Beating the Direct Sum Theorem in Communication Complexity with Implications for Sketching

Marco Molinaro*
Carnegie Mellon University
molinaro@cmu.edu

David P. Woodruff
IBM Almaden
dpwoodru@us.ibm.com

Grigory Yaroslavtsev †
Penn State University
grigory@cse.psu.edu

Abstract

A direct sum theorem for two parties and a function f states that the communication cost of solving k copies of f simultaneously with error probability $1/3$ is at least $k \cdot R_{1/3}(f)$, where $R_{1/3}(f)$ is the communication required to solve a single copy of f with error probability $1/3$. We improve this for a natural family of functions f, showing that the 1-way communication required to solve k copies of f simultaneously with probability $2/3$ is $\Omega(k \cdot R_{1/k}(f))$. Since $R_{1/k}(f)$ may be as large as $\Omega(R_{1/3}(f) \cdot \log k)$, we asymptotically beat the direct sum bound for such functions, showing that the trivial upper bound of solving each of the k copies of f with probability $1 - O(1/k)$ and taking a union bound is optimal! In order to achieve this, our direct sum involves a novel measure of information cost which allows a protocol to abort with constant probability, and otherwise must be correct with very high probability. Moreover, for the functions considered, we show strong lower bounds on the communication cost of protocols with these relaxed guarantees; indeed, our lower bounds match those for protocols that are not allowed to abort.

In the distributed and streaming models, where one wants to be correct not only on a single query, but simultaneously on a sequence of n queries, we obtain optimal lower bounds on the communication or space complexity. Lower bounds obtained from our direct sum result show that a number of techniques in the sketching literature are optimal, including the following:

- (JL transform) Lower bound of $\Omega(\frac{1}{\epsilon^2} \log \frac{n}{\delta})$ on the dimension of (oblivious) Johnson-Lindenstrauss transforms.
- (ℓ_p-estimation) Lower bound for the size of encodings of n vectors in $[\pm M]^d$ that allow ℓ_1 or ℓ_2-estimation of $\Omega(n\epsilon^{-2} \log \frac{n}{\delta}(\log d + \log M))$.
- (Matrix sketching) Lower bound of $\Omega(\frac{1}{\epsilon^2} \log \frac{n}{\delta})$ on the dimension of a matrix sketch S satisfying the entrywise guarantee $|(ASS^TB)_{i,j} - (AB)_{i,j}| \leq \epsilon \|A_i\|_2 \|B^j\|_2$.
- (Database joins) Lower bound of $\Omega(n \frac{1}{\epsilon^2} \log \frac{n}{\delta} \log M)$ for sketching frequency vectors of n tables in a database, each with M records, in order to allow join size estimation.

*Supported by an IBM PhD fellowship. Work done while visiting IBM Almaden Research Center.

†Work done during an internship at IBM Almaden Research Center. Supported by College of Engineering Fellowship at Pennsylvania State University.

1 Introduction

We study the two-party communication complexity of solving multiple instances of a function $f(x, y)$. In this setting, Alice has $x_1, \ldots, x_k$, while Bob has $y_1, \ldots, y_k$, and they would like to communicate as few bits as possible in order to compute the list $(f(x_1, y_1), \ldots, f(x_k, y_k))$ with probability at least $2/3$. We call this problem $f^k((x_1, \ldots, x_k), (y_1, \ldots, y_k))$. A natural protocol for f^k would be for Alice and Bob to run an independent protocol for each $i \in [k]$ to compute $f(x_i, y_i)$ with probability at least $1 - 1/(3k)$. Then, by a union bound, the entire list is computed correctly with probability at least $2/3$. If we let $R_\delta(f)$ denote the minimal communication cost of a randomized protocol for computing f with probability at least $1 - \delta$, this gives us the upper bound $R_{1/3}(f^k) = O(kR_{1/(3k)}(f))$. A natural question is whether this is optimal.

A direct sum theorem in communication complexity states that solving k copies of f with probability at least $2/3$ requires at least k times as much communication as solving a single copy with probability at least $2/3$, that is, $R_{1/3}(f^k) = \Omega(kR_{1/3}(f))$. The direct sum problem is the focus of much work [10, 35, 21, 16, 8, 26, 9]. The direct sum theorem is known to hold for a number of specific functions, though it is not true for randomized private coin communication in general, as demonstrated by the Equality function. For this function, Alice and Bob have $x \in \{0, 1\}^k$ and $y \in \{0, 1\}^k$ respectively, and $f(x, y) = 1$ if $x = y$, otherwise $f(x, y) = 0$. In this case, $R_{1/3}(f^k) = \Theta(k)$ [15], yet $R_{1/3}(f) = \Theta(\log k)$ [28].

One of the most general known results about direct sums for communication is the following. Letting $D^{\mu}_{1/3}(f^k)$ denote the distributional complexity of f^k, that is, the minimal cost of a deterministic protocol for computing f^k which errs on at most a $1/3$ fraction of inputs, weighted according to distribution μ, then $D^{\mu,r}_{1/3-\epsilon}(f^k) = \Omega(k(D^{\mu}_{1/3}(f) - r\log(1/\epsilon) - O(\sqrt{D^{\mu}_{1/3}(f)r})))$, where $D^{\mu,r}_{1/3-\epsilon}(f^k)$ denotes the minimal cost of a deterministic protocol for computing f^k

with error probability at most $1/3 - \epsilon$ according to μ, and which uses at most r rounds of communication [9]. Moreover, this holds even if the protocol is only required to individually solve each of the n copies of f with probability at least $1/3 - \epsilon$, rather than to simultaneously solve all copies. Other related work includes direct product theorems, which state that any protocol which uses $o(kR_{1/3}(f))$ communication has success probability at most $\exp(-k)$ in solving f^k. Direct product theorems are known to hold for several functions, such as disjointness [26] and bounded round public-coin randomized communication [19, 20]. For more discussion on the latter two works, see below.

The starting point of our work is the following observation: even if a direct sum or direct product theorem were to hold for a function f, this is not enough to obtain optimal communication bounds, since one would only be able to conclude that:

$$\Omega(kR_{1/3}(f)) = R_{1/3}(f^k) = O(kR_{1/(3k)}(f)).$$

The ratio of the upper and lower bounds is $O(R_{1/(3k)}(f)/R_{1/3}(f))$, which can be as large as $\Theta(\log k)$. This $\Theta(\log k)$ factor is important in applications, which we describe below. Generic direct sum (or direct product) theorems are not suitable for addressing this gap, since such theorems do not take into account whether or not f becomes harder with smaller error probability, i.e., whether $R_\delta(f) \gg R_{\delta'}(f)$ for $\delta \ll \delta'$. For many functions, $R_{1/3}(f) = \Theta(R_0(f))$, e.g., for the disjointness problem, and there is nothing to prove in this case. Still for other functions, such as equality on n-bit strings, one can show that $R_\delta(f) = \Theta(\log 1/\delta + \log n)$, and so $R_\delta(f) \gg R_{\delta'}(f)$ for $\delta \ll \delta' \ll 1/n$.

Our Results: Our main theorem is a strengthening of the direct sum theorem for two-party randomized communication complexity to address this gap. We note that although our applications are for 1-way protocols, our main theorem holds for general 2-way communication. For that, we introduce the notion of deterministic protocols Π with the following "abortion" property.

Definition 1.1. (Protocols with abortion) *Consider a communication problem given by $f : \mathcal{X} \times \mathcal{Y} \to \mathcal{Z}$ and a probability distribution μ over $\mathcal{X} \times \mathcal{Y}$. We say that a deterministic protocol Π_D (β, δ)-computes f with respect to μ if it satisfies the following (where $(X, Y) \sim \mu$):*

1. *(Abortion probability)* $\Pr[\Pi_D(X, Y) = \text{`abort'}] \leq \beta$
2. *(Success probability)* $\Pr[\Pi_D(X, Y) \neq f(X, Y) \mid \Pi_D(X, Y) \neq \text{`abort'}] \leq \delta$.

We can view randomized protocols as distributions over deterministic protocols (both for private-coin and public-coin protocols). We say that a randomized protocol Π (α, β, δ)-computes f with respect to μ if $\Pr_{\Pi_D \sim \Pi}[\Pi_D\ (\beta, \delta)\text{-computes } f] \geq 1 - \alpha$. The probability is taken over all randomness of the parties.

One should think of $\beta \gg \delta$. Notice that a protocol that (β, δ)-computes f is more powerful than a deterministic protocol which errs with probability at most $\approx \beta$ on the input distribution μ, since it "knows when it is wrong". On the other hand, it is less powerful than a deterministic protocol which errs with probability at most δ on distribution μ.

In our proofs we use the information complexity framework (see Section 2 for an introduction and basic definitions), developed and used for many important communication complexity problems [6, 8, 10, 16]. Our definitions are most closely related to those in [7] (see also [8]). Let λ be a distribution on $\mathcal{X} \times \mathcal{Y} \times \mathcal{D}$ with marginals μ on $\mathcal{X} \times \mathcal{Y}$ and ν on $\mathcal{D}$. Let $(X, Y, D) \sim \lambda$ and suppose for any value of $d \in \mathcal{D}$ that X and Y are independent conditioned on $D = d$. The conditional information cost of Π under λ is defined as $\mathrm{I}(\Pi(X, Y); X, Y | D)$, where $(X, Y, D) \sim \lambda$. Let $\mathrm{IC}_{\mu,\alpha,\beta,\delta}(f|\nu)$ denote the minimum, over all protocols Π that (α, β, δ)-compute f, of $\mathrm{I}(X, Y; \Pi \mid D)$, where with some abuse of notation, Π is also used to denote the transcript of the protocol (that is, the set of messages exchanged together with the output). We also use the notation $\mathrm{IC}_{\mu,\delta}(f|\nu)$ to denote the minimum of $\mathrm{I}(X, Y; \Pi \mid D)$ over all randomized protocols Π, which $(0, 0, \delta)$-compute f (that is, which err with probability at most δ on every input, where the probability is over the random coins of Π). Notice that $\mathrm{I}(X, Y; \Pi \mid D) \leq \mathrm{H}(\Pi) \leq |\Pi|$ (where $|\Pi|$ is the maximum number of bits transmitted by Π over all inputs and choices of randomness), and so $\mathrm{IC}_{\mu,\delta}(f|\nu)$ is a lower bound on $R_\delta(f)$. As we will also be interested in 1-way protocols, we use $\mathrm{IC}^{\rightarrow}_{\mu,\alpha,\beta,\delta}(f|\nu)$ and $\mathrm{IC}^{\rightarrow}_{\mu,\delta}(f|\nu)$ to denote the above notions, where the minimum is taken over only 1-way protocols Π.

The following is our main theorem.

Theorem 1.1. *(Informal) For all $\delta \leq 1/3$,*

$$\mathrm{IC}_{\mu^k,\delta}(f^k|\nu^k) \geq \Omega(k)\, \mathrm{IC}_{\mu,\frac{1}{20},\frac{1}{10},\frac{\delta}{k}}(f|\nu),$$

and also $\mathrm{IC}^{\rightarrow}_{\mu^k,\delta}(f^k|\nu^k) \geq \Omega(k)\, \mathrm{IC}^{\rightarrow}_{\mu,\frac{1}{20},\frac{1}{10},\frac{\delta}{k}}(f|\nu)$.

As an example usage of our theorem, we can apply it to the Equality function f on k-bit strings. Namely, we

are able to show for certain distributions μ and ν that $\mathrm{IC}^{\rightarrow}_{\mu,1/20,1/10,1/k}(f|\nu) = \Omega(\log k)$, matching the lower bound for protocols that are not allowed to abort. Our theorem therefore implies that $R^{\rightarrow}_{1/3}(f^k) = \Omega(k \log k)$, that is, the randomized 1-way complexity of solving k copies of Equality simultaneously is $\Omega(k \log k)$. This is matched by a trivial $O(k \log k)$ upper bound which solves Equality independently on each instance with probability $1 - O(1/k)$. To the best of our knowledge, no such result was known in the literature.

More importantly, we are able to apply our theorem to the augmented indexing problem on large domains with low error [23], denoted by $\mathrm{Ind}^a(k, N)$. In this problem, Alice has a list $x_1, x_2, \ldots, x_N$, each item belonging to the set $[k] = \{1, 2, \ldots, k\}$, while Bob has input $j \in [N], x_1, x_2, \ldots, x_{j-1}$, and $y \in [k]$. The function f evaluates to 1 if $x_j = y$, and otherwise it evaluates to 0. We consider 1-way protocols Π, where the message is sent from Alice to Bob. It is known that $R^{\rightarrow}_{1/k}(f) = \Theta(N \log k)$ [23]. We are able to show that for certain distributions μ and ν, we in fact have $\mathrm{IC}^{\rightarrow}_{\mu,1/20,1/10,1/k}(f|\nu) = \Omega(N \log k)$. Plugging this in to our main theorem, we obtain that $R^{\rightarrow}_{1/3}(f^k) = \Omega(kN \log k)$. Previously, it was only known that $R^{\rightarrow}_{1/3}(f^k) = \Omega(kN)$, which can be shown by using that $\mathrm{IC}^{\rightarrow}_{\mu,1/3}(f|\nu) = \Omega(N)$ [6], and applying a standard direct sum argument [7].

Our lower bound is optimal in light of a trivial upper bound in which Alice sends her entire input to Bob. The augmented indexing problem is known to have a long list of applications to data streams and sketching, some of the most recent applications appearing in [25, 24, 23], and so our lower bound on solving k copies of this problem applies to solving multiple copies of these problems, as described below.

Applications: [1] Our first application is to the sketching complexity [18, 30] of n-point Johnson-Lindenstrauss transforms. Here one wants to design a distribution over $k \times d$ matrices S so that given any n points $\mathbf{p}^1, \mathbf{p}^2, \ldots, \mathbf{p}^n$ in $\mathbb{R}^d$, with probability at least $1 - \delta$, for all i and j, $\|S\mathbf{p}^i - S\mathbf{p}^j\|_2 = (1 \pm \epsilon)\|\mathbf{p}^i - \mathbf{p}^j\|_2$. See Definition 1 of [34], where this is referred to as $\mathrm{JLT}(\epsilon, \delta, n)$. Alon [3] has shown that this problem requires $k = \Omega(\frac{1}{\epsilon^2}\frac{1}{\log 1/\epsilon}\log\frac{n}{\delta})$ dimensions. Jayram and the second author show that for a constant number of points ($n = O(1)$), $k = \Omega(\epsilon^{-2}\log\frac{1}{\delta})$, which is also achievable by applying known Johnson-Lindenstrauss transforms [23]. We note that such work does not imply that for general n there is a lower bound of $k = \Omega(\epsilon^{-2}\log\frac{n}{\delta})$. Indeed, for all we knew, it could have been that $k = O(\frac{1}{\epsilon^2}\frac{1}{\log 1/\epsilon}\log\frac{n}{\delta})$, since there may be a better strategy than setting the failure probability to $O(\delta/n^2)$ and taking a union bound over the $\binom{n}{2}$ pairs of points. Our main theorem rules out this possibility, showing that $k = \Omega(\epsilon^{-2}\log\frac{n}{\delta})$ (Theorem 4.3). In fact, the main theorem shows that even if S is allowed to depend on the first $n/2$ points in an arbitrary way, the same lower bound still holds. In addition, we show that any encoding $\phi(\mathbf{p}^1), \ldots, \phi(\mathbf{p}^n)$ that allows pairwise ℓ_p-distance estimation for $p \in \{1, 2\}$ requires bit size $\Omega(n\epsilon^{-2}\log\frac{n}{\delta}(\log d + \log M))$, where M is the largest entry in absolute value of the vectors $\mathbf{p}^i$'s (Theorem 4.1); this is again optimal and achieved by known dimension reduction techniques [17].

A related problem is that of sketching matrix product, initiated in [34]. Here one wants to design a distribution over $n \times k$ matrices S, so that given $n \times n$ matrices A and B, one can "sketch" the matrices to obtain AS and S^TB such that the matrix $C = ASS^TB$ approximates the product AB for some measure of error. Ideally, we would like k to be as small as possible, and obtain an entrywise error guarantee, namely, for all $i, j \in [n]$, we would like $|(AB)_{i,j} - C_{i,j}| \leq \epsilon\|A_i\|_2\|B^j\|_2$, where A_i denotes the i-th row of A and B^j the j-th column of B. This notion of error has been used in several works, see the end of Section 1.1 of [33] for a discussion. In particular, Sárlos [34] achieves this error guarantee with $k = O(\epsilon^{-2}\log\frac{n}{\delta})$, for success probability $1 - \delta$. Later, Clarkson and the second author [12] were able to achieve $k = O(\epsilon^{-2}\log\frac{1}{\delta})$ with the weaker guarantee that $\|AB - C\|_F \leq \epsilon\|A\|_F\|B\|_F$. A natural question left open is whether $k = O(\epsilon^{-2}\log\frac{1}{\delta})$ is possible with the entrywise error guarantee. Using our main theorem, we show that this is not possible, namely that $k = \Omega(\epsilon^{-2}\log\frac{n}{\delta})$ is required in order to achieve the entrywise error guarantee (Theorem 4.5). We therefore separate the complexity of the two problems. Moreover, we show that sketches that satisfy the weaker guarantee that there is a procedure f outputting a matrix such that $|f(AS, B)_{i,j} - (AB)_{i,j}| \leq \epsilon\|A_i\|\|B^j\|$ for all matrices $A, B \in [\pm M]^{n\times n}$, then the bit size of AS is at least $\Omega(n\frac{1}{\epsilon^2}\log\frac{n}{\delta}(\log n + \log M))$, which is achieved in [34].

The final application we discuss is to multiple aggregation queries. While much of the data stream literature involves sequentially processing a large database to answer a single query, such as the number of distinct elements in a certain column or the join size of two tables, what one usually wants is to perform a sequence of such queries to different parts of the database. This issue was raised in [4], where the authors consider the setting of a relation which holds multiple tables, each of which has multiple columns of attributes. The authors

[1] All logarithms are base 2, unless otherwise specified. To simplify the notation, we assume throughout that quantities like $1/\epsilon^2$ and $1/\delta$ are always integral.

consider the problem of sketching each of the columns of attributes in each of the different tables, so that the storage size of the database can be reduced, yet at any later time, a user can ask for the join size along an attribute shared by two tables. They show that if the number of tables and attributes is poly(n), then each column can be compressed to $O(\epsilon^{-2}\log\frac{n}{\delta}\log M)$ bits, where M is an upper bound on the number of records in each table. It was left open whether or not this is optimal. Using our main theorem, we can show that $\Omega(\epsilon^{-2}\log\frac{n}{\delta}\log M)$ bits are in fact necessary (Theorem 4.6).

All of our results concerning linear sketches also hold for the *turnstile model* of data streaming [32, 5] and for more general data structures which, given their current state, and an update to the underlying vector, can produce a new state. Such data structures are sometimes referred to as *mergeable summaries* [2].

Our Techniques: Our starting point is the direct sum framework of [7]. There the authors show that for $(X,Y,D)\sim\lambda$ with $(X,Y)\sim\mu$ and $D\sim\nu$, if X and Y are independent conditioned on $D=d$ for any $d\in\mathcal{D}$, then $\mathrm{IC}_{\mu^k,\delta}(f^k|\nu^n)=\Omega(k\,\mathrm{IC}_{\mu,\delta}(f|\nu))$. To show this, they start with any randomized private coin protocol Π for f^k, with inputs $(X_1,Y_1),\ldots,(X_k,Y_k)$ and values $D_1,\ldots,D_k$, so that the X_i and Y_i are independent conditioned on D_i. They study the mutual information between the transcript and the inputs, conditioned on $D_1,\ldots,D_k$, namely $I(\mathbf{X},\mathbf{Y};\Pi|\mathbf{D})=H(\mathbf{X},\mathbf{Y}|\mathbf{D})-H(\mathbf{X},\mathbf{Y}|\Pi,\mathbf{D})$, where $\mathbf{X}=(X_1,\ldots,X_n)$, and $\mathbf{Y}$ and $\mathbf{D}$ are defined similarly. By the chain rule for mutual information,

$$\mathrm{I}(\mathbf{X},\mathbf{Y};\Pi\mid\mathbf{D})=\sum_{i=1}^{k}\mathrm{I}(X_i,Y_i;\Pi\mid\mathbf{D},\mathbf{X}_{<i},\mathbf{Y}_{<i}),$$

where $\mathbf{X}_{<i}$ and $\mathbf{Y}_{<i}$ denote the first $i-1$ coordinates of $\mathbf{X}$ and $\mathbf{Y}$, respectively. For each summand, we further have

$$\mathrm{I}(X_i,Y_i;\Pi\mid\mathbf{D},\mathbf{X}_{<i},\mathbf{Y}_{<i})=\sum_{\substack{\mathbf{x}_{<i},\mathbf{y}_{<i},\\ \mathbf{d}_{-i}}}\mathrm{I}(X_i,Y_i;\Pi\mid D_i,\mathbf{X}_{<i}=\mathbf{x}_{<i},\mathbf{Y}_{<i}=\mathbf{y}_{<i},\mathbf{D}_{-i}=\mathbf{d}_{-i})\cdot\Pr[\mathbf{X}_{<i}=\mathbf{x}_{<i},\mathbf{Y}_{<i}=\mathbf{y}_{<i},\mathbf{D}_{-i}=\mathbf{d}_{-i}],$$

where $\mathbf{D}_{-i}$ denotes $\mathbf{D}$ with its i-th coordinate removed. The next step is the embedding step, which argues that for any choice of $\mathbf{x}_{<i},\mathbf{y}_{<i}$, and $\mathbf{d}_{-i}$, $\mathrm{I}(X_i,Y_i;\Pi\mid D_i,\mathbf{X}_{<i}=\mathbf{x}_{<i},\mathbf{Y}_{<i}=\mathbf{y}_{<i},\mathbf{D}_{-i}=\mathbf{d}_{-i})$ is at least $\mathrm{IC}_{\mu,\delta}(f|\nu)$. This step works by building a protocol Π' for solving f by hardwiring the values $\mathbf{x}_{<i},\mathbf{y}_{<i}$ and $\mathbf{d}_{-i}$ into Π'. Then given inputs (A,B) to Π' distributed according to μ, the parties set $X_i=A$, $Y_i=B$, and generate $\mathbf{X}_{>i}$ and $\mathbf{Y}_{>i}$ using private randomness without any communication. This is possible given the conditioning $\mathbf{D}_{-i}=\mathbf{d}_{-i}$. A randomized protocol Π for f^k, for every input, solves f in each coordinate simultaneously with probability at least $1-\delta$, and therefore Π' is a correct protocol for f with probability at least $1-\delta$. Moreover, this simulation guarantees that $\mathrm{I}(X_i,Y_i;\Pi\mid D_i,\mathbf{X}_{<i}=\mathbf{x}_{<i},\mathbf{Y}_{<i}=\mathbf{y}_{<i},\mathbf{D}_{-i}=\mathbf{d}_{-i})=\mathrm{I}(A,B;\Pi'\mid D_i)\geq\mathrm{IC}_{\mu,\delta}(f|\nu)$.

Our main idea is to change the embedding step as follows. Observe that

$$1-\delta\leq\Pr(\Pi(\mathbf{X},\mathbf{Y})=f^k(\mathbf{X},\mathbf{Y}))=\prod_{i=1}^{k}\Pr(\Pi_i(\mathbf{X},\mathbf{Y})=f_i^k(\mathbf{X},\mathbf{Y})\mid\Pi_{<i}(\mathbf{X},\mathbf{Y})=f_{<i}^k(\mathbf{X},\mathbf{Y})),$$

where $\Pi_i(\mathbf{X},\mathbf{Y})$ denotes the i-th coordinate of the output of Π, and $f_i^k(\mathbf{X},\mathbf{Y})$ the i-th coordinate of the output of f^k, and similarly define $\Pi_{<i}(\mathbf{X},\mathbf{Y})$ and $f_{<i}^k(\mathbf{X},\mathbf{Y})$. Hence, by averaging, most of the k terms in the product are at least $1-O\left(\frac{\delta}{k}\right)$. Qualitatively speaking, conditioned on Π succeeding on a typical prefix of the first $i-1$ coordinates, it is much more likely to succeed on the i-th coordinate than it would be without this conditioning.

This motivates the following change to the embedding step: since $\mathbf{x}_{<i}$ and $\mathbf{y}_{<i}$ are hardwired into Π, the parties know the value $f(x_j,y_j)$ for all $j<i$, and given the output of Π, can first verify whether or not $\Pi_{<i}(\mathbf{X},\mathbf{Y})=(f(x_1,y_1),\ldots,f(x_{i-1},y_{i-1}))$. If this condition holds, then they can output $\Pi_i(\mathbf{X},\mathbf{Y})$ as usual. However, if this condition fails to hold, the parties output 'abort'. We prove that for a typical prefix $\mathbf{x}_{<i},\mathbf{y}_{<i}$, for most of the random seeds of the protocol and for most choices of random suffixes $\mathbf{X}_{>i}$ and $\mathbf{Y}_{>i}$, the following holds: the parties only abort with constant probability over the inputs $(A,B)\sim\mu$, and given that they do not abort, the output is correct with a very large $1-O(1/k)$ probability. Moreover, we still have that the information revealed by this protocol can be used to lower bound the term $\mathrm{I}(X_i,Y_i;\Pi\mid\mathbf{D},\mathbf{X}_{<i},\mathbf{Y}_{<i})$.

To complete the direct sum argument, we need a way of lower-bounding the information revealed by a protocol with this abortion property. For this, we directly bound the information revealed by designing an estimation procedure for predicting (X_i,Y_i) from the transcript of Π, and applying Fano's inequality. This part generalizes the approach used in [23], who show lower bounds for $(0,0,\delta)$-protocols. In our case we show that allowing constant probability of failure over the choice or randomness of the protocol and constant probability of abortion over the choice of the input doesn't change the asymptotic dependence of information cost

as a function of the success probability over the choice of the input conditioned on non-abortion.

Other Related Work: In [19, 20], the authors show that for $O(1)$-round public-coin randomized communication complexity $R^{\text{pub}}_{1-(1-\epsilon/2)^{\Omega(k\epsilon^2)}}(f^k) = \Omega\left(\epsilon k\left(R^{\text{pub}}_{\epsilon}(f) - O\left(\frac{1}{\epsilon^2}\right)\right)\right)$, where $\epsilon > 0$ is arbitrary. One cannot apply this theorem to our problem, as one would need to set $\epsilon = 1/k$ to obtain our results, at which point the theorem gives a trivial bound. A similar problem occurs trying to apply the direct sum theorem of [22]. These are not drawbacks of these works, since their study is for a vastly different regime of parameters, namely, for constant ϵ, and for every relation f. We instead only consider functions f for which we can lower bound the conditional information cost of protocols with the abortion property. These are of particular importance for sketching and streaming applications and for these functions we obtain the first optimal bounds.

2 The Direct Sum Theorem

We give a summary of basic properties of the entropy of a discrete random variable X, denoted as $H(X)$, and the mutual information between two discrete random variables X and Y, denoted as $I(X;Y) = H(X) - H(X|Y)$, below (see Chapter 2 in [13] for the proofs):

Proposition 2.1. *1. Entropy span:* $0 \leq H(X) \leq \log|supp(X)|$.

2. $I(X;Y) \geq 0$ *because* $H(X|Y) \leq H(X)$.

3. *Chain rule:* $I(X_1, X_2, \ldots, X_n; Y|Z) = \sum_{i=1}^{n} I(X_i; Y|X_1, \ldots, X_{i-1}, Z)$.

4. *Subadditivity:* $H(X,Y|Z) \leq H(X|Z) + H(Y|Z)$, *where the equality holds if and only if* X *and* Y *are independent conditioned on* Z.

5. *Fano's inequality: Let* A *be a random variable, which can be used as "predictor" of* X*, namely there exists a function* g *such that* $\Pr[g(A) = X] \geq 1 - \delta$ *for some* $\delta < 1/2$*. If* $|supp(X)| \geq 2$ *then* $H(X|A) \leq \delta\log(|supp(X)| - 1) + h_2(\delta)$*, where* $h_2(\delta) = \delta\log(1/\delta) + (1-\delta)\log\frac{1}{1-\delta}$ *is the binary entropy.*

We recall standard definitions from information complexity and introduce the information complexity for protocols with abortion, denoted as $\mathrm{IC}_{\mu,\alpha,\beta,\delta}(f|\nu)$, more formally. Given a communication problem $f : \mathcal{X} \times \mathcal{Y} \to \mathcal{Z}$, consider the augmented space $\mathcal{X} \times \mathcal{Y} \times \mathcal{D}$ for some $\mathcal{D}$. Let λ be a distribution over $\mathcal{X} \times \mathcal{Y} \times \mathcal{D}$, which induces marginals μ on $\mathcal{X} \times \mathcal{Y}$ and ν on $\mathcal{D}$. We say that ν *partitions* μ, if μ is a *mixture* of product distributions, namely for a random variable $(X, Y, D) \sim \lambda$, conditioning on any value of D makes the distribution of (X, Y) product.

To simplify the notation, a δ-*protocol* for f is one that for all inputs $(x, y) \in \mathcal{X} \times \mathcal{Y}$ computes $f(x, y)$ with probability at least $1 - \delta$ (over the randomness of the protocol).

Definition 2.1. *Let* Π *be a protocol, which computes* f*. The* conditional information cost *of* Π *under* λ *is defined as* $\mathrm{I}(\Pi(X,Y); X, Y \mid D)$*, where* $(X, Y, D) \sim \lambda$*. The* conditional information complexity *of* f *with respect to* λ*, denoted by* $\mathrm{IC}_{\mu,\delta}(f|\nu)$*, is defined as the minimum conditional information cost of a* δ*-protocol for* f*. The* information complexity with aborts*, denoted by* $\mathrm{IC}_{\mu,\alpha,\beta,\delta}(f|\nu)$*, is the minimum conditional information cost of a protocol that* (α, β, δ)*-computes* f*. The analogous quantities* $\mathrm{IC}^{\rightarrow}_{\mu,\delta}(f|\nu)$ *and* $\mathrm{IC}^{\rightarrow}_{\mu,\alpha,\beta,\delta}(f|\nu)$ *are defined by taking the respective minimums over only one-way protocols.*

Our main theorem gives a lower bound the conditional information cost of a δ-protocol for k copies of a communication problem. More precisely, for a function $f : \mathcal{X} \times \mathcal{Y} \to \mathcal{Z}$ let $f^k : (\mathcal{X} \times \mathcal{Y})^k \to \mathcal{Z}^k$ denote its k-fold version $f^k(\mathbf{x}, \mathbf{y}) = (f(x_1, y_1), f(x_2, y_2), \ldots, f(x_k, y_k))$.

Theorem 2.1. *Let* $\delta \leq 1/3$*. Then for every function* $f\colon \mathcal{X} \times \mathcal{Y} \to \mathcal{Z}$ *and distribution* λ *on* $\mathcal{X} \times \mathcal{Y} \times \mathcal{D}$ *with marginal* μ *on* $\mathcal{X} \times \mathcal{Y}$ *and marginal* ν *on* $\mathcal{D}$*, such that* μ *is partitioned by* ν*, it holds that* $\mathrm{IC}_{\mu^k,\delta}(f^k|\nu^k) \geq \Omega(k)\,\mathrm{IC}_{\mu,\frac{1}{20},\frac{1}{10},\frac{\delta}{k}}(f|\nu)$*. Moreover, this result also holds for 1-way protocols:* $\mathrm{IC}^{\rightarrow}_{\mu^k,\delta}(f^k|\nu^k) \geq \Omega(k)\,\mathrm{IC}^{\rightarrow}_{\mu,\frac{1}{20},\frac{1}{10},\frac{\delta}{k}}(f|\nu)$.

For the remaining part of the section we prove this theorem. Amplifying the success probability by repeating the protocol a constant number of times, it is easy to see that $\mathrm{IC}_{\mu^k,\delta}(f^k|\nu^k) = \Omega(1)\,\mathrm{IC}_{\mu^k,\delta/2000}(f^k|\nu^k)$, and similarly for one-way protocols (see Appendix A). Thus, without loss of generality we work with $(\delta/2000)$-protocols instead.

We focus on the first part of the theorem. For each $i \in [k]$, consider independent random variables $(X_i, Y_i, D_i) \sim \lambda$; to simplify the notation, we use W_i to denote the pair (X_i, Y_i). Let Π be a $(\delta/2000)$-protocol for f^k with private randomness R that achieves $\mathrm{I}(\Pi(\mathbf{W}, R); \mathbf{W}|\mathbf{D}) = \mathrm{IC}_{\mu^k,\frac{1}{2000}}(f^k|\nu^k)$. Our goal is to lower bound the mutual information $\mathrm{I}(\Pi(\mathbf{W}, R); \mathbf{W}|\mathbf{D})$ by $\frac{k}{8}\,\mathrm{IC}_{\mu,\frac{1}{20},\frac{\delta}{10},\frac{\delta}{k}}(f|\nu)$, by essentially showing that Π needs to compute most of the k instances with probability $1 - O(\delta/k)$.

To make this precise, the guarantee of the protocol gives that

$$1 - \frac{\delta}{2000} \leq \Pr(\Pi(\mathbf{W}, R) = f^k(\mathbf{W}))$$
$$= \prod_{i=1}^{k} \Pr(\Pi_i(\mathbf{W}, R) = f_i^k(\mathbf{W}) \mid \Pi_{<i}(\mathbf{W}, R) = f_{<i}^k(\mathbf{W})).$$

Using the bound $p \leq e^{-(1-p)}$ (valid for all $p \in [0,1]$) to each term in the right-hand side, we can then use Markov's inequality to show that for at least half of the indices $i \in [k]$ we have the strong conditional guarantee

$$\Pr(\Pi_i(\mathbf{W}, R) = f_i^k(\mathbf{W}) \mid \Pi_{<i}(\mathbf{W}, R) = f_{<i}^k(\mathbf{W})) \geq 1 - \frac{2\ln(1 - \frac{\delta}{2000})^{-1}}{k} \geq 1 - \frac{\delta}{200k}, \tag{2.1}$$

where the last inequality uses the first-order approximation of ln at 1. We call these indices *good*. Moreover, using the chain rule, we can express the mutual information $\mathrm{I}(\Pi(\mathbf{W}, R); \mathbf{W} \mid \mathbf{D})$ in terms of the information revealed of each component of $\mathbf{W}$:

$$\mathrm{I}(\Pi(\mathbf{W}, R); \mathbf{W} \mid \mathbf{D}) = \sum_{i=1}^{k} \mathrm{I}(\Pi(\mathbf{W}, R); W_i \mid \mathbf{D}, \mathbf{W}_{<i}). \tag{2.2}$$

The idea is then, for each good index i, to obtain from Π a protocol that $(1/20, \delta/10, \delta/k)$-computes $f_i^k(\mathbf{W})$ and which reveals only $O(\mathrm{I}(\Pi(\mathbf{W}, R); W_i \mid \mathbf{D}, \mathbf{W}_{<i}))$ conditional information about W_i, effectively showing that $\mathrm{I}(\Pi(\mathbf{W}, R); W_i \mid \mathbf{D}, \mathbf{W}_{<i}) \geq \Omega(\mathrm{IC}_{\mu, \frac{1}{20}, \frac{\delta}{10}, \frac{\delta}{k}}(f|\nu))$. This is accomplished by simulating Π over some of its input. We show next that we can "hardwire" the first $i-1$ inputs of Π while preserving the relevant properties of the protocol. Unfortunately hardwiring the last $k-i$ inputs of Π and its random seed (and thus leaving only input i free) might change the mutual information with W_i drastically; but we show that there is a large set of suffixes that still preserve most properties that we need. The existence of such suffixes is proved via the probabilistic method.

Lemma 2.1. *Consider a good index $i \in [k]$. Then there exists a prefix $\mathbf{w}_{<i} \in (\mathcal{X} \times \mathcal{Y})^{i-1}$ and a set G of fixings of the suffix $\mathbf{W}_{>i}$ and the random bits used by Π with the following properties:*

1. *(Low information cost)* $\mathrm{I}(\Pi(\mathbf{W}, R); W_i \mid \mathbf{D}, \mathbf{W}_{<i} = \mathbf{w}_{<i}) \leq 4\,\mathrm{I}(\Pi(\mathbf{W}, R); W_i \mid \mathbf{D}, \mathbf{W}_{<i})$.
2. *(Large set of fixings)* $\Pr((\mathbf{W}_{>i}, R) \in G) \geq 1 - \frac{1}{20}$.
3. *(Success probability) For every $(\mathbf{w}_{>i}, r)$ in G we have*
$$\Pr\left[\Pi_{<i}(\mathbf{w}_{<i} W_i \mathbf{w}_{>i}, r) \neq f_{<i}^k(\mathbf{w}_{<i} W_i \mathbf{w}_{>i})\right] \leq \frac{\delta}{10}.$$
4. *(Conditional success probability) For every $(\mathbf{w}_{>i}, r)$ in G we have*
$$\Pr[\Pi_i(\mathbf{w}_{<i} W_i \mathbf{w}_{>i}, r) \neq f_i^k(\mathbf{w}_{<i} W_i \mathbf{w}_{>i}) \mid \Pi_{<i}(\mathbf{w}_{<i} W_i \mathbf{w}_{>i}, r) = f_{<i}^k(\mathbf{w}_{<i} W_i \mathbf{w}_{>i})] \leq \frac{\delta}{k}.$$

Proof. We start by proving the following proposition.

Proposition 2.2. *Consider a good index $i \in [k]$. Then there exists $\mathbf{w}_{<i} \in (\mathcal{X} \times \mathcal{Y})^{i-1}$ such that the following hold:*

- $\mathrm{I}(\Pi(\mathbf{w}_{<i}\mathbf{W}_{\geq i}, R); W_i \mid \mathbf{D}, \mathbf{W}_{<i} = \mathbf{w}_{<i}) \leq 4\,\mathrm{I}(\Pi(\mathbf{W}, R); W_i \mid \mathbf{D}, \mathbf{W}_{<i})$
- $\Pr[\Pi_{<i}(\mathbf{w}_{<i}\mathbf{W}_{\geq i}, R) \neq f_{<i}^k(\mathbf{w}_{<i}\mathbf{W}_{\geq i})] \leq \frac{\delta}{500}$
- $\Pr[\Pi_i(\mathbf{w}_{<i}\mathbf{W}_{\geq i}, R) \neq f_i^k(\mathbf{w}_{<i}\mathbf{W}_{\geq i}) \mid \Pi_{<i}(\mathbf{w}_{<i}\mathbf{W}_{\geq i}, R) = f_{<i}^k(\mathbf{w}_{<i}\mathbf{W}_{\geq i})] \leq \frac{\delta}{50k}$.

Proof. We use the probabilistic method, so first we analyze the expected value of the quantities in the left-hand side of the above expression with respect to the random variable $\mathbf{W}_{<i}$.

For Item 1, it follows from the definition of conditional mutual information that

$$\mathop{\mathbb{E}}_{\mathbf{W}_{<i}} [\mathrm{I}(\Pi(\mathbf{W}_{<i}\mathbf{W}_{\geq i}, R); W_i \mid \mathbf{D}, \mathbf{W}_{<i})] = \mathop{\mathbb{E}}_{\mathbf{W}_{<i}} \left[\mathop{\mathbb{E}}_{\mathbf{D}} [\mathrm{I}(\Pi(\mathbf{W}, R); W_i \mid \mathbf{D}, \mathbf{W}_{<i}) \mid \mathbf{W}_{<i}]\right] = \mathrm{I}(\Pi(\mathbf{W}, R); W_i | \mathbf{D}, \mathbf{W}_{<i}).$$

For Item 2, the product structure of μ^k and the guarantee of Π give

$$\mathop{\mathbb{E}}_{\mathbf{W}_{<i}} \left[\Pr\left(\Pi_{<i}(\mathbf{W}_{<i}\mathbf{W}_{\geq i}, R) \neq f_{<i}^k(\mathbf{w}_{<i}\mathbf{W}_{\geq i})\right)\right] = \Pr(\Pi_{<i}(\mathbf{W}, R) \neq f_{<i}^k(\mathbf{W})) \leq \Pr(\Pi(\mathbf{W}, R) \neq f^k(\mathbf{W})) \leq \frac{\delta}{2000}.$$

For Item 3, we now use the fact that i is good to obtain

$$\sum_{\mathbf{w}_{<i}} \Pr\left[\Pi_i(\mathbf{w}_{<i}\mathbf{W}_{\geq i}, R) \neq f_i^k(\mathbf{w}_{<i}\mathbf{W}_{\geq i}) \mid \Pi_{<i}(\mathbf{w}_{<i}\mathbf{W}_{\geq i}, R) = f_{<i}^k(\mathbf{w}_{<i}\mathbf{W}_{\geq i})\right] \cdot \Pr(\mathbf{W}_{<i} = \mathbf{w}_{<i} \mid \Pi_{<i}(\mathbf{W}, R) = f_{<i}^k(\mathbf{W}))$$
$$= \Pr[\Pi_i(\mathbf{W}, R) \neq f_i^k(\mathbf{W}) \mid \Pi_{<i}(\mathbf{W}, R) = f_{<i}^k(\mathbf{W})] \leq \frac{\delta}{200k}.$$

Although this last expectation is with respect to the distribution conditioned on $\Pi_{<i}(\mathbf{W}, R) = f_{<i}^k(\mathbf{W})$,

because of the guarantee of Π, this conditioning does not change the distribution by much; more precisely, for every event E we have $\Pr(E) \leq \Pr(E \mid \Pi_{<i}(\mathbf{W}, R) = f^k_{<i}(\mathbf{W})) + \delta/2000 < \Pr(E \mid \Pi_{<i}(\mathbf{W}, R) = f^k_{<i}(\mathbf{W})) + 1/4$.

Using Markov's inequality to upper bound the probability of being 4 times larger than the expectation in each of the 3 items and taking a union bound, we obtain that the there is a $\mathbf{w}_{<i}$ satisfying the desired properties in the proposition. This concludes the proof. □

The proof of Lemma 2.1 then follows from Proposition 2.2 above and again from the application of Markov's inequality and the union bound. □

Now we use the protocol Π hardwiring $\mathbf{W}_{<i} = \mathbf{w}_{<i}$ (for a $\mathbf{w}_{<i}$ as above) and $\mathbf{D}_{-i} = \mathbf{d}_{-i}$ to obtain a protocol to $(1/20, \delta/10, \delta/k)$-compute f under the distribution μ. The idea is to simulate the inputs $\mathbf{W}_{>i}$ (conditioned on $\mathbf{D} = \mathbf{d}$) and run the protocol $\Pi(\mathbf{w}_{<i} W_i \mathbf{W}_{\geq i}, R)$, aborting whenever $\Pi_{<i}(\mathbf{w}_{<i} W_i \mathbf{W}_{\geq i}, R) \neq f^k_{<i}(\mathbf{w}_{<i} W_i \mathbf{W}_{\geq i})$.

Lemma 2.2. *Consider a good $i \in [k]$, let $\mathbf{w}_{<i}$ satisfy Lemma 2.1 and let $\mathbf{d}_{-i}$ be such that $\Pr(\mathbf{W}_{<i} = \mathbf{w}_{<i}, \mathbf{D}_{-i} = \mathbf{d}_{-i}) \neq 0$. Then there exists a protocol $\bar{\Pi}$ with input in $\mathcal{X} \times \mathcal{Y}$ and only private randomness $\bar{R}$ satisfying the following:*

- *$\bar{\Pi}$ $(\frac{1}{20}, \frac{\delta}{10}, \frac{\delta}{k})$-computes f with respect to the distribution μ*
- *For $(\bar{W}, \bar{D}) \sim \lambda$, $\mathrm{I}(\bar{\Pi}(\bar{W}, \bar{R}); \bar{W} \mid \bar{D}) = \mathrm{I}(\Pi(\mathbf{W}, R); W_i \mid D_i, \mathbf{D}_{-i} = \mathbf{d}_{-i}, \mathbf{W}_{<i} = \mathbf{w}_{<i})$.*

Moreover, if Π is 1-way, then $\bar{\Pi}$ is also 1-way.

Proof. The protocol $\bar{\Pi}$ is constructed as follows. Suppose that Alice has input $x \in \mathcal{X}$ and Bob has input $y \in \mathcal{Y}$. Since ν partitions μ, Alice and Bob use their private randomness to sample respectively $\mathbf{X}'_{>i}$ and $\mathbf{Y}'_{>i}$ according to the distribution μ^{k-i} conditioned on $\mathbf{D}_{-i} = \mathbf{d}_{-i}$; more precisely, the random variable $(\mathbf{X}'_{>i}, \mathbf{Y}'_{>i})$ has the same distribution as $(\mathbf{X}_{>i}, \mathbf{Y}_{>i}) \mid (\mathbf{D}_{-i} = \mathbf{d}_{-i})$. They also use their private randomness to obtain a random variable R' with same distribution as the random coins used in Π.

Using these random variables, the players run the protocol $\Pi(\mathbf{w}_{<i}, (x, y), (\mathbf{X}'_{>i}, \mathbf{Y}'_{>i}), R')$ to obtain estimates of the vector-valued function $f^k(\mathbf{w}_{<i}, (x, y), (\mathbf{X}'_{>i}, \mathbf{Y}'_{>i}))$. Finally, since $\mathbf{w}_{<i}$ is known to Bob, he checks whether Π gave the correct values of the first $i-1$ coordinates of $f^k(\mathbf{w}_{<i}, (x, y), (\mathbf{X}'_{>i}, \mathbf{Y}'_{>i})$, namely if

$$\Pi_j(\mathbf{w}_{<i}, (x, y), (\mathbf{X}'_{>i}, \mathbf{Y}'_{>i}), R') = f^k_j(w_j)$$

for all $j < i$; if so, he outputs the estimate of $f(x, y) = f^k_i(\mathbf{w}_{<i}, (x, y), (\mathbf{X}'_{>i}, \mathbf{Y}'_{>i}))$ given by $\Pi_i(\mathbf{w}_{<i}, (x, y), (\mathbf{X}'_{>i}, \mathbf{Y}'_{>i}), R')$, and otherwise he aborts. Let

$$\bar{\Pi}(x, y, \bar{R}) = \Pi(\mathbf{w}_{<i}, (x, y), (\mathbf{X}'_{>i}, \mathbf{Y}'_{>i}), R')$$

to denote the transcript exchanged with (and output of) this protocol.

We first analyze the information revealed by the protocol. Consider $(\bar{W}, \bar{D}) \sim \lambda$. Using the definition of our random variables and the product structure of λ^k, it follows by substitution of random variables that

$$\begin{aligned}&\mathrm{I}(\bar{\Pi}(\bar{W}, \bar{R}); \bar{W} \mid \bar{D}) \\ &\quad = \mathrm{I}(\Pi(\mathbf{W}, R); W_i \mid D_i, \mathbf{D}_{-i} = \mathbf{d}_{-i}, \mathbf{W}_{<i} = \mathbf{w}_{<i}),\end{aligned}$$

which gives the second part of the lemma.

For the correctness of the protocol, let the set G be defined as in Lemma 2.1. Take any $(\mathbf{w}_{>i}, r) \in G$; we claim that, conditioned on $((\mathbf{X}'_{>i}, \mathbf{Y}'_{>i}), R') = (\mathbf{w}_{>i}, r)$, the protocol $\bar{\Pi}$ $(\delta/10, \delta/k)$-computes f (notice that conditioned on $((\mathbf{X}'_{>i}, \mathbf{Y}'_{>i}), R') = (\mathbf{w}_{>i}, r)$ the protocol is indeed a deterministic one). Since the event $((\mathbf{X}'_{>i}, \mathbf{Y}'_{>i}), R') \in G$ only depends on the randomness of the protocol, and since $\Pr(((\mathbf{X}'_{>i}, \mathbf{Y}'_{>i}), R') \in G) \geq 1 - 1/20$, this implies that $\bar{\Pi}$ $(1/20, \delta/10, \delta/k)$-computes f.

To prove the claim, let E denote the event $((\mathbf{X}'_{>i}, \mathbf{Y}'_{>i}), R') = (\mathbf{w}_{>i}, r)$. It follows again from the definition of our random variables that the probability that $\bar{\Pi}(\bar{X}, \bar{Y}, \bar{R})$ aborts conditioned on E is equal to the probability that $\Pi_{<i}(\mathbf{w}_{<i} \mathbf{W}_{\geq i}, R) \neq f^k_{<i}(\mathbf{w}_{<i} \mathbf{W}_{\geq i})$ conditioned on $(\mathbf{D}_{-i}, \mathbf{W}_{>i}, R) = (\mathbf{d}_{-i}, \mathbf{w}_{>i}, r)$. Using the mutual independence between W_i, $\mathbf{W}_{>i}$ and R, this is the same as the probability that $\Pi_{<i}(\mathbf{w}_{<i} W_i \mathbf{w}_{>i}, r) \neq f^k_{<i}(\mathbf{w}_{<i} W_i \mathbf{w}_{\geq i})$; by definition of G (Item 3 of Lemma 2.1), this probability is at most $\delta/10$. Similarly, we obtain that

$$\begin{aligned}&\Pr\left[\bar{\Pi}(\bar{W}, \bar{R}) \neq f(W') \mid \bar{\Pi} \text{ does not abort}, E\right] \\ &\quad = \Pr\Big[\Pi_i(\mathbf{w}_{<i} \mathbf{W}_{\geq i}, R) \neq f^k_i(\mathbf{w}_{<i} \mathbf{W}_{\geq i}) \mid \\ &\qquad \Pi_{<i}(\mathbf{w}_{<i} \mathbf{W}_{\geq i}, R) = f^k_{<i}(\mathbf{w}_{<i} \mathbf{W}_{\geq i}), \\ &\qquad\qquad (\mathbf{D}_{-i}, \mathbf{W}_{>i}, R) = (\mathbf{d}_{-i}, \mathbf{w}_{>i}, r)\Big] \leq \frac{\delta}{k},\end{aligned}$$

where the last inequality follows again from the product structure of λ^k, independence of R from the other random variables, and from the definition of G. This proves the claim and shows that $\bar{\Pi}$ $(1/20, \delta/10, \delta/k)$-computes f, giving the second item in the lemma.

Finally, notice that if Π is one-way then $\bar{\Pi}$ is also one-way. This concludes the proof of the lemma. □

The previous lemma (averaged out over all $\mathbf{d}_{-i}$), together with the first part of Lemma 2.1, gives that

for every good index $i \in [k]$ we can lower bound $\mathrm{I}(\Pi(\mathbf{W}, R); W_i \mid \mathbf{D}, \mathbf{W}_{<i})$ by $\frac{1}{4}\,\mathrm{IC}_{\mu,\frac{1}{20},\frac{\delta}{10},\frac{\delta}{k}}(f|\nu)$. Since at least half of the i's in $[k]$ are good, plugging this bound on (2.2) gives that $\mathrm{IC}_{\mu^k,\frac{\delta}{2000}}(f^k|\nu^k) \geq \frac{k}{8}\,\mathrm{IC}_{\mu,\frac{1}{20},\frac{\delta}{10},\frac{\delta}{k}}(f|\nu)$, and similarly for the one-way information complexity. This concludes the proof of Theorem 2.1.

3 Lower Bounds for Protocols with Abortion

In this section we prove lower bounds on the information cost of one-way protocols with abortion. To illustrate the techniques, we first consider the equality problem. In order to make the argument more formal, we introduce the following formalization of one-way protocols. Alice has a (possibly random) function $M : \mathcal{X} \to \mathcal{M}$ and Bob has a (also possibly random) function $B : \mathcal{M} \times \mathcal{Y} \to \mathcal{Z}$ that depends on the received message and on its input, and $B(M(x), y)$ is the estimate for $f(x, y)$ output by Bob. Consider the augmented space $\mathcal{X} \times \mathcal{Y} \times \mathcal{D}$ and let λ be a distribution on it that has marginal μ over $\mathcal{X} \times \mathcal{Y}$ and marginal ν over $\mathcal{D}$. Notice that, whenever ν partitions μ, the conditional information cost of the protocol (M, B) is given by $\mathrm{I}(M(X); X \mid D) = \mathrm{I}(M(X); X, Y \mid D)$, where $(X, Y, D) \sim \lambda$. For such distributions, $\mathrm{IC}^{\to}_{\mu,\delta}(f|\nu)$ is the minimum of $\mathrm{I}(M(X); X \mid D)$ over all one-way δ-protocols (M, B) for f.

3.1 Equality Problem

Let EQ^ℓ denote the equality problem: Alice and Bob have respectively the binary strings $\mathbf{x}$ and $\mathbf{y}$ of length ℓ and their goal is to check whether $\mathbf{x} = \mathbf{y}$ or not.

Lemma 3.1. *For $\ell = \log(1/20\delta)$, with $\delta \in (0, 1)$, there exists a distribution with marginals μ and ν, such that ν partitions μ and*

$$\mathrm{IC}^{\to}_{\mu,\frac{1}{20},\frac{1}{10},\delta}(\mathrm{EQ}^\ell|\nu) = \Omega\left(\log(1/\delta)\right).$$

Proof. To construct μ and ν, let D_0 be a random variable uniformly distributed on $\{0, 1\}$ and let $\mathbf{D}$ be a random variable uniformly distributed on $\{0, 1\}^\ell$. Let $(\mathbf{X}, \mathbf{Y})$ be a random variable supported on $\{0, 1\}^\ell \times \{0, 1\}^\ell$ such that, conditioned on $D_0 = 0$ we have $\mathbf{X}$ and $\mathbf{Y}$ distributed independently and uniformly on $\{0, 1\}^\ell$, and conditioned on $D_0 = 1$ we have $\mathbf{X} = \mathbf{Y} = \mathbf{D}$. Let μ be the distribution of $(\mathbf{X}, \mathbf{Y})$ and let ν be the distribution of $(D_0\mathbf{D})$. Note that ν partitions μ.

Consider a one-way protocol Π for EQ^ℓ and let M denote Alice's message function. Since $\mathbf{X}$ and $\mathbf{Y}$ are independent conditioned on $D_0\mathbf{D}$, we have

$$\begin{aligned}\mathrm{I}(M(\mathbf{X}); \mathbf{X}, \mathbf{Y} \mid D_0\mathbf{D}) &= \mathrm{I}(M(\mathbf{X}); \mathbf{X} \mid D_0\mathbf{D}) \\ &= \mathrm{H}(\mathbf{X} \mid D_0\mathbf{D}) - \mathrm{H}(\mathbf{X} \mid M(\mathbf{X}), D_0\mathbf{D}).\end{aligned}$$

Notice that $\mathrm{H}(\mathbf{X} \mid D_0\mathbf{D}) \geq \frac{1}{2}\,\mathrm{H}(\mathbf{X} \mid D_0 = 0, \mathbf{D}) = \frac{1}{2}\log(1/20\delta)$.

From Fano's inequality [13] we also have

$$\mathrm{H}(\mathbf{X} \mid M(\mathbf{X}), D_0\mathbf{D}) \leq \mathrm{H}(\mathbf{X} \mid M(\mathbf{X})) \leq 1 + p_e \log(|supp(\mathbf{X})|),$$

where $p_e = \min_g \Pr[g(M(\mathbf{X})) \neq \mathbf{X}]$ is the minimum error over all predictors g. Thus, to prove the lemma it suffices to show that if Π $(1/20, 1/10, \delta)$-computes EQ^ℓ then we can obtain a predictor with error at most $2/5$.

First assume Π is a deterministic protocol that $(1/10, \delta)$-computes EQ^ℓ. We say that an input $\mathbf{x}$ for Alice is *good* if $\Pi(\mathbf{x}, \mathbf{y}) = 1$ iff $\mathbf{x} = \mathbf{y}$; we claim that many inputs are good. Note that the probability mass that our distribution assigns to every input $(\mathbf{x}, \mathbf{x})$ is

$$\begin{aligned}p_1 &= \Pr[D_0 = 0]\Pr[\mathbf{X} = \mathbf{Y} = \mathbf{x} \mid D_0 = 0] \\ &\quad + \Pr[D_0 = 1]\Pr[\mathbf{X} = \mathbf{Y} = \mathbf{x} \mid D_0 = 1] = 200\delta^2 + 10\delta.\end{aligned}$$

The probability assigned to every input $(\mathbf{x}, \mathbf{y})$ for $\mathbf{x} \neq \mathbf{y}$ is equal to $p_2 = 200\delta^2$. So the number of $\mathbf{x}$'s such that $\Pi(\mathbf{x}, \mathbf{x}) = \mathit{abort}$ is at most $\Pr[\Pi = \mathit{abort}]/p_1 = 1/(10p_1) \leq 1/(100\delta)$. Similarly, the number of $\mathbf{x}$'s such that there is at least one $\mathbf{y}$ where the protocol does not abort but makes a mistake is at most $\Pr[\Pi \neq \mathrm{EQ}^\ell, \Pi \neq \mathit{abort}]/p_2 \leq \delta/(200\delta^2) = 1/(200\delta)$. Finally notice that if $\mathbf{x}$ does not satisfy either of these two conditions then $\mathbf{x}$ is good. This implies that there are at most $\frac{3}{200\delta}$ not good $\mathbf{x}$'s, and hence the probability that $\mathbf{X}$ is not good is at most $3/10$.

Now notice that if $\mathbf{x}$ is good then we can recover $\mathbf{x}$ itself from $M(\mathbf{x})$ using Π: simply find the unique $\mathbf{y}$ such that Bob outputs 1 upon receiving message $M(\mathbf{x})$. This then gives a predictor g with error probability $p_e \leq 3/10$ as desired.

For the case where Π only $(1/20, 1/10, \delta)$-computes EQ^ℓ, we can use the same argument as before and run Bob's part of the protocol over all $\mathbf{y}$ upon receiving message $M(\mathbf{x})$, but now we need Bob's private coins R^B to do it. This gives a predictor for $\mathbf{X}$ using $M(\mathbf{X})$ and R^B with error at most $3/10 + 1/20 \leq 2/5$, which shows that $\mathrm{H}(\mathbf{X} \mid M(\mathbf{X}), D_0\mathbf{D}) = \mathrm{H}(\mathbf{X} \mid M(\mathbf{X}), D_0\mathbf{D}, R^B) \leq 1 + \frac{2}{5}\log(1/20\delta)$. This concludes the proof. □

3.2 Augmented Indexing

In order to obtain the desired lower bound for our applications we need a generalization of EQ^ℓ, namely the augmented indexing problem on large domain with low error $\mathrm{Ind}^a(k, N)$, presented in the introduction.

Theorem 3.1. *Consider an integer k and a parameter δ such that k is at least a sufficiently large constant and $\delta \leq \frac{1}{20k}$. Then there is a distribution with marginals μ and ν such that ν partitions μ and $\mathrm{IC}^{\to}_{\mu,\frac{1}{20},\frac{1}{10},\delta}(\mathrm{Ind}^a(k, N)|\nu) \geq \Omega\left(N \log k\right)$.*

In the remainder of this section we prove Theorem 3.1. To do so, we consider the following hard distribution for $\mathrm{Ind}^a(k,N)$. First we have the random variable $\mathbf{D}$ uniformly distributed in $[k]^N$ and a random variable D_0 taking value 0 or 1 with equal probability. The distribution of Alice's input is given by $\mathbf{X}$ and the distribution of Bob's input is given by $(I, \mathbf{Y}_{<I}, Y)$ as follows: when $D_0 = 1$, we set I uniformly at random from $[N]$, $\mathbf{Y}_{<I} = \mathbf{X}_{<I} = \mathbf{D}_{<I}$, $Y = X_I = D_I$ and $\mathbf{X}_{>I}$ uniformly in $[k]^{N-I}$; when $D_0 = 0$, we again set I uniformly at random from $[N]$, $\mathbf{Y}_{<I} = \mathbf{X}_{<I} = \mathbf{D}_{<I}$, $\mathbf{X}_{>I}$ uniformly in $[k]^{N-I}$, but now Y and X_I are picked independently and uniformly at random in $[k]$.

Let λ denote the joint distribution of $(\mathbf{X}, I, \mathbf{Y}_{<I}, Y, D_0, \mathbf{D}_{\leq I})$, with marginal μ over $(\mathbf{X}, I, \mathbf{Y}_{<I}, Y)$ and marginal ν over $(D_0, \mathbf{D}_{\leq I})$ (notice that the we use $\mathbf{D}_{\leq I}$ and not $\mathbf{D}$). We remark that μ is partitioned by ν.

Now we show that Theorem 3.1 holds with the distribution defined above. For that, consider a private-randomness one-way protocol given by Alice's message function M and Bob's output function B that $(1/20, 1/10, \delta)$-computes $\mathrm{Ind}^a(k,N)$ with respect to μ and has conditional information cost $\mathrm{I}(M(\mathbf{X}); \mathbf{X} \mid D_0\mathbf{D}_{\leq I}) = \mathrm{IC}^{\rightarrow}_{\mu,\frac{1}{20},\frac{1}{10},\delta}(\mathrm{Ind}^a(k,N) \mid \nu)$. We show that the mutual information $\mathrm{I}(M(\mathbf{X}); \mathbf{X} \mid D_0\mathbf{D}_{\leq I})$ is $\Omega(N \log k)$.

First, using the chain rule for mutual information, we express the above conditional information in terms of the conditional information of each X_i revealed by $M(\mathbf{X})$:

$$\mathrm{I}(M(\mathbf{X}); \mathbf{X} \mid D_0\mathbf{D}_{\leq I}) = \sum_{i=1}^{N} \mathrm{I}(M(\mathbf{X}); X_i \mid D_0\mathbf{D}_{\leq I}, \mathbf{X}_{<i}) \tag{3.3}$$

$$= \sum_{i=1}^{N} \mathrm{H}(X_i \mid D_0\mathbf{D}_{\leq I}, \mathbf{X}_{<i}) - \sum_{i=1}^{N} \mathrm{H}(X_i \mid M(\mathbf{X}), D_0\mathbf{D}_{\leq I}, \mathbf{X}_{<i}).$$

We first claim that for each i, the term $\mathrm{H}(X_i \mid D_0\mathbf{D}_{\leq I}, \mathbf{X}_{<i})$ is at least $(\frac{1}{2N} + \frac{i-1}{N}) \log k$. To see this, notice that conditioned on $I = i$ and $D_0 = 0$, X_i is independent of $\mathbf{D}_{\leq I}$, and $\mathrm{H}(X_i \mid D_0 = 0, \mathbf{D}_{\leq I}, \mathbf{X}_{<i}, I = i) = \log k$. Similarly, conditioned on $I < i$, X_i is independent of $\mathbf{D}_{\leq I}$ and hence $\mathrm{H}(X_i \mid D_0\mathbf{D}_{\leq I}, \mathbf{X}_{<i}, I < i) = \log k$. Since the first event holds with probability $1/2N$ and the second holds with probability $(i-1)/N$, it follows that $\mathrm{H}(X_i \mid D_0\mathbf{D}_{\leq I}, \mathbf{X}_{<i}) \geq (\frac{1}{2N} + \frac{i-1}{N}) \log k$. Adding over all i's then gives that

$$\sum_{i=1}^{N} \mathrm{H}(X_i \mid D_0\mathbf{D}_{\leq I}, \mathbf{X}_{<i}) \geq \frac{N}{2} \log k.$$

Now we need to upper bound the second summation in (3.3). For that, we will show that the guarantee of the protocol implies that $M(\mathbf{X})$ together with the prefix $\mathbf{X}_{<i}$ leads to a good predictor of X_i (for most i's); an application of Fano's inequality will then give the desired upper bound.

To make things more explicit, let R^A and R^B denote respectively Alice's and Bob's private randomness, and define $R = (R^A, R^B)$. To simplify the notation we use $\Pi(\mathbf{x}, j, y, r)$ to denote the transcript (and, as usual, also the output) of the protocol when Alice get $\mathbf{x}$, Bob gets $(j, \mathbf{x}_{<j}, y)$ and the random seed is $r = (r^A, r^B)$, namely $\Pi(\mathbf{x}, j, y, r) = B(M(\mathbf{X}, r^A), j, \mathbf{x}_{\leq j}, y, r^B)$. We also use $f(\mathbf{x}, j, y)$ to denote the function of the associated communication game, namely $f(\mathbf{x}, j, y)$ equals 0 if $x_j \neq y$ and 1 if $x_j = y$.

We first focus on tuples $(i, \mathbf{x}, r)$ that allows for a good predictor of x_i. To capture the bad tuples, let U_1 be the set of tuples $(i, \mathbf{x}, r)$ such that the protocol with random seed r aborts on the instances where Alice has input $\mathbf{x}$ and Bob has input $(i, \mathbf{x}_{<i}, \mathbf{x}_i)$ (so it is an 'equal' input), namely $U_1 = \{(i, \mathbf{x}, r) : \Pi(\mathbf{x}, i, x_i, r) =$ 'abort'$\}$. Also define U_2 as the tuples $(i, \mathbf{x}, r)$ where the protocol with random seed r makes a mistake (but does not abort) when Alice gets input $\mathbf{x}$ and Bob gets input $(i, \mathbf{x}_{<i}, y)$ for *some* y, namely $U_2 = \{(i, \mathbf{x}, r) : \exists y \text{ st } \Pi(\mathbf{x}, i, y, r) \neq f(\mathbf{x}, i, y) \text{ and } \Pi(\mathbf{x}, i, y, r) \neq \textit{abort}\}$. We say that a tuple $(i, \mathbf{x}, r)$ is *good* if it does not belong to either U_1 or U_2.

Notice that if $(i, \mathbf{x}, r)$ is good, then: (i) $\Pi(\mathbf{x}, i, x_i, r) = 1$; (ii) for every $y \neq x_i$, $\Pi(\mathbf{x}, i, y, r) \neq 1$. Good tuples render a good predictor for X_i.

Lemma 3.2. *For every index $i \in [N]$, there is a predictor g_i such that*

$$\Pr\left[g_i(M(\mathbf{X}, R^A), \mathbf{X}_{<i}) = X_i\right] \geq \Pr((i, \mathbf{X}, R) \textit{ is good}).$$

Proof. We are first going to use the protocol and the information $M(\mathbf{x}, r), \mathbf{x}_{<i}, r^B$ to estimate x_i as follows: let $\tilde{g}_i(M(\mathbf{x}, r^A), \mathbf{x}_{<i}, r^B)$ be any value y such that $\Pi(\mathbf{x}, i, y, r) = B(M(\mathbf{x}, r^A), i, \mathbf{x}_{<i}, y, r^B) = 1$. (If no such y exists, set the function value to any arbitrary value). It follows directly from the paragraph before the statement of the lemma that $\tilde{g}_i(M(\mathbf{x}, r), \mathbf{x}_{<i}, r^B) = x_i$ for all good $(i, \mathbf{x}, r)$, and hence

$$\mathop{\mathbb{E}}_{R^B}\left[\Pr\left[\tilde{g}_i(M(\mathbf{X}, R^A), \mathbf{X}_{<i}, R^B) = X_i\right]\right]$$
$$= \Pr\left[\tilde{g}_i(M(\mathbf{X}, R^A), \mathbf{X}_{<i}, R^B) = X_i\right] \geq \Pr((i, \mathbf{X}, R) \text{ is good}).$$

To remove the dependence on R^B, simply choose an outcome r^B such that

$$\Pr\left[\tilde{g}_i(M(\mathbf{X}, R^A), \mathbf{X}_{<i}, r^B) = X_i\right] \geq \Pr((i, \mathbf{X}, R) \text{ is good}),$$

and set $g_i(m, \mathbf{x}_{<i}) = \tilde{g}_i(m, \mathbf{x}_{<i}, r^B)$. □

Using this lemma and Fano's inequality [13], we obtain that

$$\sum_{i=1}^{N} \mathrm{H}(X_i \mid M(\mathbf{X}, R^A), D_0 \mathbf{D}_{\leq I}, \mathbf{X}_{<i}) \leq N + \log k \sum_{i=1}^{N} \Pr((i, \mathbf{X}, R) \text{ is not good}).$$

Since we have assumed that k is at least a sufficiently large constant, it suffices to show that $\sum_{i=1}^{N} \Pr((i, \mathbf{X}, R) \text{ is not good}) \leq 9N/20$. The following lemma then concludes the proof.

LEMMA 3.3. $\Pr((I, \mathbf{X}, R)$ *is not good*$) \leq 9/20$.

Proof. Using the union bound, we get that the probability that $(I, \mathbf{X}, R)$ is not good is at most the probability that it belongs to U_1 plus the probability that it belongs to U_2. We claim that $\Pr((I, \mathbf{X}, R) \in U_1) \leq 3/10$. Using the definition of U_1 and the fact that the random variable $(\mathbf{X}, I, X_I, R)$ has the same distribution as $(\mathbf{X}, I, Y, R)|(D_0 = 1)$, we get that

$$\begin{aligned} &\Pr((I, \mathbf{X}, R) \in U_1) = \Pr(\Pi(\mathbf{X}, I, X_I, R) = \mathit{abort}) \\ &= \Pr(\Pi(\mathbf{X}, I, Y, R) = \mathit{abort} \mid D_0 = 1) \\ &= \Pr(\text{protocol aborts} \mid D_0 = 1). \end{aligned}$$

Furthermore, since the protocol $(1/20, 1/10, \delta)$-computes f, by union bound we see that the probability that it aborts is at most $3/20$. Therefore, using the fact that $\Pr(D_0 = 1) = 1/2$, we directly get that $\Pr((I, \mathbf{X}, R) \in U_1) \leq 3/10$.

Now we claim that the second term $\Pr((I, \mathbf{X}, R) \in U_2)$ is at most $3/20$. To do so, let C denote the event (which is solely determined by the random seed R) that the protocol $(1/10, \delta)$-computes f. Given that C happens with probability at least $1/20$, to prove the claim it suffices to show $\Pr((I, \mathbf{X}, R) \in U_2 \mid C) \leq 1/10$. For that, let Err denote the event that $\Pi(\mathbf{X}, I, Y, R) \neq f(\mathbf{X}, I, Y)$ and $\Pi(\mathbf{X}, I, Y, R) \neq \mathit{abort}$. Similar to the previous case, using the definition of U_2 and the fact that the random variable $(\mathbf{X}, I, y, R)|C$ has the same distribution as $(\mathbf{X}, I, Y, R)|(D_0 = 0, Y = y, C)$, we get

$$\begin{aligned} &\Pr((I, \mathbf{X}, R) \in U_2 \mid C) = \\ &\Pr\left[\bigvee_{y \in [k]} (\Pi(\mathbf{X}, I, y, R) \neq f(\mathbf{X}, I, y) \text{ and } \Pi(\mathbf{X}, I, y, R) \neq \mathit{abort}) \,|C\right] \\ &\leq \sum_{y \in [k]} \Pr\left(Err \mid D_0 = 0, Y = y, C\right) \\ &= k \cdot \mathop{\mathbb{E}}_{Y} \left[\Pr\left(Err \mid D_0 = 0, Y, C\right) \mid D_0 = 0, C\right] \\ &= k \cdot \Pr\left(Err \mid D_0 = 0, C\right), \end{aligned}$$

where the second equality follows from the fact that $\Pr(Y = y \mid D_0 = 0, C) = \Pr(Y = y \mid D_0 = 0) = 1/k$ for all y.

By definition of C, we have that $\Pr(Err \mid C) \leq \delta$, so using the fact that $\Pr(D_0 = 0 \mid C) = \Pr(D_0 = 0) = 1/2$ we obtain that $\Pr(Err \mid D_0 = 0, C) \leq 2\delta$. Plugging this bound in the last displayed equation and using the fact that $\delta \leq 1/20k$, we get that $\Pr((I, \mathbf{X}, R) \in U_2 \mid C) \leq 1/10$ as desired. This concludes the proof of the lemma. □

4 Applications

For our application we will often assume that the dimension d of the vectors that we consider satisfies $d^{1-\gamma} \geq \frac{1}{\epsilon^2} \log \frac{n}{\delta}$ for some constant $\gamma > 0$ (where n is the number of copies of the communication problem), otherwise Alive can simply send her whole input to Bob.

All of our lower bounds come from a reduction to the same hard problem, which is an n-fold version of the augmented indexing problem with a further indexing on top of it.

4.1 Hard Problem

During our reductions it will be more convenient to work with a different reformulation of the augmented indexing problem $\mathrm{Ind}^a(u, N)$. In this new problem, Alice has a set $S \subseteq [1/(\epsilon^2\delta)]$ of size exactly $1/\epsilon^2$, where the i-th element is required to belong to the range $[\frac{(i-1)}{\delta} + 1, \frac{i}{\delta}]$ (so S selects an integral element from each interval $[\frac{(i-1)}{\delta} + 1, \frac{i}{\delta}]$ with $i \in [1/\epsilon^2]$). Bob has an element $k \in [1/(\epsilon^2\delta)]$ and also the set $S' \subseteq S$ consisting of the elements in S which are strictly smaller than k. Their goal is to decide whether k belongs to S or not. Denote this problem by $\mathrm{SetInd}(\epsilon, \delta)$.

We claim that the problem $\mathrm{SetInd}(\epsilon, \delta)$ is equivalent to the problem $\mathrm{Ind}^a(u, N)$ with $N = 1/\epsilon^2$ and universe size $u = 1/\delta$. To see this, given elements $x_1, x_2, \ldots, x_N$ in $[u]$, we can "concatenate" them to form the set $\{x_1, u + x_2, \ldots, (N-1)u + x^N\} \subseteq [1/(\epsilon^2\delta)]$. Therefore, given an instance of $\mathrm{Ind}^a(u, N)$ it is easy to construct an instance of $\mathrm{SetInd}(\epsilon, \delta)$ (with the same yes/no answer) using this concatenation. Moreover, we can reverse this operation and use it to obtain the reverse mapping from an instance of $\mathrm{SetInd}(\epsilon, \delta)$ to an instance of $\mathrm{Ind}^a(u, N)$.

Using this correspondence, Theorem 3.1 directly gives the following.

COROLLARY 4.1. *Assume that δ is at most a sufficiently small constant. Then there is a distribution with marginals μ and ν such that ν partitions μ and* $\mathrm{IC}^{\rightarrow}_{\mu, \frac{1}{20}, \frac{1}{10}, \delta}(\mathrm{SetInd}(\epsilon, \delta) \mid \nu) \geq \Omega(\frac{1}{\epsilon^2} \log \frac{1}{\delta})$.

Now we consider the n-fold version of this problem: Alice and Bob receive n instances of $\mathrm{SetInd}(\epsilon, \delta/n)$ and they want, with probability at least $1 - \delta$, to solve all of them. Denote this problem by $n\mathrm{SetInd}(\epsilon, \delta)$. Our direct sum theorem directly gives the following.

COROLLARY 4.2. *Assume that δ is at most a sufficiently small constant. Then there is a distribution*

with marginals μ and ν such that ν partitions μ and $\mathrm{IC}^{\rightarrow}_{\mu^n,\delta}(n\mathit{SetInd}(\epsilon,\delta)|\nu^n) \geq \Omega(n\frac{1}{\epsilon^2}\log\frac{n}{\delta})$.

Finally, we take an augmented indexing of r copies of this problem to obtain our hard problem $\mathrm{Ind}(n\mathrm{SetInd}(\epsilon,\delta),r)$. More precisely, an instance of $\mathrm{Ind}(n\mathrm{SetInd}(\epsilon,\delta),r)$ is obtained as follows: consider r instances $(\mathcal{S}_1^A,\mathcal{S}_1^B),\ldots,(\mathcal{S}_r^A,\mathcal{S}_r^B)$ of $n\mathrm{SetInd}(\epsilon,\delta)$ (where $\mathcal{S}_i^A$ and $\mathcal{S}_i^B$ denote respectively Alice's and Bob's part of the input); then Alice receives $\mathcal{S}_1^A,\ldots,\mathcal{S}_r^A$ and Bob receives and index j and the collections $\mathcal{S}_1^A,\ldots,\mathcal{S}_{j-1}^A$ and $\mathcal{S}_j^B$; their goal is to solve the instance $(\mathcal{S}_j^A,\mathcal{S}_j^B)$.

The following lower bound follows from Corollary 4.2 and standard direct sum arguments; for completeness we present a proof in Section B.1 of the appendix.

Corollary 4.3. *Assume that δ is at most a sufficiently small constant. Then there is a distribution with marginals μ and ν such that ν partitions μ and* $\mathrm{IC}^{\rightarrow}_{\mu,\delta}(\mathrm{Ind}(n\mathit{SetInd}(\epsilon,\delta),r)|\nu) \geq \Omega(r\cdot n\frac{1}{\epsilon^2}\log\frac{n}{\delta})$.

4.2 Estimating Multiple ℓ_p Distances Consider the following communication problem: Alice has n vectors $\mathbf{v}^1,\mathbf{v}^2,\ldots,\mathbf{v}^n \in [\pm M]^d$, Bob has n vectors $\mathbf{u}^1,\mathbf{u}^2,\ldots,\mathbf{u}^n \in [\pm M]^d$, and their goal is to compute (with probability at least $1-\delta$) approximations $(1\pm\epsilon)\|\mathbf{u}^i-\mathbf{v}^i\|_p$ to the ℓ_p distances *for all* $i\in[n]$. Let $\ell_p(n,d,M,\epsilon)$ denote this problem.

Theorem 4.1. *Assume that n is at least a sufficiently large constant and that ϵ is at most a sufficiently small constant. Also assume that there is a constant $\gamma>0$ such that $d^{1-\gamma}\geq\frac{1}{\epsilon^2}\log\frac{n}{\delta}$. Then $R^{\rightarrow}_{\delta}(\ell_p(n,d,M,\epsilon))\geq\Omega\left(n\frac{1}{\epsilon^2}\log\frac{n}{\delta}(\log d+\log M))\right)$ for $p\in\{1,2\}$.*

In the remaining part of this section we prove the above theorem. Since we can amplify the success probability of a protocol by repeating it and taking majority (see Section A), we will assume throughout that δ is at most a sufficiently small constant. We separately obtain the lower bound $\Omega(n\frac{1}{\epsilon^2}\log\frac{n}{\delta}\log d)$ when the alphabet M is small (Lemma 4.1) and the lower bound $\Omega(n\frac{1}{\epsilon^2}\log\frac{n}{\delta}\log M)$ when the alphabet is large (Lemma 4.3). It is easy to verify that together these lemmas imply Theorem 4.1.

Lower Bound for Small Alphabet Size. We consider the problem with $M=1$ and prove the following.

Lemma 4.1. *Assume that n is at least a sufficiently large constant, δ is at most a sufficiently small constant and $\epsilon\leq 1/25$. Also assume that there is a constant $\gamma>0$ such that $d^{1-\gamma}\geq\frac{1}{\epsilon^2}\log\frac{n}{\delta}$. Then $R^{\rightarrow}_{\delta}(\ell_p(n,d,1,\epsilon))\geq\Omega\left(n\frac{1}{\epsilon^2}\log\frac{n}{\delta}\log d)\right)$ for $p\in\{1,2\}$.*

To prove this lemma, we show how to use the n-fold ℓ_p approximation problem $\ell_p(n,d,1,\epsilon/25)$ to solve the indexing problem $\mathrm{Ind}(n\mathrm{SetInd}(2\epsilon,\delta),c\log d)$, for some constant c. The main component of the reduction is the following lemma, which is a special case of Lemma 3.1 in [23]; although in [23] the authors present the lemma for instances of the problem $\mathrm{Ind}^a(k,N)$, the equivalence between this problem and $\mathrm{SetInd}(\epsilon,\delta)$ directly gives the following.

Lemma 4.2. ([23]) *Given $\epsilon,\eta\in(0,1]$, consider subsets $S^1,S^2,\ldots,S^r$ of $[1/(4\epsilon^2\eta)]$, each of size $1/4\epsilon^2$ (assumed to be odd). Also consider an index $j\in[r]$ and an element $k\in[1/(4\epsilon^2\eta)]$ and let S' be the set consisting of all the elements of S^j that are smaller than k. Then there is an encoding of these objects, based on a random variable R, into vectors $\mathbf{u}=\mathbf{u}(S^1,S^2,\ldots,S^r,R)$ and $\mathbf{v}=\mathbf{v}(S^1,S^2,\ldots,S^{j-1},S',j,k,R)$ with the following properties:*

1. *The vectors $\mathbf{u}$ and $\mathbf{v}$ belong to $\{0,1\}^{d'}$, where $d'=O(10^r\frac{1}{\epsilon^2}\log\frac{1}{\eta})$.*
2. *If k does not belong to the set S^j, then with probability at least $1-\eta$ we have $\|\mathbf{u}-\mathbf{v}\|_p^p\geq d'10^{-j+1}(\frac{1}{2}-\frac{3\epsilon}{10})$ for all $p>0$.*
3. *If k belongs to the set S^j, then with probability at least $1-\eta$ we have $\|\mathbf{u}-\mathbf{v}\|_p^p\leq d'10^{-j+1}(\frac{1}{2}-\frac{6\epsilon}{10})$ for all $p>0$*

Using this lemma, the reduction of $\mathrm{Ind}(n\mathrm{SetInd}(2\epsilon,\delta),c\log d)$ (for some constant c to be determined) to $\ell_p(n,d,1,\epsilon/25)$ (where Alice and Bob have shared randomness) is straightforward. Let Alice's instance for $\mathrm{Ind}(n\mathrm{SetInd}(2\epsilon,\delta),c\log d)$ be given by the sets $\{S_i^\ell\}_{i\in[n],\ell\in[c\log d]}$, where for a fixed ℓ the sets $S_1^\ell,S_2^\ell,\ldots,S_n^\ell$ correspond to the ℓ'th copy of the n-fold problem in the indexing of $\mathrm{Ind}(n\mathrm{SetInd}(2\epsilon,\delta),c\log d)$; unraveling the definition of the problem, we get that each S_i^ℓ is a subset of $[\frac{n}{4\epsilon^2\delta}]$ of size $1/4\epsilon^2$. Similarly, let Bob's instance be given by the index $j\in[c\log d]$, the elements $k_1,k_2,\ldots,k_n$, the sets $\{S_i^\ell\}_{i\in[n],\ell<j}$ and the sets $S_1',S_2',\ldots,S_n'$; again unraveling the definitions we have that for all i the set S_i' consists of all the elements of S_i^j less than k_i. The players want to decide whether $k_i\in S_i^j$ holds or not for all i.

For that, they evoke Lemma 4.2 with $\eta=\delta/n$ and use their inputs and shared randomness to make Alice compute $\mathbf{u}_i=\mathbf{u}_i(S_i^1,S_i^2,\ldots,S_i^{c\log d},R)$ for each i, and make Bob compute $\mathbf{v}_i=\mathbf{v}_i(S_i^1,S_i^2,\ldots,S_i^{j-1},S_i',j,k,R)$ for each i. Notice that these vectors have $O(d^c\frac{1}{\epsilon^2}\log\frac{n}{\delta})$ coordinates, so we can use the fact $d^{1-\gamma}\geq\frac{1}{\epsilon^2}\log\frac{n}{\delta}$ to set the constant c to be small enough (depending on

γ) so that these vectors have at most d coordinates. Then Alice and Bob use a protocol for $\ell_p(n, d, 2, \epsilon/25)$ to obtain with probability $1-\delta$ an approximation $val_i = (1 \pm \frac{\epsilon}{10})\|\mathbf{u}_i - \mathbf{v}_i\|_p^p$ for all i. Based on Items 2 and 3 of Lemma 4.2, Bob then outputs that k_i belongs to S_i^j iff $val_i \leq d10^{-j+1}(\frac{1}{2} - \frac{5\epsilon}{10})$.

It is easy to see that whenever both the guarantees of Lemma 4.2 hold for all n pairs $\{(\mathbf{u}_i, \mathbf{v}_i)\}_{i=1}^n$ and the protocol for $\ell_p(n, d, 1, \epsilon/25)$ succeeds, then Bob outputs the correct answer. Since this happens with probability at least $1 - 2\delta$, we obtain the lower bound $R_\delta^{\rightarrow}(\ell_p(n, d, 1, \epsilon/25)) \geq R_{2\delta}^{\rightarrow,\text{pub}}(\text{Ind}(n\text{SetInd}(2\epsilon, \delta), c \log d)))$, where shared randomness is allowed.

A well-know result relates the randomized complexity of private-randomness and shared-randomness protocols (using the assumption that δ is sufficiently small) [27]:

$$R_{4\delta}^{\rightarrow}(f) \leq R_{2\delta}^{\rightarrow,\text{pub}}(f) + O(\log I + \log \tfrac{1}{\delta}), \tag{4.4}$$

where I denotes the bit size of the input. Using this bound and employing our lower bound on $R_{4\delta}^{\rightarrow}(\text{Ind}(n\text{SetInd}(2\epsilon, \delta), c \log d))$ given by Corollary 4.3, we obtain that

$$\begin{aligned}
&R_{2\delta}^{\rightarrow,\text{pub}}(\text{Ind}(n\text{SetInd}(2\epsilon, \delta), c \log d)) \\
&\geq R_{4\delta}^{\rightarrow}(\text{Ind}(n\text{SetInd}(2\epsilon, \delta), c \log d)) - O\left(\log\left(\frac{n}{\epsilon\delta} + \log d\right)\right) \\
&\geq \Omega(n \tfrac{1}{\epsilon^2} \log \tfrac{n}{\delta} \log d),
\end{aligned}$$

where the last inequality uses the fact that n is at least a sufficiently large constant. This concludes the proof of Lemma 4.1.

Lower Bound for Large Alphabet Size. In this part we prove the following.

LEMMA 4.3. *Assume that n is at least a sufficiently large constant, δ is at most a sufficiently small constant and $\epsilon \leq 1/75$. Also assume that $d \geq \Omega(\frac{1}{\epsilon^2} \log \frac{n}{\delta})$ and that there is a constant $\gamma > 0$ such that $M^{1-\gamma} \geq \frac{d}{\epsilon^3} \log \frac{n}{\delta}$. Then $R_\delta^{\rightarrow}(\ell_p(n, d, M, \epsilon)) \geq \Omega\left(\frac{1}{\epsilon^2} n \log n \log M)\right)$ for $p \in \{1, 2\}$.*

For that, we need two specific statements of JL-type transforms.

THEOREM 4.2. *[1] Let V be an arbitrary set of n points in $\mathbb{R}^d$ and consider $k \geq C\frac{1}{\epsilon^2} \log \frac{n}{\delta}$ for some sufficiently large constant C. Let S be a $k \times d$ matrix with entries picked independently uniformly from $\{-1/\sqrt{k}, 1/\sqrt{k}\}$. Then with probability at least $1-\delta$ we have $\|Su - Sv\|_2^2 = (1 \pm \epsilon)\|u - v\|_2^2$ for all $u, v \in V$.*

LEMMA 4.4. *($\ell_2 \to \ell_1$ JL) Let V be an arbitrary set of n points in $\mathbb{R}^d$ and consider $k \geq C\frac{1}{\epsilon^2} \log \frac{n}{\delta}$ for some sufficiently large constant C. Let S be a $k \times d$ matrix with entries picked independently uniformly from the centered normal distribution with standard deviation $1/k$. Then with probability at least $1 - \delta$ we have $\|Su - Sv\|_1 = (1 \pm \epsilon)\|u - v\|_2$ for all $u, v \in V$.*

Proof. [Proof sketch] This result is essentially proved in [29]. More precisely, consider a vector $x \in \mathbb{R}^d$ with $\|x\| = 1$ and define $Y_i = kS_i x$, where S_i is the i-th row of S. By 2-stability of the normal distribution, Y_i is also normal with variance 1. The proof then follows exactly as in the proof of Theorem 5.1 of [29]. □

Again the lower bound is proved using a reduction from the indexing problem $\text{Ind}(n\text{SetInd}(\epsilon, \delta), r)$, but now with r set to $c \log M$, for some constant c to be determined later. Indeed, we simply modify the reduction above as follows, starting with the ℓ_2 case. Assume for now that the players can use shared randomness. As before, the players evoke Lemma 4.2 with $\eta = \delta/2n$ and make Alice compute $\mathbf{u}_i = \mathbf{u}_i(S_i^1, S_i^2, \ldots, S_i^{c \log M}, R)$ for each i, and make Bob compute $\mathbf{v}_i = \mathbf{v}_i(S_i^1, S_i^2, \ldots, S_i^{j-1}, S_i', j, k, R)$ for each i. These vectors have $O(M^c \frac{1}{\epsilon^2} \log n)$ coordinates, which is $O(M)$ for small enough c by our assumption that $M^{1-\gamma} \geq \frac{d}{\epsilon^3} \log \frac{n}{\delta}$. Now the players use their shared randomness to apply the JL transform from Theorem 4.2 and obtain the vectors $\{\mathbf{u}_i'\}_i$ and $\{\mathbf{v}_i'\}_i$ satisfying the following: (i) with probability at least $1 - \delta/2$ we have $\|\mathbf{u}_i' - \mathbf{v}_i'\|_2^2 = (1 \pm \frac{\epsilon}{20})\|\mathbf{u}_i - \mathbf{v}_i\|_2^2$ for all $i \in [n]$; (ii) the dimension of each of these vectors is $O(\frac{1}{\epsilon^2} \log \frac{n}{\delta})$, which is $O(d)$ due to the assumption $d \geq \Omega(\frac{1}{\epsilon^2} \log \frac{n}{\delta})$; (iii) all entries of these vectors belong to the set $0, \pm 1/\sqrt{k}, \ldots, \pm O(M/\sqrt{k})$.

Then Alice and Bob can use a protocol for $\ell_2(n, O(d), O(M), \epsilon/50)$ that succeeds with probability $1-\delta$ to compute $(1 \pm \frac{\epsilon}{20})$ approximations to the distances $\|\mathbf{u}_i' - \mathbf{v}_i'\|_2^2$ for all i and decide whether k_i belongs to S_i^j or not for every i just as before. It is easy to see that Alice and Bob will report the right answer with probability at least $1 - 2\delta$, and hence $R_\delta^{\rightarrow}(\ell_2(n, O(d), O(M), \epsilon/50)) \geq R_{2\delta}^{\rightarrow,\text{pub}}(\text{Ind}(n\text{SetInd}(2\epsilon, \delta), c \log M)))$. Again using (4.4) and Corollary 4.3 concludes the proof of Lemma 4.3 for the case ℓ_2.

For the case of ℓ_1 distance again the players evoke Lemma 4.2 with $\eta = \delta/2n$ and make Alice compute $\mathbf{u}_i = \mathbf{u}_i(S_i^1, S_i^2, \ldots, S_i^{c \log M}, R)$ for each i, and make Bob compute $\mathbf{v}_i = \mathbf{v}_i(S_i^1, S_i^2, \ldots, S_i^{j-1}, S_i', j, k, R)$ for each i. Again that these vectors have $O(M^c \frac{1}{\epsilon^2} \log n) = O(\epsilon M/d)$ coordinates for small enough c (due to our assumption on M). Now for each i they use their shared randomness to obtain a matrix S with $d' = O(\frac{1}{\epsilon^2} \log \frac{n}{\delta}) = O(d)$ columns satisfying the guarantees

from Lemma 4.4 (with approximation factor $(1\pm\frac{\epsilon}{75})$ and success probability $1-\delta$). Then for all i Alice computes the vector $\tilde{\mathbf{u}}_i$ by taking $S\mathbf{u}_i$ and rounding each entry to the closest additive multiple of $\epsilon/75d'$, and Bob can compute $\tilde{\mathbf{v}}_i$ similarly. One can then verify that with probability $1-\delta$ we have $\|\tilde{\mathbf{u}}_i-\tilde{\mathbf{v}}_i\|_1 = (1\pm\frac{2\epsilon}{75})\|\mathbf{u}_i-\mathbf{v}_i\|_2$ (see for instance Section C.1). Then Alice checks if $\|\tilde{\mathbf{u}}_i\|_\infty \leq 2\|\mathbf{u}_i\|_2^2$ (which is $O(\epsilon M/d)$) for all i; if so, she and Bob use a protocol for $\ell_1(n, O(d), O(M), \epsilon/75)$ to compute $(1\pm\epsilon/75)\|\tilde{\mathbf{u}}_i-\tilde{\mathbf{v}}_i\|_1$ for all i with probability $1-\delta$. It is easy to see that with probability at least $1-2\delta$ Alice and Bob compute an approximation $(1\pm\frac{\epsilon}{10})\|\mathbf{u}_i-\mathbf{v}_i\|_2^2$ for all i, which can be used as before to solve their instance of $\mathrm{Ind}(n\mathrm{SetInd}(2\epsilon,\delta), c\log M))$. The proof of the lemma then follows just as in the ℓ_2 case.

4.3 Other Applications The proof of the lower bound for the remaining applications is similar in spirit to that of Theorem 4.1, and are presented in Section C of the appendix.

JL Transforms. The main result of this section is an optimal lower bound on the dimension of a JL transform.

Definition 4.1. *A family $\mathcal{F}$ of $k\times d$ matrices together with a distribution μ on $\mathcal{F}$ forms a* Johnson-Lindenstrauss transform *with parameters ϵ,δ,n,d (or JLT(ϵ,δ,n,d) for short), if the following holds for $S\sim\mu$: for any set V of n vectors in $\mathbb{R}^d$, for all $\mathbf{u},\mathbf{v}\in V$ we have $(1-\epsilon)\|\mathbf{u}-\mathbf{v}\|^2 \leq \|S\mathbf{u}-S\mathbf{v}\|^2 \leq (1+\epsilon)\|\mathbf{u}-\mathbf{v}\|^2$ with probability at least $1-\delta$. We say that k is the* dimension *of the transform.*

Theorem 4.3. *Assume that n is at least a sufficiently large constant and that ϵ is at most a sufficiently small constant. Also assume that there is a constant $\gamma>0$ such that $d^{1-\gamma}\geq\frac{1}{\epsilon^2}\log\frac{n}{\delta}$. Then any JLT$(\epsilon,\delta,n,d)$ has dimension at least $\Omega(\frac{1}{\epsilon^2}\log\frac{n}{\delta})$. Moreover, this holds even if the guarantees of the transform only need to hold for vectors in $\{-1,0,1\}^d$.*

Sketching Multiple Inner Products. Consider the following communication problem: Alice has n vectors $\mathbf{u}^1,\mathbf{u}^2,\ldots,\mathbf{u}^n\in[\pm M]^d$ and Bob has n vectors $\mathbf{v}^1,\mathbf{v}^2,\ldots,\mathbf{v}^n\in[\pm M]^d$. Alice needs to send sketches $S\mathbf{u}^1,S\mathbf{u}^2,\ldots,S\mathbf{u}^n$ of her vectors to Bob, who then has to output (with probability at least $1-\delta$) approximations $\langle\mathbf{u}^i,\mathbf{v}^i\rangle\pm\epsilon\|\mathbf{u}^i\|\|\mathbf{v}^i\|$ for *all* $i\in[n]$. Let $\mathrm{Ip}(n,d,M,\epsilon)$ denote this problem.

Theorem 4.4. *Assume that n is at least a sufficiently large constant and that ϵ is at most a sufficiently small constant. Also assume that there is a constant $\gamma>0$ such that $(d\log M)^{1-\gamma}\geq\frac{1}{\epsilon^2}\log\frac{n}{\delta}$. Then $R_\delta^{sketch}(\mathrm{Ip}(n,d,M,\epsilon))\geq\Omega\left(n\frac{1}{\epsilon^2}\log\frac{n}{\delta}(\log d+\log M)\right)$.*

Notice that the above lower bound requires the protocol to be a sketching one: otherwise one can apply a JL transform to reduce the dimension and use ℓ_2 sampling to solve $\mathrm{Ip}(n,d,M,\epsilon)$ with communication $\tilde{O}(n\frac{1}{\epsilon^2}\log\frac{n}{\delta}\log\log_{1+\epsilon}M)$ [31, 11].

Matrix Sketching. Given a matrix A, we use A_i to denotes its i-th row and use A^j to denote its j-th column.

Theorem 4.5. *Assume that n is a sufficiently large constant and that ϵ is at most a sufficiently small constant. Also assume that there is a constant $\gamma>0$ such that $n^{1-\gamma}\geq\frac{1}{\epsilon^2}\log\frac{n}{\delta}$. Let S be a random $n\times k$ matrix which has an estimation procedure f outputting a matrix satisfying the following: for every pair of matrices $A,B\in[\pm M]^{n\times n}$, with probability at least $1-\delta$ we have $f(AS,B)_{i,j}=(AB)_{i,j}\pm\epsilon\|A_i\|\|B^j\|$ for all $i,j\in[n]$. Then the bit size of AS is at least $\Omega(n\frac{1}{\epsilon^2}\log\frac{n}{\delta}(\log n+\log M))$. Moreover, if the estimation is given by $f(AS,B)_{i,j}=(ASS^TB)_{i,j}$, then the dimension k is at least $\Omega(\frac{1}{\epsilon^2}\log\frac{n}{\delta})$.*

Database Joins. We refer the reader to [4] for more details about this application. Consider a database consisting of n tables and multiple attributes, with value domain $\mathcal{D}$. Let M denote the maximum number of records over all these tables. Given attribute j in table i, we use $f_i^j(d)$ to denote the number of records in table i whose value for attribute j is equal to d. We see f_i^j as a vector in $\{0,1,\ldots,M\}^{|\mathcal{D}|}$. Given attribute j in table i and attribute j' in table i', the *join size* of these attributes is gives by the inner product $\langle f_i^j, f_{i'}^{j'}\rangle$. For simplicity, we assume that there is only one attribute j_i in each table i that we are interested in estimating join sizes. We have the following bounds for estimating these join sizes.

Theorem 4.6. *Assume that n is at least a sufficiently large constant and that ϵ is at most a sufficiently small constant. Consider linear sketches of the n frequency vectors $f_i^{j_i}$ which allow the join size estimation $\langle f_i^{j_i}, f_{i'}^{j_{i'}}\rangle\pm\epsilon\|f_i^{j_i}\|\|f_{i'}^{j_{i'}}\|$ for all $i,i'\in[n]$ with probability at least δ. Then we have the following lower bounds for the total bit size required by these sketches:*

- *If there is a constant $\gamma>0$ such that $|\mathcal{D}|^{1-\gamma}\geq\frac{1}{\epsilon^2}\log\frac{n}{\delta}$, then we have the lower bound $\Omega(n\frac{1}{\epsilon^2}\log\frac{n}{\delta}\log|\mathcal{D}|)$.*
- *If $d\geq\frac{n}{\epsilon^2\delta}$ and M is at least a sufficiently large constant, then we have the lower bound $\Omega(n\frac{1}{\epsilon^2}\log\frac{n}{\delta}\log M)$.*

As mentioned earlier, the bounds above actually lower bound the total size of computing a mergeable summary for the n tables.

References

[1] D. Achlioptas. Database-friendly random projections: Johnson-lindenstrauss with binary coins. *Journal of Computer and System Sciences*, 66(4):671 – 687, 2003.

[2] P. K. Agarwal, G. Cormode, Z. Huang, J. M. Phillips, Z. Wei, and K. Yi. Mergeable summaries. In *PODS*, pages 23–34, 2012.

[3] N. Alon. Perturbed identity matrices have high rank: Proof and applications. *Combinatorics, Probability & Computing*, 18(1-2):3–15, 2009.

[4] N. Alon, P. B. Gibbons, Y. Matias, and M. Szegedy. Tracking join and self-join sizes in limited storage. *J. Comput. Syst. Sci.*, 64(3):719–747, 2002.

[5] N. Alon, Y. Matias, and M. Szegedy. The space complexity of approximating the frequency moments. *J. Comput. Syst. Sci.*, 58(1):137–147, 1999.

[6] Z. Bar-Yossef, T. S. Jayram, R. Krauthgamer, and R. Kumar. The sketching complexity of pattern matching. In *APPROX-RANDOM*, pages 261–272, 2004.

[7] Z. Bar-Yossef, T. S. Jayram, R. Kumar, and D. Sivakumar. An information statistics approach to data stream and communication complexity. *J. Comput. Syst. Sci.*, 68(4):702–732, 2004.

[8] B. Barak, M. Braverman, X. Chen, and A. Rao. How to compress interactive communication. In *STOC*, pages 67–76, 2010.

[9] M. Braverman and A. Rao. Information equals amortized communication. In *FOCS*, pages 748–757, 2011.

[10] A. Chakrabarti, Y. Shi, A. Wirth, and A. C.-C. Yao. Informational complexity and the direct sum problem for simultaneous message complexity. In *FOCS*, pages 270–278, 2001.

[11] K. L. Clarkson, E. Hazan, and D. P. Woodruff. Sublinear optimization for machine learning. In *Proceedings of the 2010 IEEE 51st Annual Symposium on Foundations of Computer Science*, FOCS '10, pages 449–457, Washington, DC, USA, 2010. IEEE Computer Society.

[12] K. L. Clarkson and D. P. Woodruff. Numerical linear algebra in the streaming model. In *STOC*, pages 205–214, 2009.

[13] T. M. Cover and J. A. Thomas. *Elements of information theory (2. ed.)*. Wiley, 2006.

[14] D. P. Dubhashi and A. Panconesi. *Concentration of Measure for the Analysis of Randomized Algorithms*. Cambridge University Press, 2009.

[15] T. Feder, E. Kushilevitz, M. Naor, and N. Nisan. Amortized communication complexity. *SIAM J. Comput.*, 24(4):736–750, 1995.

[16] P. Harsha, R. Jain, D. A. McAllester, and J. Radhakrishnan. The communication complexity of correlation. *IEEE Transactions on Information Theory*, 56(1):438–449, 2010.

[17] P. Indyk. Stable distributions, pseudorandom generators, embeddings, and data stream computation. *J. ACM*, 53(3):307–323, May 2006.

[18] P. Indyk. Sketching, streaming and sublinear-space algorithms. 2007. Graduate course notes, available at `http://stellar.mit.edu/S/course/6/fa07/6.895/`.

[19] R. Jain. New strong direct product results in communication complexity. *Electronic Colloquium on Computational Complexity (ECCC)*, 18:24, 2011.

[20] R. Jain, A. Pereszlényi, and P. Yao. A direct product theorem for bounded-round public-coin randomized communication complexity. *CoRR*, abs/1201.1666, 2012.

[21] R. Jain, J. Radhakrishnan, and P. Sen. A direct sum theorem in communication complexity via message compression. In *ICALP*, pages 300–315, 2003.

[22] R. Jain, P. Sen, and J. Radhakrishnan. Optimal direct sum and privacy trade-off results for quantum and classical communication complexity. *CoRR*, abs/0807.1267, 2008.

[23] T. S. Jayram and D. P. Woodruff. Optimal bounds for johnson-lindenstrauss transforms and streaming problems with sub-constant error. In D. Randall, editor, *SODA*, pages 1–10. SIAM, 2011.

[24] H. Jowhari, M. Saglam, and G. Tardos. Tight bounds for lp samplers, finding duplicates in streams, and related problems. In *PODS*, pages 49–58, 2011.

[25] D. M. Kane, J. Nelson, and D. P. Woodruff. On the exact space complexity of sketching and streaming small norms. In *SODA*, pages 1161–1178, 2010.

[26] H. Klauck. A strong direct product theorem for disjointness. In *STOC*, pages 77–86, 2010.

[27] I. Kremer, N. Nisan, and D. Ron. On randomized one-round communication complexity. *Computational Complexity*, pages 21–49, 1999.

[28] E. Kushilevitz and N. Nisan. *Communication complexity*. Cambridge University Press, 1997.

[29] J. Matousek. On variants of the johnson-lindenstrauss lemma. *Random Structures and Algorithms*, 33(2):142–156, 2008.

[30] A. McGregor. Data streams and linear sketches. 2007. STOC Workshop presentation, available at `http://people.cs.umass.edu/~mcgregor/stocworkshop/mcgregor.pdf`.

[31] M. Monemizadeh and D. P. Woodruff. 1-pass relative-error l_p-sampling with applications. In *SODA*, pages 1143–1160, 2010.

[32] S. Muthukrishnan. Data streams: algorithms and applications. *Found. Trends Theor. Comput. Sci.*, 1(2):117–236, Aug. 2005.

[33] R. Pagh. Compressed matrix multiplication. In *ITCS*, pages 442–451, 2012.

[34] T. Sarlós. Improved approximation algorithms for large matrices via random projections. In *FOCS*, pages 143–152, 2006.

[35] R. Shaltiel. Towards proving strong direct product theorems. *Computational Complexity*, 12(1-2):1–22, 2003.

A Information Cost When Amplifying Success Probability

Consider a function $f : \mathcal{X} \times \mathcal{Y} \to \mathcal{Z}$ and let λ be a distribution over $\mathcal{X} \times \mathcal{Y} \times \mathcal{D}$, with marginals μ over $\mathcal{X} \times \mathcal{Y}$ and ν over $\mathcal{D}$. We show that $\mathrm{IC}_{\mu,\delta^{\Omega(r)}}(f|\nu) \leq r\,\mathrm{IC}_{\mu,\delta}(f|\nu)$. For that, take a δ-protocol Π for f which achieves $\mathrm{I}(\Pi; X, Y \mid D) = \mathrm{IC}_{\mu,\delta}(f|\nu)$, where $(X, Y, D) \sim \lambda$. Then let $\bar{\Pi}$ be the protocol on input (x, y) that runs r copies $\Pi(x, y, R_1), \Pi(x, y, R_2), \ldots, \Pi(x, y, R_r)$ with independent coins $R_1, R_2, \ldots, R_r$ and outputs the value obtained by the majority of the runs.

It is easy to see that $\bar{\Pi}$ outputs the correct answer with probability at least $1 - \delta^{\Omega(r)}$. Moreover, by the chain rule for mutual information, we have

$$\mathrm{IC}_{\mu,\delta^{\Omega(r)}}(f|\nu) \leq \mathrm{I}(\bar{\Pi}; X, Y \mid D) = \sum_{i=1}^{r} \mathrm{I}\left[\Pi(X, Y, R_i); X, Y \mid D, \Pi(X, Y, R_1), \ldots, \Pi(X, Y, R_{i-1})\right]. \tag{A.1}$$

But we can expand the i-th term as

$$\begin{aligned}
&\mathrm{H}\left[\Pi(X, Y, R_i) \mid D, \Pi(X, Y, R_1), \ldots, \Pi(X, Y, R_{i-1})\right] \\
&\quad - \mathrm{H}\left[\Pi(X, Y, R_i) \mid D, \Pi(X, Y, R_1), \ldots, \Pi(X, Y, R_{i-1}), X, Y\right] \\
&\leq \mathrm{H}\left[\Pi(X, Y, R_i) \mid D\right] \\
&\quad - \mathrm{H}\left[\Pi(X, Y, R_i) \mid D, \Pi(X, Y, R_1), \ldots, \Pi(X, Y, R_{i-1}), X, Y\right] \\
&= \mathrm{H}[\Pi(X, Y, R_i) \mid D] - \mathrm{H}[\Pi(X, Y, R_i) \mid D, X, Y] \\
&= \mathrm{I}[\Pi(X, Y, R_i); X, Y \mid D] = \mathrm{IC}_{\mu,\delta}(f|\nu),
\end{aligned}$$

where the first equality follows from the fact that, since the R_j's are independent, then conditioned on (X, Y) we have $\Pi(X, Y, R_i)$ independent from $\Pi(X, Y, R_1) \ldots, \Pi(X, Y, R_{i-1})$. Plugging this bound on equation (A.1) gives that $\mathrm{IC}_{\mu,\delta^{\Omega(r)}}(f|\nu) \leq r\,\mathrm{IC}_{\mu,\delta}(f|\nu)$.

B Auxiliary Results for Lower Bounding Applications

Before proving the lower bound for our applications, we need to spell out some (standard) tools. In the next two subsections, we introduce the hard communication problem from there the lower bounds will come from. This hard problem is essentially based on constructing the n-fold version of augmented indexing and then doing an extra indexing over it. In the following subsection, we present, for completeness, an encoding of augmented indexing into vectors whose inner product depends whether the instance is yes/no; this was already present in the proof of Lemma 3.1 of [23].

B.1 Generic Indexing problems

A generic indexing problem can be defined as follows. Consider a function $f : \mathcal{X} \times \mathcal{Y} \to \mathcal{Z}$ and the associated (one-way) communication problem where Alice and Bob get respectively an element of $\mathcal{X}$ and $\mathcal{Y}$ and want to compute the value of the f over this pair; we use f to also denote this problem. Let $\mathrm{Ind}(f, N)$ denote the communication problem where Alice has input $x_1, x_2, \ldots x_N \in \mathcal{X}$ and Bob has input $j \in [N]$, $x_1, x_2, \ldots, x_{j-1} \in X$ and $y \in \mathcal{Y}$, and they want to compute $f(x_j, y)$. To simplify the notation, let $\tilde{\mathcal{X}} = \mathcal{X}^N$ denote the space of Alice's input, let $\tilde{\mathcal{Y}} = \bigcup_{i=0}^{N-1}(\mathbb{N} \times \mathcal{X}^i \times \mathcal{Y})$ denote the space of Bob's input.

It is folklore that the information complexity of an indexing problem $\mathrm{Ind}(f, N)$ is typically $\Omega(N)$ times the complexity of the base problem of computing f.

Lemma B.1. *Let λ be a probability distribution over $\mathcal{X} \times \mathcal{Y} \times \mathcal{D}$ with marginal μ on $\mathcal{X} \times \mathcal{Y}$ and marginal ν on $\mathcal{D}$, such that μ is partitioned by ν. Then there is a distribution $\tilde{\lambda}$ over $\tilde{\mathcal{X}} \times \tilde{\mathcal{Y}} \times \tilde{\mathcal{D}}$ (where $\tilde{\mathcal{D}} = \mathbb{N} \times \mathcal{D}^N$) with the following property. Let $\tilde{\mu}$ denote the marginal of $\tilde{\lambda}$ on $\tilde{\mathcal{X}} \times \tilde{\mathcal{Y}}$ and let $\tilde{\nu}$ denote the marginal of $\tilde{\lambda}$ on $\tilde{\mathcal{D}}$. Then for all $\delta \in [0, 1]$*

$$\mathrm{IC}^{\rightarrow}_{\tilde{\mu},\delta}(\mathrm{Ind}(f, N)|\tilde{\nu}) \geq N \cdot \mathrm{IC}^{\rightarrow}_{\mu,\delta}(f|\nu).$$

Moreover, $\tilde{\nu}$ partitions $\tilde{\mu}$.

To make the presentation self-contained, in the remaining part of this section we prove the above lemma. For that, we start by constructing the distribution $\tilde{\lambda}$. First let J be the uniform random variable over $[N]$, and, to makes things formal, let $D_0 = J$. Now let the random variables $\mathbf{X} = (X_1, X_2, \ldots, X_N)$, Y and $\mathbf{D} = (D_1, D_2, \ldots, D_N)$ have the following distribution: conditioned on $J = j$, we have $(X_j, Y, D_j) \sim \lambda$ and (X_i, D_i) distributed according to the marginal of λ on $\mathcal{X} \times \mathcal{D}$ for all $i \neq j$ (so the conditioning on J only specifies which variables will be correlated with Y). The distribution $\tilde{\lambda}$ is then defined as the distribution of the random variable $(\mathbf{X}, (J, \mathbf{X}_{<J}, Y), (D_0, \mathbf{D}))$. It is easy to see that $\tilde{\nu}$ partitions $\tilde{\mu}$.

Recall the notation for one-way protocols used in Section 3. Consider a private-randomness one-way δ-protocol (M, B) for $\mathrm{Ind}(f, N)$ (with Alice's and Bob's private coins respectively denoted by R^A and R^B) and attains $\mathrm{IC}^{\rightarrow}_{\tilde{\mu},\delta}(\mathrm{Ind}(f, N)|\tilde{\nu})$; that is, for the random variables above, we have $\mathrm{I}(M(\mathbf{X}, R^A); \mathbf{X} \mid D_0\mathbf{D}) = \mathrm{IC}^{\rightarrow}_{\tilde{\mu},\delta}(\mathrm{Ind}(f, N)|\tilde{\nu})$. We start lower bounding the left-hand side.

First notice that the random variable $(\mathbf{X}, R^A, \mathbf{D})$ is

independent of D_0. Therefore, we have

$$\begin{aligned}
&\mathrm{I}(M(\mathbf{X}, R^A); \mathbf{X} \mid D_0\mathbf{D}) = \mathrm{I}(M(\mathbf{X}, R^A); \mathbf{X} \mid \mathbf{D}) \\
&= \sum_{j=1}^{N} \mathrm{I}(M(\mathbf{X}, R^A); X_j \mid \mathbf{D}, \mathbf{X}_{<j}) \\
&= \sum_{j=1}^{N} \mathrm{I}(M(\mathbf{X}, R^A); X_j \mid \mathbf{D}_{\geq j}, \mathbf{X}_{<j}) \\
&= \sum_{j=1}^{N} \sum_{\mathbf{d}_{>j}, \mathbf{x}_{<j}} \mathrm{I}(M(\mathbf{X}, R^A); X_j \mid D_j, \mathbf{D}_{>j} = \mathbf{d}_{>j}, \mathbf{X}_{<j} = \mathbf{x}_{<j}) \cdot \\
&\qquad \Pr(\mathbf{D}_{>j} = \mathbf{d}_{>j}, \mathbf{X}_{<j} = \mathbf{x}_{<j}) \\
&= \sum_{j=1}^{N} \sum_{\mathbf{d}_{>j}, \mathbf{x}_{<j}} \mathrm{I}(M(\mathbf{x}_{<j}\mathbf{X}_{\geq j}, R^A); X_j \mid D_j, \mathbf{D}_{>j} = \mathbf{d}_{>j}) \cdot \\
&\qquad \Pr(\mathbf{D}_{>j} = \mathbf{d}_{>j}, \mathbf{X}_{<j} = \mathbf{x}_{<j}),
\end{aligned} \tag{B.2}$$

where the second equality follows from the chain rule for conditional mutual information, and the others follows from the product structure of $\mathbf{D}$ and $\mathbf{X}$ and independence from R^A. Now we lower bound each term in this expression using a standard simulation argument.

Claim 1. *For every index $j \in [N]$ and fixing $\mathbf{d}_{>j}$ and $\mathbf{x}_{<j}$, there is a private-randomness one-way protocol $(\bar{M}, \bar{B})$ with domain $\mathcal{X} \times \mathcal{Y}$ satisfying the following (where $\bar{R}^A$ and $\bar{R}^B$ denote Alice's and Bob's private coins respectively):*

- *$(\bar{M}, \bar{B})$ is a δ-protocol for f.*
- *For the random variable $(\bar{X}_j, \bar{Y}, \bar{D}_j) \sim \lambda$, we have $\mathrm{I}(\bar{M}(\bar{X}_j, \bar{R}^A); \bar{X}_j \mid \bar{D}_j) = \mathrm{I}(M(\mathbf{x}_{<j}\mathbf{X}_{\geq j}, R^A); X_j \mid D_j, \mathbf{D}_{>j} = \mathbf{d}_{>j})$.*

Proof. The desired protocol $(\bar{M}, \bar{B})$ is the following. Alice uses her private randomness $\bar{R}^A$ to obtain the random variable $\tilde{R}^A$ with the same distribution as R^A, and also the random variable $\tilde{\mathbf{X}}_{>j}$ with the same distribution as the conditioned random variable $\mathbf{X}_{>j} \mid (\mathbf{D}_{>j} = \mathbf{d}_{>j})$; Bob uses his private randomness $\bar{R}^B$ to obtain the random variable $\tilde{R}^B$ with same distribution as R^B. Then for every input $(x, y) \in \mathcal{X} \times \mathcal{Y}$, we set Alice's message to be $\bar{M}(x, \bar{R}^A) = M(\mathbf{x}_{<j}x\tilde{\mathbf{X}}_{>j}, \tilde{R}^A)$ and Bob's output upon receiving message m is $\bar{B}(m, y, \bar{R}^B) = B(m, j, \mathbf{x}_{<j}, y, \tilde{R}^B)$.

For every input $(x, y) \in \mathcal{X} \times \mathcal{Y}$, we can use the fact (M, B) is a δ-protocol for $\mathrm{Ind}(f, N)$ to obtain that

$$\begin{aligned}
1 - \delta &\leq \Pr\left[B(M(\mathbf{x}_{<j}x\tilde{\mathbf{X}}_{>j}, R^A), j, \mathbf{x}_{<j}, y, R^B) = f(x, y)\right] \\
&= \Pr((\bar{M}, \bar{B}) \text{ outputs } f(x, y)),
\end{aligned}$$

where the equality follows from the definition of our random variables. This gives the first part of the claim.

For the second part, let $(\bar{X}_j, \bar{Y}_j, \bar{D}_j) \sim \lambda$. By the definition of our random variables, $(\bar{X}_j\tilde{\mathbf{X}}_{>j}, \bar{D}_j, \tilde{R}^A)$ has the same distribution as $(\mathbf{X}_{\geq j}, D_j, R^A) \mid (\mathbf{D}_{>j} = \mathbf{d}_{>j})$, so by substitution we have

$$\begin{aligned}
&\mathrm{I}(\bar{M}(\bar{X}_j, \bar{R}^A); \bar{X}_j \mid \bar{D}_j) = \mathrm{I}(M(\mathbf{x}_{<j}\bar{X}_j\tilde{\mathbf{X}}_{>j}, \tilde{R}^A); \bar{X}_j \mid \bar{D}_j) \\
&= \mathrm{I}(M(\mathbf{x}_{<j}\mathbf{X}_{\geq j}, R^A); X_j \mid D_j, \mathbf{D}_{>j} = \mathbf{d}_{>j}),
\end{aligned}$$

which concludes the proof of the claim. □

Lemma B.1 then follows directly from the above claim and equation (B.2).

B.2 Encoding of Indexing Over Augmented Set Indexing In this section we present the following encoding of $\mathrm{Ind}(\mathrm{SetInd}(\epsilon, \eta), r)$, which was already present in the proof of Lemma 3.1 in [23]. Notice that we consider the problem $\mathrm{Ind}(\mathrm{SetInd}(\epsilon, \eta), r)$ and not the n-fold problem $\mathrm{Ind}(n\mathrm{SetInd}(\epsilon, \eta), r)$, but we can use this encoding for each of the n copies present in the latter.

Lemma B.2. *Given $\epsilon, \eta \in (0, 1]$, consider subsets $S^1, S^2, \ldots, S^r$ of $[1/(\epsilon^2\eta)]$, each of size $1/\epsilon^2$ (assumed to be odd). Also consider an index $j \in [r]$ and an element $k \in [1/(\epsilon^2\eta)]$. Then there is an encoding of these objects, based on a random variable R, into vectors $\mathbf{u} = \mathbf{u}(S^1, S^2, \ldots, S^r, R)$, $\underline{\mathbf{u}} = \underline{\mathbf{u}}(S^1, S^2, \ldots, S^{j-1}, R)$ and $\mathbf{v} = \mathbf{v}(j, k, R)$ with the following properties:*

1. *The vectors $\mathbf{u}, \underline{\mathbf{u}}$ and $\mathbf{v}$ belong to $\{0, 1\}^{2t \cdot \frac{10^r - 1}{9}}$, where $t = \frac{72}{\epsilon^2} \log \frac{1}{\eta}$.*
2. *$\|\mathbf{u} - \underline{\mathbf{u}}\|^2 \leq 2 \cdot 10^{r-j}t$ and $\|\mathbf{v}\|^2 = 10^{r-j}t$.*
3. *If k does not belong to the set S^j, then with probability at least $1 - \eta$ we have $\langle \mathbf{u} - \underline{\mathbf{u}}, \mathbf{v} \rangle \leq 10^{r-j}t(\frac{1}{2} + \frac{\epsilon}{12})$.*
4. *If k belongs to the set S^j, then with probability at least $1 - \eta$ we have $\langle \mathbf{u} - \underline{\mathbf{u}}, \mathbf{v} \rangle \geq 10^{r-j}t(\frac{1}{2} + \frac{2\epsilon}{12})$.*

To prove this lemma, we first define and analyze an encoding scheme for the case where we only have one set, i.e., $r = 1$.

So consider a set $S \subseteq [1/(\epsilon^2\eta)]$. Let $\mathbf{X}$ be a uniform random vector in $\{-1, +1\}^{1/(\epsilon^2\eta)}$. We define $\mathrm{enc}_1(S, \mathbf{X})$ to be the majority of the set $\{X_i\}_{i \in S}$; this is well-defined since $1/\epsilon^2$ is odd. We contrast this with the encoding $\mathrm{enc}_2(k, \mathbf{X})$ which is just the k-th component of $\mathbf{X}$.

Notice that if $k \notin S$, then the encodings are independent and hence

$$\Pr[\mathrm{enc}_1(S, \mathbf{X}) = \mathrm{enc}_2(k, \mathbf{X})] = \tfrac{1}{2}.$$

On the other hand, suppose $k \in S$. Then, using the fact that $\mathrm{enc}_1(S, \mathbf{X})$ depends on only $1/\epsilon^2$ coordinates

of $\mathbf{X}$ (since $|S| = 1/\epsilon^2$), standard arguments involving the binomial coefficients give that

$$\Pr[\mathrm{enc}_1(S, \mathbf{X}) = \mathrm{enc}_2(k, \mathbf{X})] \geq \tfrac{1}{2}(1 + \tfrac{\epsilon}{2}).$$

We repeat the above scheme to amplify the gap between the two cases. Let $\mathbb{X} = (\mathbf{X}^1, \mathbf{X}^2, \ldots, \mathbf{X}^t)$ be a collection of $t = \frac{72}{\epsilon^2} \log \frac{1}{\eta}$ uniform i.i.d. random variables in $\{-1, +1\}^{1/(\epsilon^2 \eta)}$. Define

$$\mathrm{enc}_1(S, \mathbb{X}) = (\mathrm{enc}_1(S, \mathbf{X}^1), \mathrm{enc}_1(S, \mathbf{X}^2), \ldots, \mathrm{enc}_1(S, \mathbf{X}^t)),$$

and

$$\mathrm{enc}_2(k, \mathbb{X}) = (\mathrm{enc}_2(k, \mathbf{X}^1), \mathrm{enc}_2(k, \mathbf{X}^2), \ldots, \mathrm{enc}_2(k, \mathbf{X}^t)).$$

Fact B.1. (Chernoff bounds, [14]) *Let $Y_1, Y_2, \ldots, Y_t$ be a collection of i.i.d. 0-1 Bernoulli random variables with success probability p. Set $\bar{Y} = \sum_{i=1}^t Y_i/t$. Then,*

$$\Pr[\bar{X} < p - h] < exp(-2h^2 t), \text{ and}$$
$$\Pr[\bar{X} > p + h] < exp(-2h^2 t).$$

In the above fact with $t = \frac{72}{\epsilon^2} \log \frac{1}{\eta}$ and $h = \epsilon/12$, we obtain that the tail probabilities are at most η. In the case $k \notin S$ we use $p = \frac{1}{2}$ to get

$$\text{(B.3)} \quad \Pr[\#(\mathrm{enc}_1(S, \mathbb{X}), \mathrm{enc}_2(k, \mathbb{X})) > t(\tfrac{1}{2} + \tfrac{\epsilon}{12})] \leq \eta,$$

where # denotes the number of coordinates where the vectors agree. In the case $k \in S$ we use $p = \frac{1}{2}(1 + \frac{\epsilon}{2})$ to get

$$\text{(B.4)} \quad \Pr[\#(\mathrm{enc}_1(S, \mathbb{X}), \mathrm{enc}_2(k, \mathbb{X})) < t(\tfrac{1}{2} + \tfrac{2\epsilon}{12})] \leq \eta.$$

Finally, we convert the ± 1 vector $\mathrm{enc}_1(S, \mathbb{X})$ (resp. $\mathrm{enc}_2(k, \mathbb{X})$) into the 0/1 vector $\mathrm{enc}'_1(S, \mathbb{X})$ (resp. $\mathrm{enc}'_2(k, \mathbb{X})$) by replacing the occurrence of each 1 by the pattern 01 and the occurrence of each -1 by 10; so the new vectors have exactly twice as many coordinates as the original ones. Moreover, $\langle \mathrm{enc}'_1(S, \mathbb{X}), \mathrm{enc}'_2(k, \mathbb{X}) \rangle = \#(\mathrm{enc}_1(S, \mathbb{X}), \mathrm{enc}_2(k, \mathbb{X}))$ and $\|\mathrm{enc}'_1(S, \mathbb{X})\| = \|\mathrm{enc}'_2(k, \mathbb{X}))\| = \sqrt{t}$. Setting $\mathbf{u} = \mathrm{enc}'_1(S, \mathbb{X})$, $\underline{\mathbf{u}} = 0$ and $\mathbf{v} = \mathrm{enc}'_2(k, \mathbb{X}))$ gives the desired encoding for the case where we have only one set S.

Now we adapt this encoding to handle multiple sets. For each $i \in [r]$, define the vector $\mathbf{u}^i \in \{0, 1\}^{10^{r-i} 2t}$ by appending 10^{r-i} copies of $\mathrm{enc}'_1(S^i, \mathbb{X})$. Then define the vector $\mathbf{u}$ by appending the vectors $\mathbf{u}^i$ for $i = 1, 2, \ldots, r$. Also define the vector $\underline{\mathbf{u}}$ by appending the vectors $\mathbf{u}^i$ for $i = 1, 2, \ldots, j-1$ and the appending 0's to obtain a vector with the same number of coordinates as $\mathbf{u}$.

Now for each $i \in [r]$ define the vector $\mathbf{v}^i$ to be equal to $\mathbf{0} \in \{0, 1\}^{10^{r-i} 2t}$ if $i \neq j$, and to be equal to 10^{r-j} copies of $\mathrm{enc}'_2(k, \mathbb{X})$ otherwise. Then define $\mathbf{v}$ by appending the vectors $\mathbf{v}^i$ for $i = 1, 2, \ldots, r$.

It is easy to see that $\mathbf{u}$, $\underline{\mathbf{u}}$ and $\mathbf{v}$ have the desired properties. First, notice that these vectors have exactly $2t \sum_{i=1}^r 10^{r-i} = 2t \cdot \frac{10^r - 1}{9}$ coordinates. Also,

$$\|\mathbf{u} - \underline{\mathbf{u}}\|^2 = \sum_{i=j}^r \|\mathbf{u}^i\|^2 = \sum_{i=j}^r 10^{r-i} t \leq 2 \cdot 10^{r-j} t$$

and $\|\mathbf{v}\|^2 = 10^{r-j} t$. Moreover,

$$\langle \mathbf{u} - \underline{\mathbf{u}}, \mathbf{v} \rangle = \langle \mathbf{u}^j, \mathbf{v}^j \rangle = 10^{r-j} \langle \mathrm{enc}'_1(S^j, \mathbb{X}), \mathrm{enc}'_2(k, \mathbb{X}) \rangle.$$

Equations (B.3) and (B.4) conclude the proof of Lemma B.2.

C Proof for Other Applications

C.1 Proof of Theorem 4.3 The proof follows the same line as the proof of Theorem 4.3 in [23], where we use a JL transform to provide a solution for the $(n/2)$-fold ℓ_2 estimation.

Consider an instance of $\ell_2(n/2, d, 1, 4\epsilon)$ where Alice has vectors $\mathbf{u}^1, \mathbf{u}^2, \ldots, \mathbf{u}^{n/2} \in \{-1, 0, 1\}^d$ and Bob has vectors $\mathbf{v}^1, \mathbf{v}^2, \ldots, \mathbf{v}^{n/2} \in \{-1, 0, 1\}^d$. We consider the following shared-randomness protocol for this problem. Let $(\mathcal{F}, \mu)$ be a JLT(ϵ, δ, n, d) transform with dimension k as small as possible. The players use their shared randomness to agree upon a matrix S sampled from $\mathcal{F}$ according to μ. Then Alice computes $S\mathbf{u}^1, S\mathbf{u}^2, \ldots, S\mathbf{u}^{n/2}$ and stops if $\|S\mathbf{u}^i\|^2 > (1+\epsilon)\|\mathbf{u}^i\|^2$ for some i; otherwise, she rounds each entry of these vectors to the nearest additive multiple of $\epsilon/\sqrt{k}$ and sends the rounded vectors $\{\tilde{\mathbf{u}}^i\}_i$ to Bob. Bob then computes $S\mathbf{v}^1, S\mathbf{v}^2, \ldots, S\mathbf{v}^{n/2}$ and rounds their entries just as Alice did to obtain the vectors $\{\tilde{\mathbf{v}}^i\}_i$. Finally Bob outputs $\|\tilde{\mathbf{u}}^i - \tilde{\mathbf{v}}^i\|$ for each $i \in [n/2]$.

By the triangle inequality,

$$\|\tilde{\mathbf{u}}^i - \tilde{\mathbf{v}}^i\| = \|S\mathbf{u}^i - S\mathbf{v}^i\| \pm \left(\|\tilde{\mathbf{u}}^i - S\mathbf{u}^i\| + \|\tilde{\mathbf{v}}^i - S\mathbf{v}^i\|\right),$$

or using the definition of $\tilde{\mathbf{u}}^i$ and $\tilde{\mathbf{v}}^i$, $\|\tilde{\mathbf{u}}^i - \tilde{\mathbf{v}}^i\| = \|S\mathbf{u}^i - S\mathbf{v}^i\| \pm \epsilon$. But notice that whenever $\mathbf{u}^i = \mathbf{v}^i$, we have exactly $\|\tilde{\mathbf{u}}^i - \tilde{\mathbf{v}}^i\| = \|S\mathbf{u}^i - S\mathbf{v}^i\| = 0$, and $\mathbf{u}^i \neq \mathbf{v}^i$ implies $\|\mathbf{u}^i - \mathbf{v}^i\| \geq 1$ (because of the discrete domain $\{-1, 0, 1\}^d$). This then gives that

$$\text{(C.5)} \qquad \|\tilde{\mathbf{u}}^i - \tilde{\mathbf{v}}^i\| = (1 \pm \epsilon)\|S\mathbf{u}^i - S\mathbf{v}^i\|.$$

Now suppose $\|S(\mathbf{u}^i - \mathbf{v}^i)\| = (1 \pm 3\epsilon)\|\mathbf{u}^i - \mathbf{v}^i\|$ for all i and also $\|S\mathbf{u}^i\|^2 = (1 \pm \epsilon)\|\mathbf{u}^i\|^2$ for all i, which happens with probability at least $1 - 2\delta$. In this case, it follows directly from equation (C.5) that Bob outputs the desired estimate $(1 \pm 4\epsilon)\|\mathbf{u}^i - \mathbf{v}^i\|$ for all i. Moreover, Alice does not send too many bits:

because the input vectors have entries in $\{-1,0,1\}$, $\|S\mathbf{u}^i\|^2 = (1\pm\epsilon)\|\mathbf{u}^i\|^2 \le 2d$ and so every entry of $S\mathbf{u}^i$ (in absolute value) is upper bounded by $2d$; it then takes Alice $O\left(nk\log\left(\frac{dk}{\epsilon}\right)\right)$ bits to send all $\tilde{\mathbf{u}}^i$'s.

Therefore, the above protocol solves $\ell_2(n/2,d,1,4\epsilon)$ with probability at least $1-2\delta$ and communication $O\left(nk\log\left(\frac{dk}{\epsilon}\right)\right)$, using shared randomness. But using the lower bound that follows from Theorem 4.1 and equation (4.4) (and the fact that n is sufficiently large), we get $R_{2\delta}^{\rightarrow,\mathrm{pub}}(\ell_2(n/2,d,1,4\epsilon)) \ge \Omega\left(n\frac{1}{\epsilon^2}\log\frac{n}{\delta}\log d\right)$. The fact that $k \le d$ and the assumption that (in particular) $d \ge 1/\epsilon$ give that $k \ge \Omega(\frac{1}{\epsilon^2}\log\frac{n}{\delta})$ as desired.

C.2 Proof of Theorem 4.4 As in Section 4.2, we obtain the lower bound by analyzing the case of small and large alphabet sizes separately, and we also assume without loss of generality that δ is at most a sufficiently small constant.

Lower Bound For Small Alphabet Size. In this section we consider $M=1$ and obtain the following.

Lemma C.1. *Assume that n is at least a sufficiently large constant and that δ and ϵ are at most a sufficiently small constant. Also assume that there is a constant $\gamma > 0$ such that $d^{1-\gamma} \ge \frac{1}{\epsilon^2}\log\frac{n}{\delta}$. Then $R_\delta^{sketch}(\mathrm{Ip}(n,d,1,\epsilon)) \ge \Omega(n\frac{1}{\epsilon^2}\log\frac{n}{\delta}\log d)$.*

As in the problem of approximating the ℓ_2 norm, the lower bound is again based on the problem $\mathrm{Ind}(n\mathrm{SetInd}(\epsilon,\delta),\log d)$, but now the main element in the reduction is the encoding provided by Lemma B.2.

We show how to use the n-fold inner product problem $\mathrm{Ip}(n,d,1,\epsilon/25)$ to solve $\mathrm{Ind}(n\mathrm{SetInd}(\epsilon,\delta),c\log d)$, for a constant c depending on γ. As in Section 4.2, let Alice's instance for $\mathrm{Ind}(n\mathrm{SetInd}(\epsilon,\delta),c\log d)$ be given by the sets $\{S_i^\ell\}_{i\in[n],\ell\in[c\log d]}$ and let Bob's instance be given by the index $j \in [c\log d]$, the elements $k_1,k_2,\ldots,k_n$, the sets $\{S_i^\ell\}_{i\in[n],\ell<j}$ and the sets $S_1',S_2',\ldots,S_n'$ (although we will ignore the latter sets). They want to decide whether $k_i \in S_i^j$ holds or not for each $i\in[n]$.

To do so, independently for each $i\in[n]$ the players use the reduction from Lemma B.2 with success probability $\eta = \delta/n$ and $r = c\log d$ to make Alice obtain the vector $\mathbf{u}_i = \mathbf{u}_i(S_i^1,S_i^2,\ldots,S_i^{c\log d},R)$ and Bob obtain the vectors $\underline{\mathbf{u}}_i = \underline{\mathbf{u}}_i(S_i^1,S_i^2,\ldots,S_i^{j-1},R)$ and $\mathbf{v}_i = \mathbf{v}_i(j,k_i,R)$, using their shared randomness to simulate R. Notice that these vectors have at most $O(d^c\frac{1}{\epsilon^2}\log\frac{n}{\delta})$ coordinates, which is at most $O(d)$ for a sufficiently small c depending only on γ. Then the players can use a sketching protocol that solves $\mathrm{Ip}(n,O(d),1,\epsilon/25)$ with success probability $1-\delta$ to compute, for all $i\in[n]$, $val_i = \langle\mathbf{u}_i-\underline{\mathbf{u}}_i,\mathbf{v}_i\rangle \pm \frac{\epsilon}{25}\|\mathbf{u}_i-\underline{\mathbf{u}}_i\|\|\mathbf{v}_i\| = \langle\mathbf{u}_i-\underline{\mathbf{u}}_i,\mathbf{v}_i\rangle \pm \frac{\epsilon}{25}10^{r-j}t$, where $t=\frac{72}{\epsilon^2}\log\frac{n}{\delta}$; they do so by having Alice sending Bob the linear sketches of the vectors $\mathbf{u}_i$'s, then Bob updating these sketches to obtain sketches of the vectors $\{\mathbf{u}_i-\underline{\mathbf{u}}_i\}_i$, and finally executing Bob's part of the protocol to approximate the values $\langle\mathbf{u}_i-\underline{\mathbf{u}}_i,\mathbf{v}_i\rangle$. Having these approximations at hand, for each $i\in[n]$ Bob outputs that $k_i\in S_i^j$ iff $val_i \ge 10^{r-j}t(\frac{1}{2}+\frac{3\epsilon}{24})$.

Since the guarantees from Lemma B.2 hold for all triples $(\mathbf{u}_i\underline{\mathbf{u}}_i,\mathbf{v}_i)$ for $i\in[n]$ with probability at least $1-\delta$, it is easy to see that Bob outputs the correct answer with probability at least $1-2\delta$. This implies that $R_\delta^{\mathrm{sketch}}(\mathrm{Ip}(n,O(d),1,\epsilon/25)) \ge R_{2\delta}^{\rightarrow,\mathrm{pub}}(\mathrm{Ind}(n\mathrm{SetInd}(\epsilon,\delta),c\log d)$. Corollary 4.3, equation (4.4) and the assumption that n is sufficiently large conclude the proof of Lemma C.1.

Lower Bound for Small Dimension. In this section we obtain the following bound.

Lemma C.2. *Assume that n is at least a sufficiently large constant and that δ and ϵ are at most a sufficiently small constant. Also assume that $d \ge \Omega(\frac{1}{\epsilon^2}\log\frac{n}{\delta})$ and that there is a constant $\gamma>0$ such that $M^{1-\gamma} \ge \frac{1}{\epsilon^2}\log\frac{n}{\delta}$. Then $R_\delta^{sketch}(\mathrm{Ip}(n,d,M,\epsilon)) \ge \Omega\left(n\frac{1}{\epsilon^2}\log\frac{n}{\delta}\log M)\right)$.*

As observed, for instance, in [34], JL transforms also approximate inner products.

Proposition C.1. *Let $(\mathcal{F},\mu)$ be a JLT(ϵ,δ,n,d). Consider $S\sim\mu$. Then for every collection of n vectors $\mathbf{u}^1,\mathbf{u}^2,\ldots,\mathbf{u}^n$ in $\mathbb{R}^d$, with probability at least $1-\delta$ we have $\langle S\mathbf{u}^i,S\mathbf{u}^j\rangle = \langle\mathbf{u}^i,\mathbf{u}^j\rangle \pm \epsilon\|\mathbf{u}^i\|\|\mathbf{u}^j\|$ for all $i,j\in[n]$.*

Lemma C.2 is then obtained from the previous reduction by applying the JL transform of Theorem 4.2 to the vectors $\mathbf{u}_i$, $\underline{\mathbf{u}}_i$ and $\mathbf{v}_i$, just as done in the second part of Section 4.2.

C.3 Proof of Theorem 4.5 To prove the first part of the theorem, notice that $(AB)_{i,i} \pm \epsilon\|A_i\|\|B^i\| = \langle A_i,B^i\rangle \pm \epsilon\|A_i\|\|B^i\|$. Therefore, a sketch S that allows (with probability at least $1-\delta$) the approximation $(AB)_{i,j} \pm \|A_i\|\|B^j\|$ for all $i,j\in[n]$ can be used to solve the inner product problem $\mathrm{Ip}(n,n,M,\epsilon)$ with probability $1-\delta$ and communication equal to the bit size of AS; the desired lower bound then follows directly from Theorem 4.4.

For the second part of the theorem, we can set $B=A^T$ to obtain $(AA^T)_{i,i}\pm\epsilon\|A_i\|^2 = (1\pm\epsilon)\|A_i\|^2$, and hence the sketch S is a JLT(ϵ,δ,n,n); the lower bound $k \ge \Omega(\frac{1}{\epsilon^2}\log\frac{n}{\delta})$ then follows from Theorem 4.3.

C.4 Proof of Theorem 4.6 The lower bound of $\Omega(n\frac{1}{\epsilon^2}\log\frac{n}{\delta}\log|\mathcal{D}|)$ follows directly from Lemma C.1

(notice that the hard instances in this lemma are provided by $\{0,1\}$ vectors). However, the lower bound $\Omega(n\frac{1}{\epsilon^2}\log\frac{n}{\delta}\log M)$ does not follow directly from Lemma C.2, because there the hard instances are given by vectors which can have negative coordinates. The latter lower bound comes from the following modification of the hard instances for inner products.

Lemma C.3. *Assume that n and M are at least a sufficiently large constant and assume that δ and ϵ are at most a sufficiently small constant. Also assume that $d \geq n/(\epsilon^2\delta)$. Then $R_\delta^{sketch}(\mathrm{Ip}(n,d,M,\epsilon)) \geq \Omega\left(n\frac{1}{\epsilon^2}\log\frac{n}{\delta}\log M)\right)$. Moreover, this holds even if the protocol only offers guarantees for vectors in $\{0,1,\ldots,M\}^d$.*

Proof. Consider the problem $\mathrm{Ind}(n\mathrm{SetInd}(\epsilon,\delta),\log M)$. Let Alice's instance for this problem be given by the sets $\{S_i^\ell\}_{i\in[n],\ell\in[\log M]}$, and let Bob's instance be given by the index $j\in[\log M]$, the elements $k_1,k_2,\ldots,k_n$, the sets $\{S_i^\ell\}_{i\in[n],\ell<j}$ and the sets $S'_1,S'_2,\ldots,S'_n$. They want to decide whether $k_i\in S_i^j$ holds or not for each i. A trivial but important observation is that $k_i\in S_i^j$ iff the inner product $\langle\chi_{S_i^j},e_{k_i}\rangle$ equals 1, where $\chi_{S_i^j}\in\{0,1\}^{n/(\epsilon^2\delta)}$ is the incidence vector of S_i^j and e_{k_i} is the k_i'th canonical vector.

To solve this problem, for each $i\in[n]$ Alice makes the vector $\mathbf{u}^i \triangleq \sum_{\ell=1}^{\log M} 10^{\log M-\ell}\chi_{S_i^\ell}$, and for every $i\in[n]$ Bob makes the vectors $\underline{\mathbf{u}}^i \triangleq \sum_{\ell=1}^{j-1} 10^{\log M-\ell}\chi_{S_i^\ell}$ and $\mathbf{v}^i \triangleq 10^{\log M-j}e_{k_i}$. Notice that the constructed vectors lie in $\{0,1,\ldots,M'\}^d$, where $M'=M^{10}$ and $d=n/(\epsilon^2\delta)$. Then using the shared randomness, Alice runs a sketching protocol for $\mathrm{Ip}(n,d,M',\epsilon/4)$ to send Bob sketches $\mathcal{S}\mathbf{u}^1,\mathcal{S}\mathbf{u}^2,\ldots,\mathbf{u}^n$ that allows computation of n-fold $(\epsilon/4)$-approximations for dot products with probability at least $1-\delta$. Then Bob updates the sketches to obtain $\mathcal{S}(\mathbf{u}^1-\underline{\mathbf{u}}^1),\mathcal{S}(\mathbf{u}^2-\underline{\mathbf{u}}^2),\ldots,\mathcal{S}(\mathbf{u}^n-\underline{\mathbf{u}}^n)$, and use them to compute the inner product approximations $val_i=\langle(\mathbf{u}^i-\underline{\mathbf{u}}^i),\mathbf{v}^i\rangle\pm\frac{\epsilon}{4}\|\mathbf{u}^i-\underline{\mathbf{u}}^i\|\|\mathbf{v}^i\|$ for all $i\in[n]$, with probability at least $1-\delta$. Finally, for each i Bob reports that $k_i\in S_i^j$ iff $val_i\geq 10^{2(\log M-j)}/2$.

We claim that: (i) $\langle(\mathbf{u}^i-\underline{\mathbf{u}}^i),\mathbf{v}^i\rangle = 10^{2(\log M-j)}\langle\chi_{S_i^j},e_{k_i}\rangle \pm \frac{10^{2(\log M-j)}}{9}$ and (ii) $\|\mathbf{u}^i-\underline{\mathbf{u}}^i\|\|\mathbf{v}^i\| \leq \frac{10^{2(\log M-j)+1}}{9\epsilon}$; since $\langle\chi_{S_i^j},e_{k_i}\rangle$ equals 1 if $k_i\in S_i^j$ and 0 otherwise, these bounds show that the above protocol solves the instance of $\mathrm{Ind}(n\mathrm{SetInd}(\epsilon),\log M)$ with probability at least $1-\delta$.

To prove (i), by bilinearity of inner products we have

$$\langle(\mathbf{u}^i-\underline{\mathbf{u}}^i),\mathbf{v}^i\rangle = \sum_{\ell=j}^{\log M} 10^{2\log M-\ell-j}\langle\chi_{S_i^\ell},e_{k_i}\rangle$$
$$= 10^{2(\log M-j)}\langle\chi_{S_i^j},e_{k_i}\rangle \pm \frac{10^{2(\log M-j)}}{9},$$

where the inequality follows from the fact that $\langle\chi_{S_i^\ell},e_{k_i}\rangle\leq 1$ for all i,ℓ. To prove (ii), notice that $|S_i^\ell|=1/\epsilon^2$ and hence $\|\chi_{S_i^\ell}\|=1/\epsilon$ for all i,ℓ. Using triangle inequality, we obtain that $\|\mathbf{u}^i-\underline{\mathbf{u}}^i\|\leq(1/\epsilon)\sum_{\ell=j}^{\log M}10^{\log M-\ell}\leq\frac{10^{\log M-j+1}}{9\epsilon}$. Since $\|\mathbf{v}^i\|=10^{\log M-j}$, part (ii) directly follows.

The above protocol shows that $R_\delta^{\rightarrow,\mathrm{pub}}(\mathrm{Ind}(n\mathrm{SetInd}(\epsilon),\log M)) \leq R_\delta^{\mathrm{sketch}}(\mathrm{Ip}(n,d,M',\epsilon/4))$. Then Corollary 4.3 together with equation (4.4) gives that $R_\delta^{\mathrm{sketch}}(\mathrm{Ip}(n,d,M,\epsilon))\geq\Omega(n\frac{1}{\epsilon^2}\log\frac{n}{\delta}\log M)$ as desired. □