6\varepsilon' \cdot \Pr\big[(\mathbf{x}, \mathbf{y}) \in R\big] \cdot \frac{\ell_1 \ell_2}{2^{2c'}} &\geq 6\varepsilon' \cdot \Pr\big[(\mathbf{x}', \mathbf{y}') \in R\big] && \text{by definition of buckets} \\
&\geq \Pr\big[(\mathbf{x}', \mathbf{y}') \in B\big] && \text{by Lemma 14} \\
&\geq \Pr\big[(\mathbf{x}, \mathbf{y}) \in B\big] \cdot \frac{(\ell_1 - 1)(\ell_2 - 1)}{2^{2c'}} && \text{by definition of buckets}
\end{aligned}$$

Thus, by rearranging we get

$$\Pr\big[(\mathbf{x}, \mathbf{y}) \in B\big] \leq 6\varepsilon' \cdot \Pr\big[(\mathbf{x}, \mathbf{y}) \in R\big] \cdot \frac{\ell_1 \ell_2}{(\ell_1 - 1)(\ell_2 - 1)} \leq 24\varepsilon' \cdot \Pr\big[(\mathbf{x}, \mathbf{y}) \in R\big],$$

as $\frac{\ell_1 \ell_2}{(\ell_1 - 1)(\ell_2 - 1)} \leq 4$ for $\ell_1, \ell_2 > 1$ by Lemma 14.   ◀

Proof of Theorem 5. Suppose that $t_{\mathrm{u}} \cdot n \leq o\big(k/w\big)$. Let R be the rectangle given by Corollary 15. It holds that the rectangle R is $24\varepsilon'$-error b-monochromatic for g under μ. Therefore, for the function g holds that

$$\mathsf{mono}_\mu^{24\varepsilon'}(g) \geq \Pr\big[(\mathbf{x}, \mathbf{y}) \in R\big] \geq \frac{1}{36 \cdot 2^{26c'}}. \tag{1}$$

We need to argue that ε' is $O(\varepsilon/\alpha)$. By definition,

$$\varepsilon' = \frac{5}{\alpha - \varepsilon} \cdot \varepsilon + \delta.$$

We recall that

$$\delta = O\left(\sqrt{\frac{t_{\mathrm{u}} \cdot n \cdot w}{p}}\right) \leq \sqrt{\frac{o(k)}{p}}.$$

Thus, we can set p to be large enough so that δ is smaller than arbitrary constant and still $p \leq o(k)$. By the assumption we have $2\varepsilon < \alpha$. Thus, $\frac{\varepsilon}{\alpha-\varepsilon} \leq \frac{2\varepsilon}{\alpha}$ and we conclude that ε' is $O(\varepsilon/\alpha)$. Since $c' = O\left(\frac{t_{\mathrm{q}} \cdot w}{\alpha \cdot (1-\varepsilon)}\right) = O\left(\frac{t_{\mathrm{q}} \cdot w}{\alpha}\right)$, we get the result by rearranging Inequality (1). ◀

4 Applications

In this section we apply Theorem 5 to derive lower bounds for several explicit functions – Inner Product (IP), Disjointness (DISJ), Gap Orthogonality (ORT) and Gap Hamming Distance (GHD):

$$\mathsf{IP}(x,y) = \sum_{i \in n} x_i \cdot y_i \mod 2,$$

$$\mathsf{GHD}_n(x,y) = \begin{cases} 1 & \text{if } \Delta_H(x,y) \geq \frac{n}{2} + \sqrt{n}, \\ 0 & \text{if } \Delta_H(x,y) \leq \frac{n}{2} - \sqrt{n}. \end{cases}$$

The function Δ_H is the Hamming Distance of two strings, i.e., $\Delta_H(x,y)$ is a number of indices $i \in [n]$ such that $x_i \neq y_i$. For $\mathsf{IP}_\mathbb{R}(x,y) = \sum_{i \in [n]} (-1)^{x_i + y_i}$ we define

$$\mathsf{ORT}_{n,d}(x,y) = \begin{cases} 1 & \text{if } \left|\mathsf{IP}_\mathbb{R}(x,y)\right| \geq 2d \cdot \sqrt{n} \\ 0 & \text{if } \left|\mathsf{IP}_\mathbb{R}(x,y)\right| \leq d \cdot \sqrt{n}. \end{cases}$$

The standard value for d is 1, thus we denote $\mathsf{ORT}_n = \mathsf{ORT}_{n,1}$. Note that $\Delta_H(x,y) = \frac{n - \mathsf{IP}_\mathbb{R}(x,y)}{2}$ and $\mathsf{IP}_\mathbb{R}(x,y)$ is the Inner Product of x', y' over $\mathbb{R}$ where x' and y' arise from x and y by replacing 0 by 1 and 1 by -1. We present previous results with bounds for measures of interest under hard distributions.

▶ **Theorem 16** ([12]). *Let μ_1 be a uniform distribution on $\{0,1\}^n \times \{0,1\}^n$. Then,*

$$\mathsf{disc}_{\mu_1}(\mathsf{IP}) \leq \frac{1}{2^{n/2}}.$$

▶ **Theorem 17** (Babai et al. [2]). *Let $\rho < 1/100$ and μ_2 be a a uniform distribution over $S \times S$, where S consists of n-bit strings containing exactly $\sqrt{n}$ 1's. Then,*

$$\mathsf{mono}_{\mu_2}^\rho(\mathsf{DISJ}) \leq \frac{1}{2^{\Omega(\sqrt{n})}}.$$

Sherstov [20] provided a lower bound of communication complexity of GHD by lower bound of corruption bound of $\mathsf{ORT}_{n,\frac{1}{8}}$ following by reduction to GHD.

▶ **Theorem 18** (Sherstov [20]). *Let $\rho > 0$ be sufficiently small and μ_3 be a uniform distribution over $\{0,1\}^n \times \{0,1\}^n$. Then,*

$$\mathsf{cb}_{\mu_3}^\rho(\mathsf{ORT}_{n,\frac{1}{8}}) \geq \rho \cdot n.$$

By this theorem and Theorem 5 we get a lower bound for data structures for $\mathsf{ORT}_{n,\frac{1}{8}}$. By reductions used by Sherstov [20] we also get a lower bounds for ORT and GHD.

$$\mathsf{ORT}_{n,\frac{1}{8}}(x,y) = \mathsf{ORT}_{64n}\big(x^{64}, y^{64}\big)$$

$$\mathsf{ORT}_n(x,y) = \mathsf{GHD}_{10n+15\sqrt{n}}\big(x^{10}1^{15\sqrt{n}}, y^{10}0^{15\sqrt{n}}\big)$$
$$\wedge \neg\mathsf{GHD}_{10n+15\sqrt{n}}\big(x^{10}0^{15\sqrt{n}}, y^{10}0^{15\sqrt{n}}\big)$$

Where s^i denote i copies of s concatenated together. Let D be a semi-adaptive random scheme for the multiphase problem of the presented functions with sufficiently small error probability. By the theorems presented in this section and by Theorem 5, we can derive the following lower bounds for $t_q \cdot w$, assuming that $t_u \cdot n \leq o\big(k/w\big)$.

Function f	Ballancedness of the hard distribution	Lower bound of $t_q \cdot w$
IP	$\frac{1}{2}$	$\Omega(n)$
DISJ	$\sim \frac{1}{e}$	$\Omega(\sqrt{n})$
ORT_n	$\Theta(1)$	$\Omega(n)$
GHD_n	N/A (lower-bound is via reduction)	$\Omega(n)$

References

1   Amir Abboud and Virginia Vassilevska Williams. Popular conjectures imply strong lower bounds for dynamic problems. In *Proceedings of the 55th FOCS*, pages 434–443, 2014.

2   Laszlo Babai, Peter Frankl, and Janos Simon. Complexity classes in communication complexity theory. In *Proceedings of the 27th FOCS*, page 337–347, 1986.

3   Ziv Bar-Yossef, T. S. Jayram, Ravi Kumar, and D. Sivakumar. An information statistics approach to data stream and communication complexity. In *Proceedings of the 43rd FOCS*, page 209–218, 2002.

4   Paul Beame, Toniann Pitassi, Nathan Segerlind, and Avi Wigderson. A strong direct product theorem for corruption and the multiparty communication complexity of disjointness. *Computational Complexity*, 15(4):391–432, 2006.

5   Amit Chakrabarti, Ranganath Kondapally, and Zhenghui Wang. Information complexity versus corruption and applications to orthogonality and gap-hamming. In *Proceedings of the 16th RANDOM*, pages 483–494. Springer, 2012.

6   Arkadev Chattopadhyay, Jeff Edmonds, Faith Ellen, and Toniann Pitassi. A Little Advice Can Be Very Helpful. In *Proceedings of the 23rd SODA*, pages 615–625, 2012.

7   Thomas M. Cover and Joy A. Thomas. *Elements of Information Theory (Wiley Series in Telecommunications and Signal Processing)*. Wiley-Interscience, 2006.

8   Michael Fredman and Michael Saks. The cell probe complexity of dynamic data structures. In *Proceedings of the 21st STOC*, pages 345–354, 1989.

9   Rahul Jain and Hartmut Klauck. The partition bound for classical communication complexity and query complexity. In *Proceedings of the 25th CCC*, page 247–258, 2010.

10   Bala Kalyanasundaram and Georg Schintger. The Probabilistic Communication Complexity of Set Intersection. *SIAM Journal of Discrete Mathematics*, 5(4):545–557, 1992.

11   Young Kun Ko and Omri Weinstein. An Adaptive Step Toward the Multiphase Conjecture, 2019. `arXiv:1910.13543`.

12   Eyal Kushilevitz and Noam Nisan. *Communication Complexity*. Cambridge University Press, 1996.

13   Kasper Green Larsen. The cell probe complexity of dynamic range counting. In *Proceedings of the 44th STOC*, pages 85–94, 2012.

14   Kasper Green Larsen, Omri Weinstein, and Huacheng Yu. Crossing the logarithmic barrier for dynamic boolean data structure lower bounds. *SIAM Journal on Computing*, 2020.

15   Mihai Păatraşcu and Erik D Demaine. Tight bounds for the partial-sums problem. In *Proceedings of the 15th SODA*, pages 20–29, 2004.

16   Mihai Patrascu. Towards polynomial lower bounds for dynamic problems. In *Proceedings of the 42nd STOC*, pages 603–610, 2010.

17   Mihai Patrascu and Erik D Demaine. Logarithmic lower bounds in the cell-probe model. *SIAM Journal on Computing*, 35(4):932–963, 2006.

18   Mihai Pătraşcu. Towards Polynomial Lower Bounds for Dynamic Problems. In *Proceedings of the 42nd STOC*, pages 603–610, 2010.

19   Alexander A. Razborov. On the distributional complexity of disjointness. *Theoretical Computer Science*, 106:385–390, 1992.

20   Alexander A Sherstov. The communication complexity of gap hamming distance. *Theory of Computing*, 8(1):197–208, 2012.

21   Andrew Chi-Chih Yao. Some complexity questions related to distributive computing (preliminary report). In *Proceedings of the 11h STOC*, pages 209–213, 1979.