The Communication Complexity of Set Intersection and Multiple Equality Testing*

Dawei Huang[†] Seth Pettie[‡] Yixiang Zhang [§] Zhijun Zhang [¶]

Abstract

In this paper we explore fundamental problems in randomized communication complexity such as computing Set Intersection on sets of size k and Equality Testing between vectors of length k. Brody et al. [BCK+16] and Sağlam and Tardos [ST13] showed that for these types of problems, one can achieve optimal communication volume of $O(k)$ bits, with a randomized protocol that takes $O(\log^* k)$ rounds. They also proved [BCK+16, ST13] that this is one point along the optimal round-communication tradeoff curve.

Aside from rounds and communication volume, there is a *third* parameter of interest, namely the *error probability* p_{err}. It is straightforward to show that protocols for Set Intersection or Equality Testing need to send $\Omega(k+\log p_{\mathrm{err}}^{-1})$ bits. Is it possible to simultaneously achieve optimality in all three parameters, namely $O(k+\log p_{\mathrm{err}}^{-1})$ communication and $O(\log^* k)$ rounds?

In this paper we prove that there is no universally optimal algorithm, and complement the existing round-communication tradeoffs [BCK+16, ST13] with a new tradeoff between rounds, communication, and probability of error. In particular:

- Any protocol for solving Multiple Equality Testing in r rounds with failure probability $p_{\mathrm{err}} = 2^{-E}$ has communication volume $\Omega(Ek^{1/r})$.
- There exists a protocol for solving Multiple Equality Testing in $r + \log^*(k/E)$ rounds with $O(k + rEk^{1/r})$ communication, thereby essentially matching our lower bound and that of [BCK+16, ST13].
- Lower bounds on Equality Testing extend to Set Intersection, for every r, k, and p_{err} (which is trivial); in the reverse direction, upper bounds on Equality Testing for r, k, p_{err} imply similar upper bounds on Set Intersection with parameters $r + 1, k$, and p_{err}.

Our original motivation for considering p_{err} as an independent parameter came from the problem of enumerating triangles in distributed (CONGEST) networks having maximum degree Δ. We prove that this problem can be solved in $O(\Delta/\log n + \log\log\Delta)$ time with high probability $1 - 1/\mathrm{poly}(n)$. This beats the trivial (deterministic) $O(\Delta)$-time algorithm and is superior to the $\tilde{O}(n^{1/3})$ algorithm of [CPZ19, CS19] when $\Delta = \tilde{O}(n^{1/3})$.

*Supported by NSF grants CCF-1514383, CCF-1637546, and CCF-1815316.

[†]University of Michigan.

[‡]University of Michigan.

[§]IIIS, Tsinghua University

[¶]IIIS, Tsinghua University

1 Introduction

Communication Complexity was defined by Yao [Yao79] in 1979 and has become an indispensible tool for proving lower bounds in models of computation in which the notions of *parties* and *communication* are not direct. See, e.g., books and monographs [Rou16, RY, KN97] and surveys [CP10, Lov89] on the subject. In this paper we consider some of the most fundamental and well-studied problems in this model, such as SetDisjointness, SetIntersection, ExistsEqual, and EqualityTesting. Let us briefly define these problems formally since the terminology is not completely standard.

SetDisjointness and SetIntersection. In the SetDisjointness problem Alice and Bob receive sets $A \subset U$ and $B \subset U$ where $|A|, |B| \le k$ and must determine whether $A \cap B = \emptyset$. Define $\mathsf{SetDisj}(k, r, p_{\mathrm{err}})$ to be the minimum communication complexity of an r-round randomized protocol for this problem that errs with probability at most p_{err}. We can assume that $|U| = O(k^2/p_{\mathrm{err}})$ without loss of generality.[1] The input to the SetIntersection problem is the same, except that the parties must *report the entire set* $A \cap B$. Define $\mathsf{SetInt}(k, r, p_{\mathrm{err}})$ to be the minimum communication complexity of an r-round protocol for SetIntersection.

EqualityTesting and ExistsEqual. In the EqualityTesting problem Alice and Bob hold vectors $\mathbf{x} \in U^k$ and $\mathbf{y} \in U^k$ and must determine, for each index $i \in [k]$, whether $x_i = y_i$ or $x_i \neq y_i$. A potentially easier version of the problem, ExistsEqual, is to determine if there *exists* at least one index $i \in [k]$ for which $x_i = y_i$. Define $\mathsf{Eq}(k, r, p_{\mathrm{err}})$ to be the randomized communication complexity of any r-round protocol for EqualityTesting that errs with probability p_{err}, and $\exists\mathsf{Eq}(k, r, p_{\mathrm{err}})$ the corresponding complexity of ExistsEqual. Once again, we can

[1]Before the first round of communication, pick a pairwise independent $h : U \mapsto [O(k^2/p_{\mathrm{err}})]$ and check whether $h(A) \cap h(B) = \emptyset$ with error probability $p_{\mathrm{err}}/2$. Thus, having SetDisj depend additionally on $|U|$ is somewhat redundant, at least when $|U|$ is large.

assume that $|U| = O(k/p_{\text{err}})$ without loss of generality.

The deterministic communication complexity of these problems is well understood [KN97],[2] so we consider randomized complexity exclusively. Although these problems are well studied [HW07, FKNN95, BCK+16, ST13, KS92], most prior work has focused on the relationship between *round complexity* and *communication volume*, and paid relatively little attention to the role of p_{err}. [HW07, FKNN95, ST13, KS92]. Brody et al.[BCK+16] incorporated p_{err} into the round vs communication trade-off and in particular distinguished between the role of false positives and false negatives.

History. Håstad and Wigderson [HW07] gave an $O(\log k)$-round protocol for SetDisjointness in which Alice and Bob communicate $O(k)$ bits, which matched an $\Omega(k)$ lower bound of Kalyanasundaram and Schnitger [KS92]; see also [Raz92, BGMdW13, DKS12]. Feder et al. [FKNN95] proved that EqualityTesting can be solved with $O(k)$ communication by an $O(\sqrt{k})$-round protocol that errs with probability $\exp(-\sqrt{k})$. The round complexity and error probability were later improved to $\log k$ and $\exp(-k/\text{polylog}(k))$, respectively [Nik13].

Improving [HW07], Sağlam and Tardos [ST13] gave an r-round protocol for SetDisjointness that uses $O(k \log^{(r)} k)$ communication, where $\log^{(r)}$ is the r-fold iterated logarithm function. For $r = \log^* k$ the error probability of this algorithm is $\exp(-\sqrt{k})$, coincidentally matching [FKNN95]. In independent work, Brody et al. [BCK+16] gave r-round and $O(r)$-round protocols for ExistsEqual and SetIntersection, respectively, that use $O(k \log^{(r)} k)$ communication and err with probability $1/\text{poly}(k)$.

Sağlam and Tardos [ST13] proved that this $O(k \log^{(r)} k)$ round vs communication tradeoff is optimal, using a combinatorial round elimination technique. In particular, this lower bound applies to any ExistsEqual protocol with constant error probability. Independently, Brody et al. [BCK+16] gave a simpler proof for the EqualityTesting problem with the same tradeoff curve, but only holds for protocols with error probability of $1/\text{poly}(k)$. Brody et al. [BCK+16] also introduced a *randomized* reduction from SetIntersection to EqualityTesting, which carries a probability of error that is only tolerable if $p_{\text{err}} > \exp(-\tilde{O}(\sqrt{k}))$.

1.1 Contributions

First, we observe that a simple *deterministic* reduction shows that SetIntersection is essentially equivalent to EqualityTesting for any p_{err}, up to one round of communication, and SetDisjointness is essentially equivalent to ExistsEqual for any p_{err}. Theorem 1.1 is proved in Appendix A; it is inspired by the randomized reduction of Brody et al. [BCK+16].

THEOREM 1.1. *For any parameters $k \geq 1, r \geq 1$, and $p_{\text{err}} > 0$, it holds that*

$$\begin{aligned}
\mathsf{Eq}(k, r, p_{\text{err}}) &\leq \mathsf{SetInt}(k, r, p_{\text{err}}),\\
\exists\mathsf{Eq}(k, r, p_{\text{err}}) &\leq \mathsf{SetDisj}(k, r, p_{\text{err}}),\\
\mathsf{SetInt}(k, r+1, p_{\text{err}}) &\leq \mathsf{Eq}(k, r, p_{\text{err}}) + \zeta,\\
\mathsf{SetDisj}(k, r+1, p_{\text{err}}) &\leq \exists\mathsf{Eq}(k, r, p_{\text{err}}) + \zeta,
\end{aligned}$$

where $\zeta = O(k + \log\log p_{\text{err}}^{-1})$.

Second, we prove that in any of the four problems, it is impossible to simultaneously achieve communication volume $O(k+\log p_{\text{err}}^{-1})$ in $O(\log^* k)$ rounds for all k, p_{err}. Specifically, for $p_{\text{err}} = 2^{-E}$, any r-round protocol needs $\Omega(Ek^{1/r})$ communication. A key takeaway for this result is that for any $E > k$, if one wishes to achieve error probability 2^{-E} with the optimal $O(k + E)$ communication, one needs $\Omega(\log k)$ rounds, instead of $\log^* k$ rounds. We complement this lower bound with an upper bound showing that in $r + \log^*(k/E)$ rounds, we can solve EqualityTesting with $O(k+rEk^{1/r})$ communication. This matches our lower bound when $E \geq k$ and r is constant, but is slightly suboptimal when $r = \omega(1)$. We illustrate two ways to shave off this factor of r. We give an $(r+\log^*(k/E))$-round ExistsEqual protocol that communicates $O(k + Ek^{1/r})$ bits, as well as an EqualityTesting protocol that communicates $O(k + Ek^{1/r})$ bits, but with round complexity $O(r) + \log^*(k/E)$.

Our original interest in SetIntersection came from distributed subgraph detection in CONGEST[3] networks, which has garnered significant interest in recent years [CS19, CPZ19, IG17, ACKL17, DKO14, KR18, FGKO18, CK18, GO18]. Izumi and LeGall [IG17] proved that triangle enumeration[4] requires $\Omega(n^{1/3}/\log n)$ rounds in the CONGEST model, and further showed that *local* triangle enumeration[5] re-

[2]When $p_{\text{err}} = 0$, the deterministic complexity must be expressed in terms of k and $|U|$.

[3]In the CONGEST model there is a graph $G = (V, E)$ whose vertices are identified with processors and whose edges represent bidirectional communication links. Each vertex v does not know G, and is only initially aware of an $O(\log n)$-bit $\mathsf{ID}(v)$, $\deg(v)$, and global parameters $n \geq |V|$ and $\Delta \geq \max_{u \in V} \deg(u)$. Communication proceeds in synchronized rounds; in each round, each processor can send a (different) $O(\log n)$-bit message to each of its neighbors.

[4]Every triangle (3-cycle) in G must be reported by *some* vertex.

[5]Every triangle in G must be reported by at least one of the three constituent vertices. Izumi and LeGall [IG17] only stated the $\Omega(n/\log n)$ lower bound but it can also be expressed in terms of Δ.

Problem	Commun.	Rounds	Error Probability	Ref.
EqualityTesting	$O(k)$	$O(\sqrt{k})$	$\exp(-\sqrt{k})$	[FKNN95]
EqualityTesting	$O(k)$	$\log k$	$\exp(-k/\text{polylog}(k))$	[Nik13]
SetDisjointness	$O(k)$	$O(\log k)$	Constant	[HW07]
SetDisjointness	$O(k\log^{(r)} k)$	r	$\geq \exp(-\sqrt{k})$	[ST13]
ExistsEqual	$O(k\log^{(r)} k)$	r	$1/\text{poly}(k)$	[BCK+16]
SetIntersection		$O(r)$		
ExistsEqual / [SetDisjointness]	$O(k+Ek^{1/r})$	$r+\log^*(k/E)$ [+1]	2^{-E}	**new**
EqualityTesting / [SetIntersection]	$O(k+rEk^{1/r})$	$r+\log^*(k/E)$ [+1]	2^{-E}	
	$O(k+Ek^{1/r})$	$O(r)+\log^*(k/E)$ [+1]		

Lower Bounds

Problem	Commun.	Rounds	Error Probability	Ref.
SetDisjointness	$\Omega(\sqrt{k})$	∞	Constant	[BFS86]
SetDisjointness	$\Omega(k)$	∞	Constant	[KS92]
ExistsEqual	$\Omega(k\log^{(r)} k)$	r	Constant	[ST13]
ExistsEqual	$\Omega(k\log^{(r)} k)$	r	$1/\text{poly}(k)$	[BCK+16]
ExistsEqual	$\Omega(Ek^{1/r})$	r	2^{-E}	**new**

Table 1: Upper and Lower bounds on SetDisjointness, SetIntersection, EqualityTesting, and ExistsEqual. Via trivial reductions, lower bounds on ExistsEqual extend to all four problems, and upper bounds on SetIntersection extend to all four problems. From Theorem 1.1, the upper bounds on SetIntersection and SetDisjointness follow from those of EqualityTesting and ExistsEqual, respectively, +1 round of communication. The log-star function is defined as $\log^*(x) = \min\{i : \log^{(i)}(x) \leq 1\}$, e.g., $\log^*(k/E) = 0$ if $E \geq k$.

quires $\Omega(\Delta/\log n)$ rounds in CONGEST, which can be as large as $\Omega(n/\log n)$.

The most natural way to solve (local) triangle enumeration is, for every edge $\{u, v\} \in E(G)$, to have u and v run a two-party SetIntersection protocol in which they compute $N(u) \cap N(v)$, where $N(u) = \{\mathsf{ID}(x) \mid \{u, x\} \in E(G)\}$ and $\mathsf{ID}(x) \in \{0,1\}^{O(\log n)}$ is x's unique identifier. Any r-round protocol with communication volume $O(\Delta)$ can be simulated in CONGEST in $O(\Delta/\log n + r)$ rounds since the message size is $O(\log n)$ bits. However, to guarantee a *global* probability of success at least $1 - 1/\text{poly}(n)$, the failure probability of each SetIntersection instance must be $p_{\text{err}} = 2^{-E}$, $E = \Theta(\log n)$, which is independent of Δ. Our communication complexity lower bound suggests that to achieve this error probability, we would need $\Omega((\Delta+E\Delta^{1/r})/\log n+r)$ CONGEST rounds, i.e., with $r = \log\Delta$ we should not be able to do better than $O(\Delta/\log n+\log\Delta)$. We prove that (local) triangle enumeration can actually be solved *exponentially* faster, in $O(\Delta/\log n + \log\log\Delta)$ CONGEST rounds, without *necessarily* solving every SetIntersection instance.

Organization. The proof of Theorem 1.1 on the near-equivalence of SetIntersection/SetDisjointness and EqualityTesting/ExistsEqual appears in Appendix A. Section 2 reviews concepts from information theory and communication complexity. In Section 3 we present new lower bounds for both EqualityTesting and ExistsEqual that incorporate rounds, communication, and error probability. Section 4 presents nearly matching upper bounds for EqualityTesting and ExistsEqual, and Section 5 applies them to the distributed triangle enumeration problem. We conclude with some open problems in Section 6.

2 Preliminaries

2.1 Notational Conventions The set of positive integers at most t is denoted $[t]$. Random variables are typically written as capital letters (X, Y, M, etc.) and the values they take on are lower case (x, y, m, etc.). The letters $p, q, \mu, \mathcal{D}$ are reserved for probability mass functions (*p.m.f.*). E.g., $\mathcal{D}(x)$ denotes the probability that $X = x$ whenever $X \sim \mathcal{D}$. The *support* $\text{supp}(\mathcal{D})$ of a distribution $\mathcal{D}$ is the set of all x for which $\mathcal{D}(x) > 0$. If $\mathcal{X} \subseteq \text{supp}(\mathcal{D})$, $\mathcal{D}(\mathcal{X}) = \sum_{x\in\mathcal{X}} \mathcal{D}(x)$.

Many of our random variables are vectors. If x is a k-dimensional vector and $I \subseteq [k]$, x_I is the projection of x onto the coordinates in I and x_i is short for $x_{\{i\}}$. Similarly, if $\mathcal{D}$ is the p.m.f. of a k-dimensional random variable, $\mathcal{D}_I$ is the marginal distribution of $\mathcal{D}$ on the index set $I \subseteq [k]$.

Throughout the paper, log and exp are the base-2 logarithm and exponential functions, and $\log^{(r)}$ and

$\exp^{(r)}$ their r-fold iterated versions:

$$\log^{(0)}(x) = \exp^{(0)}(x) = x,$$
$$\log^{(r)}(x) = \log(\log^{(r-1)}(x)),$$
$$\exp^{(r)}(x) = \exp(\exp^{(r-1)}(x)).$$

The log-star function is defined to be $\log^*(x) = \min\{r \mid \log^{(r)}(x) \le 1\}$. In particular, $\log^*(x) = 0$ if $x \le 1$.

2.2 Information Theory The most fundamental concept in information theory is *Shannon entropy.* The Shannon entropy of a discrete random variable X is defined as

$$\mathrm{H}(X) = -\sum_{x \in \mathrm{supp}(X)} \Pr[X = x] \log \Pr[X = x].$$

Since there may be cases in which different distributions are defined for the "same" random variable, we use $\mathrm{H}(p)$ in place of $\mathrm{H}(X)$ if X is drawn from a p.m.f. p. We also write $\mathrm{H}(\alpha)$, $\alpha \in (0, 1)$, to be the entropy of a Bernoulli random variable with success probability α. In general, we freely use a random variable and its p.m.f. interchangeably.

The *joint entropy* $\mathrm{H}(X, Y)$ of two random variables X and Y is simply

$$\mathrm{H}(X, Y) = -\sum_{\substack{x \in \\ \mathrm{supp}(X)}} \sum_{\substack{y \in \\ \mathrm{supp}(Y)}} \Pr[X = x \wedge Y = y] \log \Pr[X = x \wedge Y = y].$$

This notion can be easily extended to cases of more than two random variables. Here, we state a well known fact about joint entropy.

Fact 2.1. *For any random variables $X_1, X_2, \ldots, X_n$, their joint entropy is at most the sum of their individual entropies, i.e.,* $\mathrm{H}(X_1, X_2, \ldots, X_n) \le \sum_{i=1}^{n} \mathrm{H}(X_i)$.

The *conditional entropy* of Y conditioned on another random variable X, denoted $\mathrm{H}(Y \mid X)$, measures the expected amount of extra information required to fully describe Y if X is known. It is defined to be

$$\begin{aligned} &\mathrm{H}(Y \mid X) \\ &= \mathrm{H}(X, Y) - \mathrm{H}(X) \\ &= -\sum_{x \in \mathrm{supp}(X)} \Pr[X = x] \\ &\quad \sum_{y \in \mathrm{supp}(Y)} \Pr[Y = y \mid X = x] \log \Pr[Y = y \mid X = x] \ \ge\ 0, \end{aligned}$$

which can be viewed as a weighted sum of entropies of a number of conditional distributions.

Finally, the *mutual information* $\mathrm{I}(X\ ;\ Y)$ between two random variables X and Y quantifies the amount of information that is revealed about one random variable through knowing the other one:

$$\begin{aligned} &\mathrm{I}(X\ ;\ Y) \\ &= \mathrm{H}(X) - \mathrm{H}(X \mid Y) \\ &= \mathrm{H}(X) + \sum_{y \in \mathrm{supp}(Y)} \Pr[Y = y] \\ &\quad \sum_{x \in \mathrm{supp}(X)} \Pr[X = x \mid Y = y] \log \Pr[X = x \mid Y = y]. \end{aligned}$$

2.3 Communication Complexity Let $f(x, y)$ be a function over domain $\mathcal{X} \times \mathcal{Y}$, and consider any two-party communication protocol $Q(x, y)$ that computes $f(x, y)$, where one party holds x and the other holds y. The *transcript* of Q on (x, y) is defined to be the concatenation of all messages exchanged by the two parties, in order, as they execute on input (x, y). The *communication cost* of Q is the maximum transcript length produced by Q over all possible inputs.

Let Q_d be a deterministic protocol for f and suppose μ is a distribution over $\mathcal{X} \times \mathcal{Y}$. The *distributional error probability of Q_d with respect to μ* is the probability $\Pr_{(x,y)\sim\mu}[Q_d(x, y) \ne f(x, y)]$. For any $0 < \epsilon < 1$, the *(μ, ϵ)-distributional deterministic communication complexity* of the function f is the minimum communication cost of any protocol Q_d that has distributional error probability at most ϵ with respect to the distribution μ.

A randomized protocol $Q_r(x, y, w)$ also takes a public random string $w \sim \mathcal{W}$ as input. The error probability of Q_r is calculated as $\max_{(x,y)\in\mathcal{X}\times\mathcal{Y}} \Pr_{w\sim\mathcal{W}}[Q_r(x, y, w) \ne f(x, y)]$. The *$\epsilon$-randomized communication complexity* of f is the minimum communication cost of Q_r over all protocols Q_r with error probability at most ϵ.

Yao's *minimax principle* [Yao77] is a common starting point for lower bound proofs in randomized communication complexity. The easy direction of Yao's minimax principle states that the communication cost of the best deterministic protocol specific to any particular distribution is at most the communication cost of any randomized protocol on its worst case input.

Lemma 2.1. (Yao's minimax principle [Yao77]) *Let $f : \mathcal{X} \times \mathcal{Y} \mapsto \mathcal{Z}$ be the function to be computed. Let $D_{\mu,\epsilon}(f)$ be the (μ, ϵ)-distributional deterministic communication complexity of f, and let $R_\epsilon(f)$ be the ϵ-randomized communication complexity of f. Then for any $0 < \epsilon < 1/2$,*

$$\max_{\mu} D_{\mu,\epsilon}(f) \le R_\epsilon(f).$$

Therefore, to show a lower bound on the ϵ-randomized communication complexity of a function f, it suffices to find a hard distribution μ on the input set and prove a lower bound for the communication cost of any deterministic protocol that has distributional error probability at most ϵ with respect to μ.

3 Lower Bounds

In this section we prove lower bounds on EqualityTesting and ExistsEqual. Theorem 3.1 obviously follows directly from Theorem 3.2, but we prove them in that order nonetheless because Theorem 3.1 is a bit simpler.

THEOREM 3.1. *Any r-round randomized protocol for EqualityTesting on vectors of length k that errs with probability $p_{\text{err}} = 2^{-E}$ requires at least $\Omega(Ek^{1/r})$ bits of communication.*

THEOREM 3.2. *Any r-round randomized protocol for ExistsEqual on vectors of length k that errs with probability $p_{\text{err}} = 2^{-E}$ requires at least $\Omega(Ek^{1/r})$ bits of communication.*

Without any constraint on the number of rounds, EqualityTesting trivially requires $\Omega(k)$ communication. ExistsEqual also requires $\Omega(k)$ communication, through a small modification to the SetDisjointness lower bounds [KS92, Raz92]. Even when $k = 1$, we need at least $\Omega(E)$ communication to solve EqualityTesting/ExistsEqual with error probability 2^{-E} [KN97]. Thus, we can assume that $E = \Omega(k^{1-1/r})$, $k^{1/r} = \Omega(1)$, and hence $r = O(\log k)$. For example, some calculations later in our proof hold when $r \le (\log k)/6$. When proving Theorem 3.2, we will further assume $E = \Omega(\log k)$ when $r = 1$, which is reasonable because of Sağlam and Tardos' $\Omega(k \log^{(r)} k) = \Omega(k \log k)$ lower bound [ST13].

3.1 Structure of the Proof

We consider deterministic strategies for ExistsEqual/EqualityTesting when Alice and Bob pick their input vectors independently from the uniform distribution on $[t]^k$, where $t = 2^{cE}$ and $c = 1/2$. Although the probability of seeing a collision in any particular coordinate is small, it is still much larger than the tolerable error probability (since $c < 1$), so it is incorrect to declare "not equal in every coordinate" without performing any communication.

We suppose, for the purpose of obtaining a contradiction, that there is a protocol for EqualityTesting with error probability 2^{-E} and communication complexity $c'Ek^{1/r}$, where $c' = c/100$. The length of the jth message is l_j, which could depend on the parameters (E, r, k, etc.) and possibly in some complicated way on the transcript of the protocol before round j.[6]

Our proof must necessarily consider transcripts of the protocol that are extremely unlikely (occurring with probability close to 2^{-E}) and also maintain a high level of uncertainty about *which* coordinates of Alice's and Bob's vectors might be equal. Consider the first message. Alice picks her input vector $x \in [t]^k$, which dictates the first message m_1. Suppose, for simplicity, that it betrays exactly $l_1/k < c'Ek^{1/r-1}$ bits of information per coordinate of x. Before Bob can respond with a message m_2 he must commit to his input, say y. Most values of y result in "good" outcomes: nearly all non-equal coordinates get detected immediately and the effective size of the problem is dramatically reduced. We are not interested in these values of y, only very "bad" values. Let I_1 be the first $k^{1-1/r}$ coordinates (or, more generally, $k^{1-1/r}$ coordinates that m_1 revealed below-average information about). With probability about $(2^{-c'Ek^{1/r-1}})^{|I_1|} = 2^{-c'E}$, Bob picks an input y that is *completely consistent* with Alice's on I_1, i.e., as far as he can tell $y_i = x_i$ for every $i \in I_1$. Rather than sample y uniformly from $[t]^k$, we sample it from a *hybrid* distribution: y_{I_1} is sampled from the same distribution that m_1 revealed about x_{I_1} (forcing the above event to happen with probability 1), and $y_{[k]\setminus I_1}$ is sampled from Bob's former distribution (in this case, the uniform distribution on $[t]^{k-|I_1|}$), conditioned on the value of y_{I_1}.

This process continues round by round. Bob's message m_2 betrays at most $l_2/|I_1| < c'Ek^{2/r-1}$ bits of information on each coordinate of y_{I_1}, and there must be an index set $I_2 \subset I_1$ with $|I_2| = k^{1-2/r}$ such that, with probability around $2^{-c'E}$, it is completely consistent that $x_{I_2} = y_{I_2}$. Alice resamples her input so that this (rare) event occurs with probability 1, generates m_3, and continues.

At the end of this process $|I_r| = k^{1-r/r} = 1$, and yet Alice and Bob have revealed less than the full cE bits of entropy about x_{I_r} and y_{I_r}. Regardless of whether they report "equal" or "not equal" (on I_r), they are wrong with probability greater than 2^{-E}. Are we done? Absolutely not! The problem is that this strange process for sampling a possible transcript of the protocol might itself only find transcripts that occur with probability $\ll 2^{-E}$, making any conclusions we make about its (probability of) correctness moot. Generally speaking, we need to show that Alice's and Bob's actions are consistent with events that occur with probability $\gg 2^{-E}$.

Let us first make every step of the above process a bit more formal. It is helpful to think about Alice's and Bob's inputs not being *fixed* vectors selected at time zero, but simply distributions over vectors that change as messages progressively reveal more information about them.

- Before the jth round of communication, the sender of the jth message's input is drawn from a discrete distribution $\widehat{\mathcal{D}}^{(j-1)}$ over $[t]^k$. The receiver of the jth message's input is drawn from the distribution $\mathcal{D}^{(j-1)}$. For example, when $j = 1$, if Alice speaks first then her initial distribution, $\widehat{\mathcal{D}}^{(0)}$, and Bob's initial distribution, $\mathcal{D}^{(0)}$, are both uniform over $[t]^k$.
- Before the jth round of communication both parties are aware of an index set I_{j-1} such that, informally, (i) the distributions $\mathcal{D}^{(j-1)}_{I_{j-1}}$ and $\widehat{\mathcal{D}}^{(j-1)}_{I_{j-1}}$ are very similar, and in particular, it is consistent that their inputs are identical on I_{j-1}, and (ii) the messages transmitted so far reveal "average" or below-average information about these coordinates. For example, $I_0 = [k]$ and it is consistent with the empty transcript that Alice's and Bob's inputs are identical on every coordinate.

[6] In the context of ExistsEqual/EqualityTesting, it is natural to think about uniform-length messages, $l_j = c'Ek^{1/r}/r$, or lengths that decay according to some convergent series, e.g., $l_j \propto c'Ek^{1/r}/2^j$ or $l_j \propto c'Ek^{1/r}/j^2$.

- The jth message is a random variable $M_j \in \{0,1\}^{l_j}$. In order to pick an m_j according to the right distribution, the sender picks an input $x \sim \widehat{\mathcal{D}}^{(j-1)}$ which, together with the history $m_1, \ldots, m_{j-1}$, determines m_j. The sender transmits m_j to the receiver and promptly forgets x. The sender's new distribution (i.e., $\widehat{\mathcal{D}}^{(j-1)}$, conditioned on $M_j = m_j$) is called $\mathcal{D}^{(j)}$.
- The distribution $\mathcal{D}^{(j)}$ may reveal information about the coordinates I_{j-1} in an irregular fashion. We find a subset $I_j \subset I_{j-1}$ of coordinates, $|I_j| = k^{1-j/r}$, for which the amount of information revealed by $\mathcal{D}^{(j)}_{I_j}$ is at most average. The receiver of m_j changes his input distribution to $\widehat{\mathcal{D}}^{(j)}$, which is defined so that it basically agrees with $\mathcal{D}^{(j)}_{I_j}$ and the marginal distribution $\widehat{\mathcal{D}}^{(j)}_{[k]\setminus I_j}$, conditioned on the value selected by $\mathcal{D}^{(j)}_{I_j}$, is identical to $\mathcal{D}^{(j-1)}_{[k]\setminus I_j}$.
- The reason $\mathcal{D}^{(j)}_{I_j}$ and $\widehat{\mathcal{D}}^{(j)}_{I_j}$ are not *identical* is due to two filtering steps. To generate $\widehat{\mathcal{D}}^{(j)}$, we remove points from the support that have tiny (but non-zero) probability, which may be too close to the error probability. Intuitively these rare events necessarily represent a small fraction of the probability mass. Second, we remove points from the support if the ratio of their probability occurring under $\mathcal{D}^{(j)}$ over $\mathcal{D}^{(j-1)}$ is too high. Intuitively, we want to conclude that if there is a high probability of an error occurring under $\mathcal{D}^{(j)}$ then the probability is also high under $\mathcal{D}^{(j-1)}$ (and by unrolling this further, under $\mathcal{D}^{(0)}$). This argument only works if the ratios are what we would expect, given how much information is being revealed about these coordinates by m_j. As a result of these two filtering steps, $\mathcal{D}^{(j)}_{I_j}(x_{I_j})$ and $\widehat{\mathcal{D}}^{(j)}_{I_j}(x_{I_j})$ differ by at most a constant factor, for any particular vector $x_{I_j} \in [t]^{|I_j|}$.

3.2 A Lower Bound on **EqualityTesting**

We begin with two general lemmas about discrete probability distributions that play an important role in our proof.

Roughly speaking, Lemma 3.1 captures and generalizes the following intuition: Suppose p is a *high entropy* distribution on some universe U and q is obtained from p by conditioning on an event $\mathcal{X} \subseteq U$ such that $p(\mathcal{X})$ is large, say some constant like $1/4$. If p's entropy is close to $\log|U|$, then q's entropy should not be much smaller than that of p. As our proof goes on round by round, we will constantly throw away part of the input distribution's support to meet certain conditions. It is Lemma 3.1 that guarantees that the input distributions continue to have relatively high entropy.

Lemma 3.2 comes into play because the error probability will be calculated backward in a round-by-round manner. Suppose the old distribution (p) has no extremely low probability point and the new distribution (q) has almost full entropy. Lemma 3.2 provides us with a useful tool to transfer a lower bound on the probability of any event w.r.t. q to a lower bound on the same event w.r.t. p.

LEMMA 3.1. *Let p and q be distributions defined on a universe of size 2^s. Suppose both of the following properties are satisfied:*

1. *The entropy of p is $\mathrm{H}(p) \geq s - g$, where $0 \leq g \leq s$;*
2. *There exists $0 < \alpha < 1$ such that $q(x) \leq p(x)/\alpha$ holds for every value $x \in \mathrm{supp}(q)$.*

The entropy of q is lower bounded by:

$$\mathrm{H}(q) \geq s - g/\alpha - \mathrm{H}(\alpha)/\alpha.$$

Proof. Let $\mathcal{X}$ be the whole universe. From our assumptions, the entropy of q can be lower bounded as follows.

$$\begin{aligned}
&\mathrm{H}(q)\\
&= \sum_{x\in\mathcal{X}} q(x)\log\frac{1}{q(x)}\\
&= \frac{1}{\alpha}\sum_{x\in\mathcal{X}} \alpha q(x)\log\frac{1}{\alpha q(x)} + \log\alpha\\
&\geq \frac{1}{\alpha}\sum_{x\in\mathcal{X}}\left[p(x)\log\frac{1}{p(x)} - (p(x)-\alpha q(x))\log\frac{1}{p(x)-\alpha q(x)}\right]\\
&\quad + \log\alpha
\end{aligned}$$

The previous step follows from Assumption 2 and the fact that $x\log x^{-1} + y\log y^{-1} \geq (x+y)\log(x+y)^{-1}$ for any $x, y \geq 0$. Continuing,

$$\begin{aligned}
&\geq \frac{1}{\alpha}\left[s - g - \sum_{x\in\mathcal{X}}(p(x)-\alpha q(x))\log\frac{1}{p(x)-\alpha q(x)}\right] + \log\alpha\\
&\geq \frac{1}{\alpha}\left[s - g - (1-\alpha)\log\frac{2^s}{1-\alpha}\right] + \log\alpha\\
&= s - \frac{g}{\alpha} + \frac{1-\alpha}{\alpha}\log(1-\alpha) + \log\alpha\\
&= s - \frac{g}{\alpha} - \frac{\mathrm{H}(\alpha)}{\alpha}.
\end{aligned}$$

∎

LEMMA 3.2. *Let p and q be distributions defined on a universe of size 2^s. Suppose both of the following properties are satisfied:*

1. *The entropy of q is $\mathrm{H}(q) \geq s - g_1$, where $0 \leq g_1 \leq s$;*
2. *There exists $g_2 \geq 0$ such that $p(x) \geq 2^{-s-g_2}$ holds for every value $x \in \mathrm{supp}(q)$.*

Then, for any $0 < \alpha < 1$,

$$\Pr_{x\sim q}\left[\frac{q(x)}{p(x)} > 2^{g_1/\alpha + g_2 - (1-\alpha)\log(1-\alpha)/\alpha}\right] \leq \alpha.$$

Proof. Let $\mathcal{X}_0 = \{x \in \mathrm{supp}(q) \mid q(x)/p(x) \leq 2^{g_1/\alpha + g_2 - (1-\alpha)\log(1-\alpha)/\alpha}\}$ and $\mathcal{X}_1 = \mathrm{supp}(q) \setminus \mathcal{X}_0$. Suppose, for the purpose of obtaining a contradiction, that the conclusion of the lemma is false, i.e., $q(\mathcal{X}_1) = \alpha_0$, for some $\alpha_0 > \alpha$. Notice that for each value $x \in \mathcal{X}_1$, Assumption 2 implies that

$$\begin{aligned}
q(x) &> p(x)\cdot 2^{g_1/\alpha + g_2 - (1-\alpha)\log(1-\alpha)/\alpha}\\
&\geq 2^{-s+g_1/\alpha - (1-\alpha)\log(1-\alpha)/\alpha}.
\end{aligned} \tag{3.1}$$

Then we can upper bound the entropy of q as follows.

$$\begin{aligned}
&\mathrm{H}(q)\\
&= \sum_{x\in\mathcal{X}_0} q(x)\log\frac{1}{q(x)} + \sum_{x\in\mathcal{X}_1} q(x)\log\frac{1}{q(x)}\\
&< \sum_{x\in\mathcal{X}_0} q(x)\log\frac{1}{q(x)} + \alpha_0\left[s - \frac{g_1}{\alpha} + \frac{1-\alpha}{\alpha}\log(1-\alpha)\right]\\
&\le (1-\alpha_0)\log\frac{2^s}{1-\alpha_0} + \alpha_0\left[s - \frac{g_1}{\alpha} + \frac{1-\alpha}{\alpha}\log(1-\alpha)\right]\\
&= s - \frac{\alpha_0}{\alpha}\cdot g_1 + \alpha_0\left[\frac{1-\alpha}{\alpha}\log(1-\alpha) - \frac{1-\alpha_0}{\alpha_0}\log(1-\alpha_0)\right]\\
&< s - g_1,
\end{aligned}$$

where the last step follows from the monotonicity of $(1-\alpha)\log(1-\alpha)/\alpha$. This contradicts Assumption 1. ∎

We are now ready to begin the proof of Theorem 3.1 proper. Fix a round j and a particular history $(m_1,\ldots,m_{j-1})$ up to round $j-1$. We let $\mu_j(m_j)$ denote the probability that the jth message is m_j, if the input to the sender is drawn from $\widehat{\mathcal{D}}^{(j-1)}$. Define $\mathcal{D}^{(j)}[m_j]$ to be the new input distribution of the sender after he commits to m_j. When m_j is clear from context, it is denoted $\mathcal{D}^{(j)}$. (The process for deriving $\widehat{\mathcal{D}}^{(j)}$ from $\mathcal{D}^{(j)}$ and $\mathcal{D}^{(j-1)}$ on the receiver's end will be explained in detail later.)

We will prove by induction that the following Invariant 3.3 holds for each $j\in[0,r]$, where the particular values of I_j, $\mathcal{D}^{(j)}$, $\widehat{\mathcal{D}}^{(j)}$, and $l_1,\ldots,l_j$ depend on the transcript $m_1,\ldots,m_j$ that is sampled. In the base case, Invariant 3.3 clearly holds when $j=0, I_0=[k]$, and both $\widehat{\mathcal{D}}^{(0)},\mathcal{D}^{(0)}$ are the uniform distribution over $[t]^k$.

Invariant 3.3. *After round $j\in[0,r]$ the partial transcript is $m_1,\ldots,m_j$, which determines the values $\{l_{j'},\widehat{\mathcal{D}}^{(j')},\mathcal{D}^{(j')},I_{j'}\}_{j'\le j}$. The index set $I_j\subseteq[k]$ satisfies all of the following:*

1. *$|I_j| = k^{1-j/r}$.*
2. *Each value $x_{I_j}\in[t]^{|I_j|}$ satisfies $\widehat{\mathcal{D}}^{(j)}_{I_j}(x_{I_j}) \le 4\mathcal{D}^{(j)}_{I_j}(x_{I_j})$.*
3. *Each nonempty subset $I'\subseteq I_j$ satisfies*

$$\mathrm{H}(\widehat{\mathcal{D}}^{(j)}_{I'}) \ge \left(cE - \sum_{u=1}^{j}\frac{16^{j-u+1}l_u}{k^{1-(u-1)/r}} - 22^j\right)|I'|.$$

In accordance with our informal discussion in Section 3.1, I_j is a subset of indices on which both parties have learned little information about each other from the partial transcript $m_1,\ldots,m_j$. Invariant 3.3(2) ensures that the two parties draw their inputs after the jth round from similar distributions. Invariant 3.3(3) is the most important property. It says that the information revealed by $\widehat{\mathcal{D}}^{(j)}$ about I' is roughly what one would expect, given the message lengths $l_1,\ldots,l_j$. Note that the uth message conveys information about $|I_{u-1}| = k^{1-(u-1)/r}$ indices so the average information-per-index should be $l_u/k^{1-(u-1)/r}$. The factor 16^{j-u+1} and the extra term 22^j come from Lemma 3.1, which throws away part of the input distribution in each round, progressively distorting the distributions in minor ways.

To begin our induction, at round j we find a large fraction of possible messages m_j that reveal little information about the sender's input, projected onto I_{j-1}. This is possible because the length of the message $l_j=|m_j|$ reflects an upper bound on the expected information gain. This idea is formalized in the following Lemma 3.4.

Lemma 3.4. *Fix $j\in[1,r]$ and suppose Invariant 3.3 holds for $j-1$. Then there exists a subset of messages $\mathcal{M}'_j$ with $\mu_j(\mathcal{M}'_j)\ge 1/2$ such that each message $m_j\in\mathcal{M}'_j$ satisfies*

$$\mathrm{H}(\mathcal{D}^{(j)}_{I_{j-1}}[m_j]) \ge \left(cE - 2\sum_{u=1}^{j}\frac{16^{j-u}l_u}{k^{1-(u-1)/r}} - 2\cdot 22^{j-1}\right)|I_{j-1}|.$$

Proof. Let $\mathcal{M}'_j$ contain all messages m_j satisfying the above inequality and $\overline{\mathcal{M}'_j}$ be its complement. Suppose, for the purpose of obtaining a contradiction, that the conclusion of the lemma is not true, i.e., $\mu_j(\overline{\mathcal{M}'_j})=\alpha>1/2$. Then the entropy of $\widehat{\mathcal{D}}^{(j-1)}_{I_{j-1}}$ can be upper bounded as follows.

$$\begin{aligned}
&\mathrm{H}(\widehat{\mathcal{D}}^{(j-1)}_{I_{j-1}})\\
&= \mathrm{I}(\widehat{\mathcal{D}}^{(j-1)}_{I_{j-1}}\,;\,M_j) + \sum_{m_j\in(\mathcal{M}'_j\cup\overline{\mathcal{M}'_j})}\mu_j(m_j)\mathrm{H}(\mathcal{D}^{(j)}_{I_{j-1}}[m_j])\\
&\le \mathrm{H}(M_j) + \sum_{m_j\in(\mathcal{M}'_j\cup\overline{\mathcal{M}'_j})}\mu_j(m_j)\mathrm{H}(\mathcal{D}^{(j)}_{I_{j-1}}[m_j])\\
&\le l_j + \sum_{m_j\in\mathcal{M}'_j}\mu_j(m_j)\mathrm{H}(\mathcal{D}^{(j)}_{I_{j-1}}[m_j])\\
&\quad + \sum_{m_j\in\overline{\mathcal{M}'_j}}\mu_j(m_j)\mathrm{H}(\mathcal{D}^{(j)}_{I_{j-1}}[m_j])\\
&< l_j + (1-\alpha)cE|I_{j-1}|\\
&\quad + \alpha\left(cE - 2\sum_{u=1}^{j}\frac{16^{j-u}l_u}{k^{1-(u-1)/r}} - 2\cdot 22^{j-1}\right)|I_{j-1}|\\
&= l_j + \left(cE - 2\alpha\sum_{u=1}^{j}\frac{16^{j-u}l_u}{k^{1-(u-1)/r}} - 2\alpha\cdot 22^{j-1}\right)|I_{j-1}|\\
&< \left(cE - \sum_{u=1}^{j-1}\frac{16^{j-u}l_u}{k^{1-(u-1)/r}} - 22^{j-1}\right)|I_{j-1}|
\end{aligned}$$

This contradicts Invariant 3.3(3) at index $j-1$. ∎

After the jth message m_j is sent, the next step is to identify a set of coordinates I_j such that $\mathcal{D}^{(j)}$ still reveals little information about I_j *and every subset of* I_j, since we need this property to hold for $I_{j+1},\ldots,I_r$ in the future, all of which are subsets of I_j. We also want I_j not to contain many low probability points w.r.t. $\mathcal{D}^{(j-1)}$, since this may stop us from applying Lemma 3.2 later on. These two constraints are captured by parts (2) and (1), respectively, of Lemma 3.5.

LEMMA 3.5. *Fix $j \in [1, r]$ and suppose Invariant 3.3 holds for $j-1$. Then there exists a subset of messages $\mathcal{M}_j \subseteq \mathcal{M}'_j$ (from Lemma 3.4) with $\mu_j(\mathcal{M}_j) \geq 1/4$ such that for each message $m_j \in \mathcal{M}_j$, there exists a subset $I_j \subseteq I_{j-1}$ of size $|I_j| = k^{1-j/r}$ satisfying both of the following properties:*

1. $\Pr_{x_{I_j} \sim \mathcal{D}^{(j)}_{I_j}} \left[\mathcal{D}^{(j-1)}_{I_j}(x_{I_j}) < (4t)^{-|I_j|}/32\right] \leq 1/2$;
2. *Each nonempty subset $I' \subseteq I_j$ satisfies*

$$\mathrm{H}(\mathcal{D}^{(j)}_{I'}) \geq \left(cE - 4\sum_{u=1}^{j} \frac{16^{j-u} l_u}{k^{1-(u-1)/r}} - 4 \cdot 22^{j-1}\right)|I'|.$$

Proof. We first prove that for each message $m_j \in \mathcal{M}'_j$ (from Lemma 3.4), there exists a subset $J_0 \subseteq I_{j-1}$ of size $|J_0| \geq |I_{j-1}|/2$ such that each nonempty subset $I' \subseteq J_0$ satisfies part (2) of the lemma. Suppose $J_1, J_2, \ldots, J_w$ are disjoint subsets of I_{j-1}, each of which *violates* the inequality of part (2), whereas none of the subsets of $J_0 = I_{j-1} \setminus (\bigcup_{v=1}^{w} J_v)$ do. Then we can upper bound the entropy of $\mathcal{D}^{(j)}_{I_{j-1}}$ as follows.

$$\begin{aligned}
&\mathrm{H}(\mathcal{D}^{(j)}_{I_{j-1}})\\
&\leq \sum_{v=0}^{w} \mathrm{H}(\mathcal{D}^{(j)}_{J_v})\\
&< cE|J_0| + \sum_{v=1}^{w}\left(cE - 4\sum_{u=1}^{j} \frac{16^{j-u} l_u}{k^{1-(u-1)/r}} - 4 \cdot 22^{j-1}\right)|J_v|\\
&= cE|I_{j-1}| - 4|I_{j-1} \setminus J_0|\left(\sum_{u=1}^{j} \frac{16^{j-u} l_u}{k^{1-(u-1)/r}} + 22^{j-1}\right).
\end{aligned}$$

On the other hand, from Lemma 3.4, having $m_j \in \mathcal{M}'_j$ guarantees that

$$\mathrm{H}(\mathcal{D}^{(j)}_{I_{j-1}}) \geq \left(cE - 2\sum_{u=1}^{j} \frac{16^{j-u} l_u}{k^{1-(u-1)/r}} - 2 \cdot 22^{j-1}\right)|I_{j-1}|.$$

The two inequalities above are only consistent if $|I_{j-1} \setminus J_0| \leq |I_{j-1}|/2$, or equivalently $|J_0| \geq |I_{j-1}|/2$. Thus, J_0 exists with the right cardinality, as claimed.

Now suppose, for the purpose of obtaining a contradiction, that the lemma is false. For every $m_j \in \mathcal{M}'_j$ there is a corresponding index set J_0 whose subsets satisfy part (2) of the lemma. If the lemma is false, that means there is a subset $\mathcal{M}''_j \subseteq \mathcal{M}'_j$ of "bad" messages with $\mu_j(\mathcal{M}''_j) > 1/4$ such that, for each $m_j \in \mathcal{M}''_j$, none of the $\binom{|J_0|}{|I_j|}$ choices for $I_j \subseteq J_0$ satisfy part (1) of the lemma. (Remember that J_0 depends on m_j but the lower bound on $|J_0| \geq |I_{j-1}|/2$ is independent of m_j.) Consider the following summation:

$$Z = \sum_{\substack{I_j \subseteq I_{j-1}:\\ |I_j| = k^{1-j/r}}} \sum_{\substack{x_{I_j} \in [t]^{|I_j|}:\\ \mathcal{D}^{(j-1)}_{I_j}(x_{I_j}) < (4t)^{-|I_j|}/32}} \mathcal{D}^{(j-1)}_{I_j}(x_{I_j}).$$

We can easily upper bound Z as follows.

$$Z < \binom{|I_{j-1}|}{|I_j|} \cdot t^{|I_j|} \cdot \frac{(4t)^{-|I_j|}}{32} = \binom{|I_{j-1}|}{|I_j|} 2^{-2|I_j|-5}.$$

Invariant 3.3(2) relates $\mathcal{D}^{(j-1)}$ and $\widehat{\mathcal{D}}^{(j-1)}$, which lets us lower bound Z.

$$Z \geq \frac{1}{4} \sum_{\substack{I_j \subseteq I_{j-1}:\\ |I_j| = k^{1-j/r}}} \sum_{\substack{x_{I_j} \in [t]^{|I_j|}:\\ \mathcal{D}^{(j-1)}_{I_j}(x_{I_j}) < (4t)^{-|I_j|}/32}} \widehat{\mathcal{D}}^{(j-1)}_{I_j}(x_{I_j})$$

By definition, $\widehat{\mathcal{D}}^{(j-1)}$ is a convex combination of the $\mathcal{D}^{(j)}[m_j]$ distributions, weighted according to $\mu_j(\cdot)$. Hence, the expression above is lower bounded by

$$\begin{aligned}
&\geq \frac{1}{4} \sum_{\substack{I_j \subseteq I_{j-1}:\\ |I_j| = k^{1-j/r}}} \sum_{\substack{x_{I_j} \in [t]^{|I_j|}:\\ \mathcal{D}^{(j-1)}_{I_j}(x_{I_j})\\ < (4t)^{-|I_j|}/32}} \sum_{m_j \in \mathcal{M}''_j} \mu_j(m_j) \cdot \mathcal{D}^{(j)}_{I_j}[m_j](x_{I_j})\\
&\geq \frac{1}{4} \sum_{m_j \in \mathcal{M}''_j} \mu_j(m_j) \sum_{\substack{I_j \subseteq J_0:\\ |I_j| = k^{1-j/r}}} \sum_{\substack{x_{I_j} \in [t]^{|I_j|}:\\ \mathcal{D}^{(j-1)}_{I_j}(x_{I_j})\\ < (4t)^{-|I_j|}/32}} \mathcal{D}^{(j)}_{I_j}[m_j](x_{I_j})
\end{aligned}$$

By definition, for every $m_j \in \mathcal{M}''_j$ and every choice of $I_j \subseteq J_0$, part (1) of the lemma is violated. Continuing with the inequalities,

$$\begin{aligned}
&> \frac{1}{4} \sum_{m_j \in \mathcal{M}''_j} \mu_j(m_j) \cdot \binom{|J_0|}{|I_j|} \cdot \frac{1}{2}\\
&> \frac{1}{32} \binom{|I_{j-1}|/2}{|I_j|}.
\end{aligned}$$

This contradicts the upper bound on Z whenever $k^{1/r}$ is at least some sufficiently large constant. $\square$

The receiver of m_j constructs a new distribution $\widehat{\mathcal{D}}^{(j)}$ in two steps. After fixing I_j, we construct $\widetilde{\mathcal{D}}^{(j)}$ by combining $\mathcal{D}^{(j-1)}$ and $\mathcal{D}^{(j)}$, filtering out some points in the space whose probability mass is too low. We then construct $\widehat{\mathcal{D}}^{(j)}$ from $\widetilde{\mathcal{D}}^{(j)}$ and $\mathcal{D}^{(j-1)}$ by filtering out points that occur under $\widetilde{\mathcal{D}}^{(j)}$ with substantially larger probability than they do under $\mathcal{D}^{(j-1)}$.

Formally, suppose Invariant 3.3 holds for $j-1$. For each message $m_j \in \mathcal{M}_j$ (from Lemma 3.5), let I_j be selected to satisfy both properties of Lemma 3.5. Define the probability mass of a vector $x \in [t]^k$ under $\widetilde{\mathcal{D}}^{(j)}$ as follows:

$$\widetilde{\mathcal{D}}^{(j)}(x) = \begin{cases} 0 & \text{if } \mathcal{D}^{(j-1)}_{I_j}(x_{I_j}) < \frac{(4t)^{-|I_j|}}{32};\\ \frac{\mathcal{D}^{(j)}_{I_j}(x_{I_j})}{\beta_1} \cdot \frac{\mathcal{D}^{(j-1)}(x)}{\mathcal{D}^{(j-1)}_{I_j}(x_{I_j})} & \text{otherwise.} \end{cases}$$

where β_1 is

$$\beta_1 = \Pr_{x_{I_j} \sim \mathcal{D}^{(j)}_{I_j}} \left[\mathcal{D}^{(j-1)}_{I_j}(x_{I_j}) \geq \frac{(4t)^{-|I_j|}}{32}\right].$$

In other words, we discard a $1-\beta_1$ fraction of the distribution $\mathcal{D}^{(j)}$, but ignoring this effect, the projection of $\widetilde{\mathcal{D}}^{(j)}$ onto I_j has the same distribution as $\mathcal{D}^{(j)}$ onto I_j, and conditioned on the value of x_{I_j}, the distribution $\widetilde{\mathcal{D}}^{(j)}$ (projected onto $[k]\backslash I_j$) is identical to $\mathcal{D}^{(j-1)}$. We derive $\widehat{\mathcal{D}}^{(j)}$ from $\widetilde{\mathcal{D}}^{(j)}$ with a similar transformation.

$$\widehat{\mathcal{D}}^{(j)}(x) = \begin{cases} 0, & \text{if } \frac{\widetilde{\mathcal{D}}^{(j)}_{I_j}(x_{I_j})}{\mathcal{D}^{(j-1)}_{I_j}(x_{I_j})} > 2^{\gamma_j}; \\ \frac{\widetilde{\mathcal{D}}^{(j)}_{I_j}(x_{I_j})}{\beta_2} \cdot \frac{\mathcal{D}^{(j-1)}(x)}{\mathcal{D}^{(j-1)}_{I_j}(x_{I_j})}, & \text{otherwise.} \end{cases}$$

where β_2 and γ_j are defined to be

$$\beta_2 = \Pr_{x_{I_j} \sim \widetilde{\mathcal{D}}^{(j)}_{I_j}} \left[\frac{\widetilde{\mathcal{D}}^{(j)}_{I_j}(x_{I_j})}{\mathcal{D}^{(j-1)}_{I_j}(x_{I_j})} \le 2^{\gamma_j} \right],$$

$$\begin{aligned} \gamma_j &= \sum_{u=1}^{j} l_u \left(\frac{16}{k^{1/r}}\right)^{j-u+1} + (16 \cdot 22^{j-1} + 6)|I_j| + 6 \\ &\le \sum_{u=1}^{j} l_u \left(\frac{16}{k^{1/r}}\right)^{j-u+1} + 22^j|I_j| + 6. \end{aligned}$$

The proofs of Lemmas 3.6 and 3.7 use several simple observations about $\widetilde{\mathcal{D}}^{(j)}$ and $\widehat{\mathcal{D}}^{(j)}$:

First, Lemma 3.5(1) states that $\beta_1 \ge 1/2$. Lemma 3.5(2) lower bounds the entropy of $\mathcal{D}^{(j)}_{I_j}$. We apply Lemma 3.1 to $\mathcal{D}^{(j)}_{I_j}$ and $\widetilde{\mathcal{D}}^{(j)}_{I_j}$ (taking the roles of p and q, respectively) with parameter $\alpha = 1/2 \le \beta_1$, and obtain the following lower bound on the entropy of $\widetilde{\mathcal{D}}^{(j)}_{I_j}$.

$$\mathrm{H}(\widetilde{\mathcal{D}}^{(j)}_{I_j}) \ge \left(cE - 8\sum_{u=1}^{j} \frac{16^{j-u} l_u}{k^{1-(u-1)/r}} - 8 \cdot 22^{j-1} - 2\right)|I_j|.$$

Second, we can then apply Lemma 3.2 to $\mathcal{D}^{(j-1)}_{I_j}$ and $\widetilde{\mathcal{D}}^{(j)}_{I_j}$ (taking the roles of p and q, respectively) with parameters

$$\begin{aligned} g_1 &= 8\sum_{u=1}^{j} \frac{16^{j-u} l_u}{k^{(j-u+1)/r}} + (8 \cdot 22^{j-1} + 2)|I_j|, \\ g_2 &= 2|I_j| + 5, \\ &\text{and } \alpha = 1/2. \end{aligned}$$

Since $g_1/\alpha + g_2 - (1-\alpha)\log(1-\alpha)/\alpha = \gamma_j$, we conclude that $\beta_2 \ge 1 - \alpha = 1/2$. Thus, for each value $x_{I_j} \in \operatorname{supp}(\widehat{\mathcal{D}}^{(j)}_{I_j})$,

$$(3.2) \qquad \widehat{\mathcal{D}}^{(j)}_{I_j}(x_{I_j}) = \frac{\widetilde{\mathcal{D}}^{(j)}_{I_j}(x_{I_j})}{\beta_2} = \frac{\mathcal{D}^{(j)}_{I_j}(x_{I_j})}{\beta_1\beta_2} \le 4\mathcal{D}^{(j)}_{I_j}(x_{I_j}).$$

Lemma 3.6 completes the inductive step by lower bounding the entropy of $\widehat{\mathcal{D}}^{(j)}_{I'}$ for every nonempty subset $I' \subseteq I_j$. To put it another way, it ensures that the values of those coordinates in I_j remain almost completely unknown to both parties.

Lemma 3.6. *Fix $j \in [1, r]$ and suppose Invariant 3.3 holds for $j-1$. Then, for each message $m_j \in \mathcal{M}_j$ (from Lemma 3.5), Invariant 3.3 also holds for j.*

Proof. Due to Lemma 3.5 and Eqn. (3.2), the first two properties of Invariant 3.3 are satisfied. For each nonempty subset $I' \subseteq I_j$, the third property of Invariant 3.3 can be derived from the second property of Lemma 3.5 and an application of Lemma 3.1 to $\mathcal{D}^{(j)}_{I'}$ and $\widehat{\mathcal{D}}^{(j)}_{I'}$ (taking the roles of p and q, respectively) with parameter $\alpha = 1/4$ as follows.

$$\begin{aligned} \mathrm{H}(\widehat{\mathcal{D}}^{(j)}_{I'}) &\ge \left(cE - 16\sum_{u=1}^{j} \frac{16^{j-u} l_u}{k^{1-(u-1)/r}} - 16 \cdot 22^{j-1} - 4\right)|I'| \\ &\ge \left(cE - \sum_{u=1}^{j} \frac{16^{j-u+1} l_u}{k^{1-(u-1)/r}} - 22^j\right)|I'|. \end{aligned}$$

□

Aside from maintaining Invariant 3.3 round by round, another important part of our proof is to compute the error probability. Lemma 3.7 shows how the error probabilities of two consecutive rounds are related after our modification to the protocol. More importantly, it also illustrates the reason to bound the *pointwise* ratio between $\widetilde{\mathcal{D}}^{(j)}_{I_j}$ and $\mathcal{D}^{(j-1)}_{I_j}$.

Lemma 3.7. *Fix a round $j \in [1, r]$ and suppose Invariant 3.3 holds for $j-1$. Fix any specific message $m_j \in \mathcal{M}_j$ (from Lemma 3.5). Define p to be the probability of error, when the protocol begins after round j with the inputs drawn from $\mathcal{D}^{(j)}$ and $\widehat{\mathcal{D}}^{(j)}$, respectively. Then the probability of error is at least $2^{-\gamma_j-1}p$ when the inputs are instead drawn from $\mathcal{D}^{(j)}$ and $\mathcal{D}^{(j-1)}$, respectively.*

Proof. From the definition of $\widehat{\mathcal{D}}^{(j)}$, for each value $x \in \operatorname{supp}(\widehat{\mathcal{D}}^{(j)})$, we have

$$(3.3) \qquad \frac{\widehat{\mathcal{D}}^{(j)}(x)}{\mathcal{D}^{(j-1)}(x)} = \frac{\widetilde{\mathcal{D}}^{(j)}_{I_j}(x_{I_j})}{\beta_2\mathcal{D}^{(j-1)}_{I_j}(x_{I_j})} \le \frac{2^{\gamma_j}}{\beta_2} \le 2^{\gamma_j+1}.$$

This concludes the proof. □

Finally, with all lemmas proved above, we have reached the point to calculate the initial error probability.

Lemma 3.8. *Recall that $c = 1/2, c' = c/100$. Fix any $r \in [1, (\log k)/6]$ and $E \ge 100k^{1-1/r}/c$. Suppose the initial input vectors are drawn independently and uniformly from $[t]^k$, where $t = 2^{cE}$. Then the error probability of the* EqualityTesting *protocol, p_{err}, is greater than 2^{-E}.*

Proof. First suppose Invariant 3.3 holds for r and consider the situation after the final round, where the inputs are drawn from $\mathcal{D}^{(r)}$ and $\widehat{\mathcal{D}}^{(r)}$, respectively. Notice that I_r is a singleton set, so the entropy of $\widehat{\mathcal{D}}^{(r)}_{I_r}$ can be lower bounded as follows.

$$\begin{aligned} \mathrm{H}(\widehat{\mathcal{D}}^{(r)}_{I_r}) &\ge cE - \sum_{u=1}^{r} \frac{16^{r-u+1} l_u}{k^{1-(u-1)/r}} - 22^r \\ &= cE - \frac{16}{k^{1/r}} \sum_{u=1}^{r} l_u \left(\frac{16}{k^{1/r}}\right)^{r-u} - 22^r \\ &\ge cE - \frac{16}{k^{1/r}} \sum_{u=1}^{r} l_u - 22k^{1-1/r} \\ &\ge cE - 16c'E - 22k^{1-1/r} > \frac{cE}{2}. \end{aligned}$$

From the lower bound on the entropy of $\widehat{\mathcal{D}}^{(r)}_{I_r}$, we can easily show that there exists no value x_{I_r} such that $\widehat{\mathcal{D}}^{(r)}_{I_r}(x_{I_r}) = \alpha > 3/4$. If there *were* such a value, then the entropy of $\widehat{\mathcal{D}}^{(r)}_{I_r}$ can also be upper bounded as

$$\begin{aligned}\mathrm{H}(\widehat{\mathcal{D}}^{(r)}_{I_r}) &\le \alpha \log \frac{1}{\alpha} + (1-\alpha)\log\frac{t}{1-\alpha}\\ &< \frac{cE}{4} + \alpha\log\frac{1}{\alpha} + (1-\alpha)\log\frac{1}{1-\alpha}\\ &< \frac{cE}{2},\end{aligned}$$

contradicting the lower bound on $\mathrm{H}(\widehat{\mathcal{D}}^{(r)}_{I_r})$.

After all r rounds of communication, the receiver of the last message has to make the decision on I_r depending only on his own input on I_r. Let $\mathcal{X}_0 \subseteq [t]$ be the subset of values x_{I_r} such that the protocol outputs "not equal" on I_r upon seeing the input x_{I_r} after r rounds of communication, $\mathcal{X}_1 = [t] \setminus \mathcal{X}_0$, and $\beta = \widehat{\mathcal{D}}^{(r)}_{I_r}(\mathcal{X}_0)$. Then, the final error probability is at least

$$\begin{aligned}&\sum_{x_{I_r}\in\mathcal{X}_0} \widehat{\mathcal{D}}^{(r)}_{I_r}(x_{I_r})\mathcal{D}^{(r)}_{I_r}(x_{I_r}) + \sum_{x_{I_r}\in\mathcal{X}_1} \widehat{\mathcal{D}}^{(r)}_{I_r}(x_{I_r})\left(1-\mathcal{D}^{(r)}_{I_r}(x_{I_r})\right)\\ &\ge \frac{1}{4}\sum_{x_{I_r}\in\mathcal{X}_0} \widehat{\mathcal{D}}^{(r)}_{I_r}(x_{I_r})^2 + \frac{1}{4}\sum_{x_{I_r}\in\mathcal{X}_1} \widehat{\mathcal{D}}^{(r)}_{I_r}(x_{I_r}) \sum_{x'_{I_r}\ne x_{I_r}} \widehat{\mathcal{D}}^{(r)}_{I_r}(x'_{I_r})\\ &= \frac{1}{4}\sum_{x_{I_r}\in\mathcal{X}_0} \widehat{\mathcal{D}}^{(r)}_{I_r}(x_{I_r})^2 + \frac{1}{4}\sum_{x_{I_r}\in\mathcal{X}_1} \widehat{\mathcal{D}}^{(r)}_{I_r}(x_{I_r})\left(1-\widehat{\mathcal{D}}^{(r)}_{I_r}(x_{I_r})\right)\\ &\ge \frac{1}{4}\sum_{x_{I_r}\in\mathcal{X}_0} \widehat{\mathcal{D}}^{(r)}_{I_r}(x_{I_r})^2 + \frac{1}{16}\sum_{x_{I_r}\in\mathcal{X}_1} \widehat{\mathcal{D}}^{(r)}_{I_r}(x_{I_r})\\ &\ge \frac{\beta^2}{4t} + \frac{1-\beta}{16} \ \ge\ \frac{1}{4t}.\end{aligned}$$

This result also meets the simple intuition that when the inputs to the two parties are almost uniformly random and no communication is allowed, the best strategy would be guessing "not equal" regardless of the actual input.

Finally, we are ready to transfer the error probability back round by round. From Lemma 3.5 through Lemma 3.7, the error probability w.r.t. $\mathcal{D}^{(j)}$ and $\widehat{\mathcal{D}}^{(j)}$ differs from the error probability w.r.t. $\mathcal{D}^{(j-1)}$ and $\widehat{\mathcal{D}}^{(j-1)}$ by at most a $4\cdot 2^{\gamma_j+1} = 2^{\gamma_j+3}$ factor. In particular, Lemma 3.5 and Lemma 3.6 say that the jth message m_j satisfies Invariant 3.3 at index j with probability at least $1/4$, provided Invariant 3.3 holds for $j-1$, and Lemma 3.7 says the error probabilities under the two measures differ by a 2^{γ_j+1} factor for any such m_j. Repeating this for each $j \in [1, r]$, we conclude that the initial error probability p_{err} is lower bounded by

$$\begin{aligned}p_{\mathrm{err}} &\ge \frac{1}{4t}\cdot\exp\left(-3r - \sum_{j=1}^{r}\gamma_j\right)\\ &= \exp\left(-cE - 2 - 3r - \sum_{j=1}^{r}\gamma_j\right)\\ &> 2^{-E},\end{aligned}$$

since

$$\begin{aligned}&cE + 2 + 3r + \sum_{j=1}^{r}\gamma_j\\ &\le cE + 2 + 3r + 6r + \sum_{j=1}^{r}\sum_{u=1}^{j} l_u\left(\frac{16}{k^{1/r}}\right)^{j-u+1}\\ &\quad + \sum_{j=1}^{r} 22^j |I_j|\\ &\le cE + 11r + \sum_{u=1}^{r}\frac{16 l_u}{k^{1/r}}\sum_{j=u}^{r}\left(\frac{16}{k^{1/r}}\right)^{j-u}\\ &\quad + 22k^{1-1/r}\sum_{j=1}^{r}\left(\frac{22}{k^{1/r}}\right)^{j-1}\\ &\le cE + 11r + \frac{32}{k^{1/r}}\sum_{u=1}^{r} l_u + 44k^{1-1/r}\\ &\le cE + \frac{11cE}{100} + \frac{32cE}{100} + \frac{44cE}{100} \ <\ E.\end{aligned}$$

∎

Proof. [Proof of Theorem 3.1] Lemma 3.8 actually shows that given integers $k \ge 1$ and $r \le (\log k)/6$, any r-round deterministic protocol for EqualityTesting on vectors of length k that has distributional error probability $p_{\mathrm{err}} = 2^{-E}$ with respect to the uniform input distribution on $[t]^k$, where $t = 2^{cE}$, requires at least $\Omega(Ek^{1/r})$ bits of communication. Notice that the additional assumption $E \ge 100k^{1-1/r}/c$ always makes sense since there is a trivial $\Omega(k)$ lower bound on the communication complexity of EqualityTesting, regardless of r. Thus, Theorem 3.1 follows directly from Yao's minimax principle. ∎

3.3 A Lower Bound on ExistsEqual

The proof of Theorem 3.2 is almost the same as that of Theorem 3.1, except for the final step, namely Lemma 3.8, in which we first compute the final error probability after all r rounds of communication and then transfer it backward round by round using Lemma 3.7. The problem with applying the same argument to ExistsEqual protocols is that the receiver of the last message may be able to announce the correct answer, even though it knows little information about the inputs on the single coordinate I_r.

In order to prove Theorem 3.2, first notice that Lemma 3.4 through Lemma 3.7 also hold perfectly well for ExistsEqual protocols as no modification is required in their proofs. Therefore, it is sufficient to prove the following Lemma 3.9, which is an analog of Lemma 3.8 for ExistsEqual. It is based mainly on Markov's inequality.

LEMMA 3.9. *Recall that $c = 1/2, c' = c/100$. Consider an execution of a deterministic r-round ExistsEqual protocol, $r \in [1, (\log k)/6]$, on input vectors drawn independently and uniformly from $[t]^k$, where $t = 2^{cE}$. Here $E \ge 100k^{1-1/r}/c$ if $r > 1$ and $E \ge (100\log k)/c$ otherwise. Then the protocol errs with probability $p_{\mathrm{err}} > 2^{-E}$.*

Proof. Similarly to the proof of Lemma 3.8, we first consider the situation after the final round. In the ExistsEqual protocol, the receiver of the last message can make the decision depending on every coordinate of his own input. Let $\mathcal{X}_0 \subseteq [t]^k$ be the subset of values x such that the protocol outputs "no" upon seeing the input x after r rounds of communication, $\mathcal{X}_1 = [t]^k \setminus \mathcal{X}_0$. Then, the final error probability is at least

$$\sum_{x\in\mathcal{X}_0} \widehat{\mathcal{D}}^{(r)}(x)\mathcal{D}^{(r)}_{I_r}(x_{I_r}) + \sum_{x\in\mathcal{X}_1} \widehat{\mathcal{D}}^{(r)}(x)\left(1-\sum_{y\in\mathcal{N}(x)}\mathcal{D}^{(r)}(y)\right),$$

where

$$\mathcal{N}(x) = \{y \in [t]^k \mid \text{there exists some } i\in[k] \text{ such that } x_i = y_i\}$$

is the subset of input vectors that agree with x on at least one coordinate.

The main difficulty here is to lower bound $1 - \sum_{y\in\mathcal{N}(x)}\mathcal{D}^{(r)}(y)$, which is potentially quite small. Consider the following summation Z_0 over *all* transcripts $m_1,\ldots,m_r$ in which $m_j \in \mathcal{M}_j$ (from Lemma 3.5), where the set $\mathcal{M}_j$ depends on $m_1,\ldots,m_{j-1}$:

$$Z_0 = \sum_{m_1\in\mathcal{M}_1}\mu_1(m_1)\sum_{m_2\in\mathcal{M}_2}\mu_2(m_2)\cdots \sum_{m_r\in\mathcal{M}_r}\mu_r(m_r)\sum_{x\in[t]^k}\widehat{\mathcal{D}}^{(r)}(x)\sum_{y\in\mathcal{N}(x)}\mathcal{D}^{(r)}(y).$$

From the proof of Lemma 3.7 (Eqn. (3.3)), we can upper bound Z_0 as follows.

$$Z_0 \le \sum_{m_1\in\mathcal{M}_1}\mu_1(m_1)\cdots\sum_{m_r\in\mathcal{M}_r}\mu_r(m_r) \sum_{\substack{x\in[t]^k,\\ y\in\mathcal{N}(x)}} 2^{\gamma_r+1}\cdot\mathcal{D}^{(r-1)}(x)\cdot\mathcal{D}^{(r)}(y)$$

Notice that γ_r and $\mathcal{D}^{(r-1)}$ are independent of the choice of m_r, hence by rearranging sums, this is equal to

$$= \sum_{m_1\in\mathcal{M}_1}\mu_1(m_1)\cdots\sum_{m_{r-1}\in\mathcal{M}_{r-1}}\mu_{r-1}(m_{r-1}) \sum_{\substack{x\in[t]^k,\\ y\in\mathcal{N}(x)}} 2^{\gamma_r+1}\cdot\mathcal{D}^{(r-1)}(x)\sum_{m_r\in\mathcal{M}_r}\mu_r(m_r)\cdot\mathcal{D}^{(r)}(y)$$

By definition, $\widehat{\mathcal{D}}^{(r-1)}$ is a convex combination of the $\mathcal{D}^{(r)}[m_r]$ distributions, weighted according to $\mu_r(\cdot)$. Hence, the expression above is upper bounded by

$$\le \sum_{m_1\in\mathcal{M}_1}\mu_1(m_1)\cdots\sum_{m_{r-1}\in\mathcal{M}_{r-1}}\mu_{r-1}(m_{r-1}) \sum_{\substack{x\in[t]^k,\\ y\in\mathcal{N}(x)}} 2^{\gamma_r+1}\cdot\mathcal{D}^{(r-1)}(x)\cdot\widehat{\mathcal{D}}^{(r-1)}(y)$$

By the symmetry of x and y, this is equal to

$$= \sum_{m_1\in\mathcal{M}_1}\mu_1(m_1)\cdots\sum_{m_{r-1}\in\mathcal{M}_{r-1}}\mu_{r-1}(m_{r-1}) \sum_{\substack{x\in[t]^k,\\ y\in\mathcal{N}(x)}} 2^{\gamma_r+1}\cdot\widehat{\mathcal{D}}^{(r-1)}(x)\cdot\mathcal{D}^{(r-1)}(y)$$

We repeat the same argument for rounds $r-1$ down to 1, upper bounding Z_0 by

$$\le \exp\left(r+\sum_{j=1}^r\gamma_j\right)\cdot\sum_{\substack{x\in[t]^k,\\ y\in N(x)}}\widehat{\mathcal{D}}^{(0)}(x)\cdot\mathcal{D}^{(0)}(y)$$
$$\le \exp\left(r+\sum_{j=1}^r\gamma_j\right)\cdot\frac{k}{t}$$

The last inequality above follows from a union bound since, under the initial distributions $\widehat{\mathcal{D}}^{(0)},\mathcal{D}^{(0)}$, each of the k coordinates is equal with probability $1/t$. Recall that $E \ge 100k^{1-1/r}/c$ when $r>1$ and $E \ge (100\log k)/c$ otherwise. Hence, using the same argument as that in the proof of Lemma 3.8, we can further bound this as

$$\le 2^{0.83cE}\cdot 2^{0.02cE}\cdot 2^{-cE} \;=\; 2^{-0.15cE},$$

since

$$\begin{aligned} r+\sum_{j=1}^r\gamma_j &\le 7r+\sum_{j=1}^r\sum_{u=1}^j l_u\left(\frac{16}{k^{1/r}}\right)^{j-u+1}+\sum_{j=1}^r 22^j|I_j| \\ &\le \frac{7cE}{100}+\frac{32cE}{100}+\frac{44cE}{100}=\frac{83cE}{100}, \end{aligned}$$

and $k \le (cE/100)^{r/(r-1)} \le (cE/100)^2 \le 2^{0.02cE}$ when $r>1$ and $k \le 2^{0.01cE}$ otherwise.

Now fix a round j and a particular history $(m_1,\ldots,m_j)$ up to round j such that $m_{j'}\in\mathcal{M}_{j'}$ holds for every $j'\le j$. Define Z_j as follows.

$$Z_j = \sum_{m_{j+1}\in\mathcal{M}_{j+1}}\mu_{j+1}(m_{j+1})\cdots\sum_{m_r\in\mathcal{M}_r}\mu_r(m_r) \sum_{x\in[t]^k}\widehat{\mathcal{D}}^{(r)}(x)\sum_{y\in\mathcal{N}(x)}\mathcal{D}^{(r)}(y).$$

By Markov's inequality, there exists a subset of messages $\widehat{\mathcal{M}}_1\subseteq\mathcal{M}_1$ with $\mu_1(\widehat{\mathcal{M}}_1)\ge\mu_1(\mathcal{M}_1)/2\ge 1/8$ such that each message $m_1\in\widehat{\mathcal{M}}_1$ satisfies $Z_1\le 2Z_0/\mu_1(\mathcal{M}_1)\le 8Z_0$ since $\mu_1(\mathcal{M}_1)\ge 1/4$ from Lemma 3.5. Similarly, conditioned on any specific $m_1\in\widehat{\mathcal{M}}_1$, by Markov's inequality, there exists a subset of messages $\widehat{\mathcal{M}}_2\subseteq\mathcal{M}_2$ with $\mu_2(\widehat{\mathcal{M}}_2)\ge\mu_2(\mathcal{M}_2)/2\ge 1/8$ such that each message $m_2\in\widehat{\mathcal{M}}_2$ satisfies $Z_2\le 2Z_1/\mu_2(\mathcal{M}_2)\le 8^2Z_0$. In general, conditioned on any specific partial transcript $m_1,\ldots,m_{j-1}$ such that $m_{j'}\in\widehat{\mathcal{M}}_{j'}$

holds for every $j' < j$, there exists a subset of messages $\widehat{\mathcal{M}}_j \subseteq \mathcal{M}_j$ with $\mu_j(\widehat{\mathcal{M}}_j) \geq \mu_j(\mathcal{M}_j)/2 \geq 1/8$ such that each message $m_j \in \widehat{\mathcal{M}}_j$ satisfies $Z_j \leq 8^j Z_j$.

After repeating the same argument r times, we get $\widehat{\mathcal{M}}_1, \ldots, \widehat{\mathcal{M}}_r$ in sequence. For any sampled transcript $m_1, \ldots, m_r$ such that $m_j \in \widehat{\mathcal{M}}_j$ for all $j \leq r$, we have

$$Z_r \leq 8^r Z_0 \leq 2^{3r} \cdot 2^{-0.15cE} \leq 2^{-0.12cE} \leq \frac{1}{4},$$

as $r \leq cE/100$ and $cE \geq 100$. Further, one more application of Markov's inequality shows that there exists a subset of values $\mathcal{X}' \subseteq [t]^k$ with $\widehat{\mathcal{D}}^{(r)}(\mathcal{X}') = \alpha \geq 1/2$ such that $\sum_{y \in \mathcal{N}(x)} \mathcal{D}^{(r)}(y) \leq 1/2$ holds for every $x \in \mathcal{X}'$.

As a result, we can then lower bound the final error probability as follows, where $\beta = \widehat{\mathcal{D}}^{(r)}(\mathcal{X}_0 \cap \mathcal{X}')$.

$$\begin{aligned}
&\sum_{x\in\mathcal{X}_0} \widehat{\mathcal{D}}^{(r)}(x)\mathcal{D}^{(r)}_{I_r}(x_{I_r}) + \sum_{x\in\mathcal{X}_1} \widehat{\mathcal{D}}^{(r)}(x)\left(1 - \sum_{y\in\mathcal{N}(x)} \mathcal{D}^{(r)}(y)\right)\\
&\geq \sum_{x\in(\mathcal{X}_0\cap\mathcal{X}')} \widehat{\mathcal{D}}^{(r)}(x)\mathcal{D}^{(r)}_{I_r}(x_{I_r})\\
&\qquad + \sum_{x\in(\mathcal{X}_1\cap\mathcal{X}')} \widehat{\mathcal{D}}^{(r)}(x)\left(1 - \sum_{y\in\mathcal{N}(x)} \mathcal{D}^{(r)}(y)\right)\\
&\geq \frac{1}{4}\sum_{x\in(\mathcal{X}_0\cap\mathcal{X}')} \widehat{\mathcal{D}}^{(r)}(x)\widehat{\mathcal{D}}^{(r)}_{I_r}(x_{I_r})\\
&\qquad + \sum_{x\in(\mathcal{X}_1\cap\mathcal{X}')} \widehat{\mathcal{D}}^{(r)}(x)\left(1 - \sum_{y\in\mathcal{N}(x)} \mathcal{D}^{(r)}(y)\right)\\
&\geq \frac{1}{4}\sum_{x\in(\mathcal{X}_0\cap\mathcal{X}')} \widehat{\mathcal{D}}^{(r)}(x)\widehat{\mathcal{D}}^{(r)}_{I_r}(x_{I_r}) + \frac{1}{2}\sum_{x\in(\mathcal{X}_1\cap\mathcal{X}')} \widehat{\mathcal{D}}^{(r)}(x)
\end{aligned}$$

In order to minimize the above expression, we can now assume without loss of generality that the partition between $\mathcal{X}_0 \cap \mathcal{X}'$ and $\mathcal{X}_1 \cap \mathcal{X}'$ depends solely on x_{I_r} as only the relative magnitude of $\widehat{\mathcal{D}}^{(r)}_{I_r}(x_{I_r})/4$ and $1/2$ matters. Continuing,

$$\geq \frac{\beta^2}{4t} + \frac{\alpha - \beta}{2} \;\geq\; \frac{\alpha^2}{4t} \;\geq\; \frac{1}{16t}.$$

Finally, we are ready to transfer the error probability back in exactly the same manner as we did in the proof of Lemma 3.8. Using a similar argument, the existence of $\widehat{\mathcal{M}}_j$ guarantees that

$$\begin{aligned}
p_{\text{err}} &\geq \frac{1}{16t} \cdot \exp\left(-4r - \sum_{j=1}^{r} \gamma_j\right)\\
&= \exp\left(-cE - 4 - 4r - \sum_{j=1}^{r} \gamma_j\right) > 2^{-E},
\end{aligned}$$

since

$$cE + 4 + 4r + \sum_{j=1}^{r} \gamma_j \leq cE + \frac{14cE}{100} + \frac{32cE}{100} + \frac{44cE}{100} \;<\; E.$$

□

Proof. [Proof of Theorem 3.2] Similarly to the proof of Theorem 3.1, Theorem 3.2 follows from Lemma 3.9 and a direct application of Yao's minimax principle. □

4 Upper Bounds on EqualityTesting and ExistsEqual

In this section, we prove upper bounds on both EqualityTesting and ExistsEqual. We first give a $(\log^*(k/E) + r)$-round EqualityTesting protocol (Theorem 4.1) that uses $O(k + rEk^{1/r})$ bits of communication and errs with probability at most $p_{\text{err}} = 2^{-E}$. The $\log^*(k/E)$ term cannot be completely eliminated, due to the lower bounds of [ST13, BCK⁺16]. Our lower bound implies that when $E \geq k$ (so $\log^*(k/E) = 0$), the second term is optimal up to a factor of r.

A natural goal is to achieve optimal communication $\Theta(k+E)$ and minimize the number of rounds subject to that constraint. When $E \geq k$ our lower bound says $r = \Omega(\log k)$, but in this case the algorithm of Theorem 4.1 only achieves $O(E \log k)$ communication. Theorems 4.2 and 4.3 illustrate two ways to shave off this factor of r. Theorem 4.2 applies to the easier ExistsEqual problem, and Theorem 4.3 applies to the general EqualityTesting problem, but blows up the round complexity to $\log^*(k/E) + O(r)$.

THEOREM 4.1. *There exists a $(\log^*(k/E) + r)$-round randomized protocol for EqualityTesting on vectors of length k that errs with probability $p_{\text{err}} = 2^{-E}$, using $O(k + rEk^{1/r})$ bits of communication.*

THEOREM 4.2. *There exists a $(\log^*(k/E) + r)$-round randomized protocol for ExistsEqual on vectors of length k that errs with probability $p_{\text{err}} = 2^{-E}$, using $O(k + Ek^{1/r})$ bits of communication.*

THEOREM 4.3. *There exists a $(\log^*(k/E) + O(r))$-round randomized protocol for EqualityTesting on vectors of length k that errs with probability $p_{\text{err}} = 2^{-E}$, using $O(k + Ek^{1/r})$ bits of communication.*

REMARK 4.1. *The $\log^*(k/E)$ terms in the round complexity of Theorems 4.1–4.3 are not absolute. They can each be replaced with $\max\{0, \log^*(k/E) - \log^*(C)\}$, at the cost of increasing the communication by $O(Ck)$.*

4.1 Generic Protocols

We start by giving a generic protocol for EqualityTesting. The protocol uses a simple subroutine for ExistsEqual/EqualityTesting when $k = 1$. Suppose Alice and Bob hold $x, y \in U = \{0,1\}^l$, respectively. Alice picks a random $w \in \{0,1\}^l$ from the shared random source and sends Bob $\check{x} = \langle x, w\rangle \bmod 2$, where $\langle \cdot, \cdot\rangle$ is the inner product operator. Bob computes $\check{y} = \langle y, w\rangle \bmod 2$ and declares "$x = y$" iff $\check{x} = \check{y}$. Clearly, Bob never errs if $x = y$; it is straightforward to show that the probability of error is exactly 1/2 when $x \neq y$. We call this protocol an *inner product test* and $\check{x}, \check{y}$ *test bits*. A *b-bit inner product test* on x and y refers to b independent inner product tests on x and y.

The entire protocol is divided into several phases. Before phase j, $j \geq 1$, Alice and Bob agree on a subset I_{j-1} of coordinates on which all previous inner product tests have passed. In other words, they have not yet witnessed that any of the coordinates in I_{j-1} are not equal. Each coordinate $i \in I_{j-1}$ represents either an actual equality ($x_i = y_i$), or a *false positive* ($x_i \neq y_i$). At the beginning of the protocol, $I_0 = [k]$. In phase j, we perform l_j independent inner product tests on each coordinate in I_{j-1} and let $I_j \subseteq I_{j-1}$ be the remaining coordinates that pass all their respective inner product tests. Notice that each coordinate in I_{j-1} corresponding to equality will always pass all the tests and enter I_j, while those corresponding to inequalities will only enter I_j with probability 2^{-l_j}. At the end of the protocol, we declare all coordinates in I_r *equal* and all other coordinates *not equal*.

This finishes the description of our generic protocol. Theorems 4.1–4.3 all use the framework of the generic protocol and mainly differ in the details, such as how Alice and Bob exchange their test bits, how they decide l_j, and when the protocol terminates.

4.1.1 A protocol for exchanging test bits

For EqualityTesting, it is possible that a constant fraction of the coordinates are actually equalities, which makes $|I_j| = \Theta(k)$ for every j. The naive implementation explicitly exchanges all $l_j|I_{j-1}|$ test bits and uses $\Omega(kE)$ bits of communication in total. All the test bits corresponding to equalities are "wasted" in a sense.

For our application, it is important that the communication volume that Alice and Bob use to exchange their test bits in phase j be proportional to the number of false positives in I_{j-1}, instead of the size of I_{j-1}. We will use a slightly improved version of a protocol of Feder et al. [FKNN95] for exchanging the test bits.

Imagine packing the test bits into vectors $\hat{x}, \hat{y} \in B^{|I_{j-1}|}$ where $B = \{0,1\}^{l_j}$. Lemma 4.1 shows that Alice can transmit $\hat{x}$ to Bob, at a cost that depends on an *a priori* upper bound on the Hamming distance $\text{dist}(\hat{x}, \hat{y})$, i.e., the number of the coordinates in I_{j-1} where they differ.

LEMMA 4.1. (CF. FEDER ET AL. [FKNN95].) *Suppose Alice and Bob hold length-K vectors $x, y \in B^K$, where $B = \{0,1\}^L$. Alice can send one $O(dL + d\log(K/d))$-bit message to Bob, who generates a string $x' \in B^K$ such that the following holds. If the Hamming distance $\text{dist}(x,y) \leq d$ then $x = x'$; if $\text{dist}(x,y) > d$ then there is no guarantee.*

Proof. Define $G = (V, E)$ to be the graph on $V = B^K$ such that $\{u, v\} \in E$ iff $\text{dist}(u,v) \leq 2d$. The maximum degree in G is clearly at most $\Delta = \binom{K}{2d} \cdot 2^{2Ld}$ since there are $\binom{K}{2d}$ ways to select the $2d$ indices and 2^{2Ld} ways to change the coordinates at those indices so that there are *at most* $2d$ different coordinates. Let $\phi : V \mapsto [\Delta + 1]$ be a proper $(\Delta + 1)$-coloring of G. Alice sends $\phi(x)$ to Bob, which requires $\log(\Delta + 1) = O(dL + d\log(K/d))$ bits. Every string in the radius-d ball around y (w.r.t. dist) is colored differently since they are all at distance at most $2d$, hence if $\text{dist}(x,y) \leq d$, Bob can reconstruct x without error. $\square$

COROLLARY 4.1. *Suppose at phase j, it is guaranteed that the number of false positives in I_{j-1} is at most k_{j-1}. Then phase j can be implemented with $O(k_{j-1}l_j + k_{j-1}\log(k/k_{j-1}))$ bits in 2 rounds.*

A naive implementation of the protocol requires $2r$ rounds if the generic protocol has r phases. In fact, the protocol can be compressed into exactly r rounds in the following way. At the beginning, both parties agree that $I_0 = [k]$. Alice generates her $l_1|I_0|$ test bits $\hat{x}^{(1)}$ for phase 1 and communicates them to Bob; Bob first generates his own test bits $\hat{y}^{(1)}$ for phase 1 and determines I_1, then generates $l_2|I_1|$ test bits $\hat{y}^{(2)}$ for phase 2 and transmits both $\hat{y}^{(1)}$ and $\hat{y}^{(2)}$ to Alice. Alice computes I_1, generates $\hat{x}^{(2)}$, computes I_2, generates $\hat{x}^{(3)}$, and sends $\hat{x}^{(2)}$ and $\hat{x}^{(3)}$ to Bob, and so on. There is no asymptotic increase in the communication volume.

4.1.2 Reducing the number of false positives

Our protocols for EqualityTesting and ExistsEqual are divided into two parts. The goal of the first part is to reduce the number of false positives from at most k to at most E; if $E \geq k$, we can skip this part. Since the number of false positives is large in this part, we can use standard Chernoff bounds to control the number of false positives surviving each phase. The details are very similar to the upper bound in Sağlam and Tardos [ST13].

THEOREM 4.4. *Let (x, y) be an instance of ExistsEqual with $|x| = |y| = k$. In $\log^*(k/E)$ rounds, we can reduce this to a new instance (x', y') of ExistsEqual where $|x'| = |y'| \leq E$, using $O(k)$ communication. The failure probability of this protocol is at most $2^{-(E+1)}$.*

For EqualityTesting, we can reduce the initial instance to a new instance (x', y') such that the Hamming distance $\text{dist}(x', y') \leq E$, with the same round complexity, communication volume, and error probability.

Proof. We first give the protocol for ExistsEqual, then apply the necessary changes to make it work for EqualityTesting.

The protocol for ExistsEqual uses our generic protocol, and imposes a strict upper bound k_j on $|I_j|$. Whenever $|I_j|$ exceeds this upper bound, we halt the entire protocol and answer *yes*. We set the parameters k_j and l_j for any $j \in [1, \log^*(k/E)]$ as follows.

$$
\begin{aligned}
k_0 &= k, \\
k_j &= \max\left\{\frac{k}{2^{j-1}\exp^{(j)}(2)}, E\right\}, \\
l_j &= 3 + \exp^{(j-1)}(2).
\end{aligned}
$$

Now suppose the input vectors share no equal coordinates. We know that $|I_{j-1}| \leq k_{j-1}$ at the beginning of phase j. The probability of any particular coordinate in I_{j-1} passing all tests in phase j is exactly $p_j = \exp(-l_j)$. Thus, the expected size of I_j is at most

$$k_{j-1}p_j = \frac{k}{2^{j-2}\exp^{(j-1)}(2)} \cdot \frac{1}{2^3\exp^{(j)}(2)} \leq \frac{k_j}{8}.$$

Let X_i be the indicator variable that the ith coordinate in I_{j-1} survives to I_j and let $X = \sum_i X_i$. By the following Chernoff bound:

$$\Pr[X \ge (1+\delta)\mu] \le \left(\frac{e^\delta}{(1+\delta)^{1+\delta}}\right)^\mu,$$

we have:

$$\Pr[X \ge k_j] < 0.3^{k_j} < 2^{-1.7k_j}.$$

Hence, the probability that there are at least k_j coordinates remaining after phase j is at most $2^{-1.7k_j}$, and the probability this happens in any phase is at most $\sum_j 2^{-1.7k_j} \le 2^{-(E+1)}$. Notice that when x and y share at least one equal coordinate, the error probability of this protocol is 0 because if it fails to reduce the number of coordinates to E it (correctly) answers *yes*. The communication volume of the protocol is asymptotic to

$$\sum_j l_j|I_{j-1}| \le \sum_j l_j k_{j-1} = \sum_j O(k/2^j) = O(k).$$

For EqualityTesting, we use the same k_j as an upper bound on the number of false positives in I_j, instead of the size of I_j. Since the number of false positives is at most k at the beginning, we can still use the same argument to show that with the same choice of k_j and l_j, after $\log^*(k/E)$ phases, the number of false positives is at most E with error probability $2^{-(E+1)}$. By Corollary 4.1, the number of bits we need to exchange in phase j is $O(k_{j-1}l_j + k_{j-1}\log(k/k_{j-1}))$. Notice that $\log(k/k_{j-1}) = j - 2 + \exp^{(j-2)}(2) = O(l_j)$, so the total communication volume is still $O(k)$. ◻

In all of our protocols, we first apply Theorem 4.4 to reduce the number of coordinates (in the case of ExistsEqual) or false positives (in the case of EqualityTesting) to be at most E. This requires no communication if $E \ge k$ to begin with. Hence, with $\log^*(k/E)$ extra rounds and $O(k)$ communication, we will assume henceforth that all instances of ExistsEqual have $E \ge k$ and instances of EqualityTesting have $\mathrm{dist}(x,y) \le E$.

4.2 An $O(k+rEk^{1/r})$-bit EqualityTesting Protocol

In light of Theorem 4.4, we can assume that the input vectors to EqualityTesting are guaranteed to differ in at most $k_0 = \min\{k, E\}$ coordinates.

THEOREM 4.5. *Fix any $k \ge 1$, $E \ge 1$, and $r \in [1, (\log k_0)/2]$, where $k_0 = \min\{k, E\}$. There exists a randomized protocol for EqualityTesting length-k vectors x, y with Hamming distance $\mathrm{dist}(x,y) \le k_0$ that uses r rounds, $O(k + rEk_0^{1/r})$ bits of communication, and errs with probability $p_{\mathrm{err}} = 2^{-(E+1)}$.*

Proof. (Sketch) The proof for Theorem 4.5 is almost identical to the one for Theorem 4.4 except we use the following parameters k_j and l_j.

$$k_j = k_0^{1-j/r},$$
$$l_j = 4Ek_0^{j/r-1}.$$

◻

Combining Theorem 4.4 and Theorem 4.5, we obtain a $(\log^*(k/E)+r)$-round randomized protocol for EqualityTesting.

4.3 An $O(k + Ek^{1/r})$-bit ExistsEqual Protocol

4.3.1 Overview of the protocol In this section, we show that we can obtain a $(\log^*(k/E) + r)$-round, $O(k+Ek^{1/r})$-bit protocol for ExistsEqual. This matches the lower bound of Theorem 3.2, asymptotically, when $E \ge k$. Theorem 4.4 covers the first part of the protocol, so we assume without loss of generality that $E \ge k$.

Suppose the inputs x and y share no equal coordinates. Imagine writing down all the possible results of the inner product tests in a matrix A of dimension $(E + \log k) \times k$, where $A_{j,i}$ is "=" if x_i, y_i pass the jth inner product test, and "$\neq$" otherwise. By a union bound, with probability $1-2^{-E}$, each column contains at least one "$\neq$". Now consider the area above the first "$\neq$" in each column. The probability that this area is at least E' is, by a union bound, at most

$$(4.4)\quad \binom{E'+k-1}{k-1}2^{-E'} < \exp(k\log(e(E'+k)/k) - E').$$

For $E' = E + O(k\log(E/k)) = O(E)$, this probability is $\ll 2^{-E}$. In our analysis it suffices to consider a situation where an *adversary* can decide the contents of A, subject to the constraint that its *error budget* (the area above the curve defined by the first "$\neq$" in each column) never exceeds $E' = O(E)$. The notion of an error budget is also essential for analyzing the protocol of Section 4.4.

In the jth phase, $j \ge 1$, our protocol exposes the fragment of A consisting of the next l_j rows of columns in I_{j-1}. The set I_j consists of those columns without any "$\neq$" exposed so far. The *communication budget* for phase j is equal to $l_j|I_{j-1}|$. In the worst case, the first exposed value in each column of $I_{j-1} \setminus I_j$ is "$\neq$", so the adversary spends at least $l_j|I_j|$ of its *error budget* in phase j.

If we witness at least one "$\neq$" in every column, we can correctly declare there does not exist an equal coordinate and answer *no*. Otherwise, if the adversary has not exceeded his error budget but there is some column without any "$\neq$", we answer *yes*. If the adversary ever exhausts his error budget, we terminate the protocol and answer *yes*. Recall that the notion of an error budget tacitly assumed that x and y differ in all coordinates. If they do not, the protocol always answers correctly, whether it halts prematurely or not. The probability that the error budget is exhausted when x and y differ in all coordinates (a false positive) is $\ll 2^{-E}$, according to Eqn. (4.4).

4.3.2 Analysis In this section we give a formal proof to the following Theorem:

THEOREM 4.6. *Fix any $k \ge 1$, $E \ge k$, and $r \in [1, (\log k)/2]$. There exists an r-round randomized protocol for ExistsEqual on vectors of length k that errs with probability $p_{\mathrm{err}} = 2^{-(E+1)}$, using $O(Ek^{1/r})$ bits of communication.*

Proof. The number of tests per coordinate in phase j is l_j:

$$l_j = 2Ek^{j/r-1}.$$

Define $E_j = \sum_{j'=1}^{j} l_{j'}|I_{j'}|$ to be the portion of the error budget spent in phases 1 through j. We can express the asymptotic communication cost of the protocol in terms of the error budget as follows.

$$\begin{aligned}\sum_{j=1}^{r} l_j|I_{j-1}| &\le l_1|I_0| + k^{1/r}\sum_{j=2}^{r} l_{j-1}|I_{j-1}| && l_j = k^{1/r}l_{j-1}.\\ &\le 2Ek^{1/r} + E_{r-1}k^{1/r} && \text{Defn. of } E_{r-1}.\end{aligned}$$

Recall that the protocol terminates immediately after phase j if $E_j \ge E'$, which indicates $E_{r-1} < E'$. Hence, the total cost is bounded by

$$\le (2E + E')k^{1/r} \;=\; O(Ek^{1/r}).$$

The protocol can only err if x and y differ in every coordinate. In this case, there are two possible sources of error. The first possibility is that the protocol answers *yes* because $|I_r| \ge 1$. By a union bound, this happens with probability at most

$$k2^{-\sum_{j=1}^{r} l_j} \le k2^{-2E}.$$

The second possibility is that the protocol terminates prematurely and answers *yes* if $E_j \ge E'$ for some $j \in [1, r]$. The probability of this event occuring is also $\ll 2^{-E}$; see Eqn. (4.4). This concludes the proof. ◻

Proof. [Proof of Theorem 4.2] Theorem 4.2 follows directly by combining Theorem 4.4 and Theorem 4.6. ◻

REMARK 4.2. *By applying the reduction of Theorem 1.1 to Theorem 4.6, we conclude that* SetDisjointness *can be solved in $r + 1$ rounds using $O(Ek^{1/r})$ bits of communication. In this particular case we actually do not need Theorem 1.1; it is possible to solve* SetDisjointness *directly in r rounds with $O(Ek^{1/r})$ communication by an algorithm along the lines of Theorem 4.6 or [ST13]. Theorem 1.1 can also be applied to Theorem 4.5 to yield a* SetIntersection *protocol using $r + 1$ rounds and $O(rEk^{1/r})$ communication, but here we do not see how to solve the problem directly in r rounds. It seems we would need some analogue of Lemma 4.1 tailored to the* SetIntersection *problem.*

4.4 A Communication Optimal EqualityTesting Protocol

Finally, we give an EqualityTesting protocol that achieves the optimal communication complexity $O(Ek^{1/r})$ and uses $O(r)$ rounds (instead of r). Due to lack of space, we defer details of the protocol to the full version of the paper [HPZZ19].

THEOREM 4.7. *Fix any $k \ge 1$, $E \ge 1$, and $r \in [1, (\log k_0)/6]$, where $k_0 = \min\{k, E\}$. There exists a randomized protocol for* EqualityTesting *length-k vectors x, y with Hamming distance $\mathrm{dist}(x, y) \le k_0$ that uses $O(r)$ rounds, $O(k + Ek_0^{1/r})$ bits of communication, and errs with probability $p_{\mathrm{err}} = 2^{-(E+1)}$.*

Theorem 4.3 can then be obtained by combining Theorem 4.4 and Theorem 4.7.

5 Application in Distributed Triangle Enumeration

One way to solve local triangle enumeration in the CONGEST model is to execute, in parallel, a SetIntersection protocol across every edge of the graph, where the set associated with a vertex is a list of its neighbors. Since there are at most $\Delta n/2$ edges, we need the SetIntersection error probability to be 2^{-E}, $E = \Theta(\log n)$, in order to guarantee a global success probability of $1 - 1/\mathrm{poly}(n)$. Our lower bound says any algorithm taking this approach must take $\Omega((\Delta + E\Delta^{1/r})/\log n + r)$ rounds since each round of CONGEST allows for one $O(\log n)$-bit message. The hardest situation seems to be when $\Delta = E = \Theta(\log n)$, in which case the optimum choice is to set $r = \log \Delta$, making the triangle enumeration algorithm run in $O(\log \Delta) = O(\log \log n)$ time. In Theorem 5.1 we show that it is possible to handle this situation exponentially faster, in $O(\log \log \Delta) = O(\log \log \log n)$ time, and in general, to solve local triangle enumeration [IG17] in optimal $O(\Delta/\log n)$ time so long as $\Delta > \log n \log \log \log n$.

THEOREM 5.1. *Local triangle enumeration can be solved in a CONGEST network $G = (V, E)$ with maximum degree Δ in $O(\Delta/\log n + \log \log \Delta)$ rounds with probability $1 - 1/\mathrm{poly}(n)$. This is optimal for all $\Delta = \Omega(\log n \log \log \log n)$.*

Proof. The algorithm consists of $\min\{\log \log \Delta, \log \log \log n\}$ phases. The goal of the first phase is to transform the original triangle enumeration problem into one with maximum degree $\Delta_1 < (\log n)^{o(1)}$, in $O(\log^* n)$ rounds of communication. The goal of every subsequent phase is to reduce the maximum degree from $\Delta' \le \sqrt{\log n}$ to $\sqrt{\Delta'}$, in $O(1)$ rounds of communication. Thus, the total number of rounds is $O(\log \log \Delta)$ rounds if the first round is skipped, and $O(\log^* n + \log \log(\Delta_1)) = O(\log \log \log n)$ otherwise.

Phase One. Suppose $\Delta \ge \sqrt{\log n}$. Each vertex u is identified with the set $A_u = \{\mathsf{ID}(v) \mid \{v, u\} \in E\}$ having size Δ. For each $\{u, v\} \in E$ we reduce SetIntersection to EqualityTesting by applying Theorem 1.1, then run the two-party EqualityTesting protocol of Theorem 4.1, with $k = \max\{\Delta, \log n\}, r = \log^* n$, and $E = r^{-1}k^{1-1/r}$. (I.e., if $\Delta < \log n$ we imagine padding each set to size $\log n$ with dummy elements.) One undesirable property of this protocol is that it can fail "silently" if the preconditions of Lemma 4.1 are not met. When the Hamming distance between two strings exceeds the threshold d, Bob generates a garbage string $x' \ne x$ but fails to detect this. To rectify this problem, we change the Lemma 4.1 protocol slightly: Alice sends the color $\phi(x)$ of her string, as well as an $O(\log n)$-bit hash $h(x)$. Bob reconstructs x' as usual and terminates the protocol if $h(x) \ne h(x')$. Clearly the probability of an undetected failure (i.e., $x \ne x'$ but $h(x) = h(x')$) is $1/\mathrm{poly}(n)$. Define $G_1 = (V, E_1)$ such that $\{u, v\} \in E_1$ iff the SetIntersection protocol over $\{u, v\}$ detected a failure. In other words, with high probability, all triangles in G have been discovered, except for those contained entirely inside G_1. The probability that any particular edge appears in E_1 is $2^{-E} = 2^{-k^{1-1/\log^* n}/\log^* n}$ and independent of all other edges. In particular, if $\Delta \gg (\log n)^{1+1/\log^* n}$ then no errors occur, with probability $1 - 1/\text{poly}(n)$. Define Δ_1 to be the maximum degree in G_1. Thus,

$$\begin{aligned}
&\Pr\left[\Delta_1 \geq (\log n)^{2\epsilon}\right] \qquad\qquad \epsilon = 1/r = 1/\log^* n.\\
&\leq n \cdot \binom{\Delta}{(\log n)^{2\epsilon}} \cdot \left(2^{-E}\right)^{(\log n)^{2\epsilon}}\\
&\leq n \cdot \exp(O((\log n)^{2\epsilon} \log\log n)) \cdot 2^{-\epsilon(\log n)^{1-\epsilon} \cdot (\log n)^{2\epsilon}}\\
&\leq 1/\text{poly}(n).
\end{aligned}$$

Phases Two and Above. Suppose that at some round, we have detected all triangles except for those contained in some subgraph $G' = (V, E')$ having maximum degree $\Delta' < \sqrt{\log n}$. Express Δ' as $(\log n)^\gamma$, where $\gamma < 1/2$. We execute the EqualityTesting protocol of Theorem 4.5 with $k = \Delta'$, $r = 2$, and $E = C(\log n)^{1-\gamma/2}$ for a sufficiently large constant C. Note that $1-\gamma/2 > \gamma$, so $E > k$, as required by Theorem 4.5. The protocol takes $O(Ek^{1/2}/\log n + r) = O(1)$ rounds since the communication volume is $O(Ek^{1/2}) = O(\log n)$ and $r = 2$. Let G'' be the subgraph of G' consisting of edges whose protocols detected a failure and Δ'' be the maximum degree in G''. Once again,

$$\begin{aligned}
&\Pr\left[\Delta'' \geq (\log n)^{\gamma/2}\right]\\
&\leq n \cdot \binom{\Delta'}{(\log n)^{\gamma/2}} \cdot \left(2^{-E}\right)^{(\log n)^{\gamma/2}}\\
&\leq n \cdot \exp(O((\log n)^{\gamma/2} \log\log n)) \cdot 2^{-C(\log n)^{1-\gamma/2} \cdot (\log n)^{\gamma/2}}\\
&\leq 1/\text{poly}(n).
\end{aligned}$$

Thus, once $\Delta \leq \sqrt{\log n}$, $\log\log\Delta \leq \log\log\log n - 1$ of these 2-round phases suffice to find all remaining triangles in G. □

Theorem 5.1 depends critically on the duality between edges and SetIntersection instances, and between edge endpoints and elements of sets. In particular, when an execution of a SetIntersection over $\{u, v\}$ is successful, this effectively *removes* $\{u, v\}$ from the graph, thereby removing many occurrences of $\mathsf{ID}(u)$ and $\mathsf{ID}(v)$ from adjacent sets.

Consider a slightly more general situation where we have a graph of *arboricity* λ (but unbounded Δ), witnessed by a given acyclic orientation having out-degree at most λ. Redefine the set A_u to be the set of out-neighbors of u.

$$A_u = \{\mathsf{ID}(v) \mid \{u, v\} \in E \text{ with orientation } u \to v\}.$$

By definition $|A_u| \leq \lambda$. Because the orientation is acyclic, every triangle on $\{x, y, z\}$ is (up to renaming) oriented as $x \to y,\ x \to z,\ y \to z$. Thus, it will *only* be detectable by the SetIntersection instance associated with $\{x, y\}$.

Theorem 5.2. *Let $G = (V, E)$ be a CONGEST network equipped with an acyclic orientation with outdegree at most λ. We can solve local triangle enumeration on G in $O(\lambda/\log n + \log\lambda)$ time.*

Proof. We apply Theorem 1.1 to reduce each SetIntersection instance to an EqualityTesting instance, then apply Theorem 4.3 with $E = \Theta(\log n)$ and $r = \log\lambda$ to solve each with $O(\lambda + E\lambda^{1/r}) = O(\lambda + E)$ communication in $O((\lambda + E)/\log n + r) = O(\lambda/\log n + \log\lambda)$ time. Note that the dependence on λ here is exponentially worse than the dependence on Δ in Theorem 5.1. □

It may be that G is known to have arboricity λ, but an acyclic orientation is unavailable. The well known "peeling algorithm" (see [CN85] or [BE10]) computes a $C\lambda$ orientation in $O(\log_C n)$ time for C sufficiently large, say $C \geq 3$. Using this algorithm as a preprocessing step, we can solve local triangle enumeration optimally when $\lambda = \Omega(\log^2 n)$.

Theorem 5.3. *Let $G = (V, E)$ be a CONGEST network having arboricity λ (with no upper bound on Δ). Local triangle enumeration can be solved in optimal $O(\lambda/\log n)$ time when $\lambda = \Omega(\log^2 n)$, and sublogarithmic time $O(\log n/\log(\log^2 n/\lambda))$ otherwise.*

Proof. The algorithm computes a $\gamma \cdot \lambda$ orientation in $O(\log_\gamma n)$ time and then applies Theorem 5.2 to solve local triangle enumeration in $O(\gamma\lambda/\log n + \log(\gamma\lambda))$ time. The only question is how to set γ. If $\lambda = \Omega(\log^2 n)$ we set $\gamma = 3$, making the total time $O(\lambda/\log n)$, which is optimal [IG17]. Otherwise we choose γ to balance the $\log_\gamma n$ and $\gamma\lambda/\log n$ terms, so that

$$\gamma\log\gamma = \log^2 n/\lambda$$

Thus, the total running time is slightly sublogarithmic $O(\log n/\log(\log^2 n/\lambda))$. Specifically, it is $O(\log n/\log\log n)$ whenever $\lambda < \log^{2-\epsilon} n$. □

6 Conclusions and Open Problems

We have established a new three-way tradeoff between rounds, communication, and error probability for many fundamental problems in communication complexity such as SetDisjointness and EqualityTesting. Our lower bound is largely incomparable to the round-communication lower bounds of [ST13, BCK+16], and stylistically very different from both [ST13] and [BCK+16]. We believe that our method *can* be extended to recover Sağlam and Tardos's [ST13] tradeoff (in the constant error probability regime), but with a more "direct" proof that avoids some technical difficulties arising from their round-elimination technique. It is still open whether EqualityTesting can be solved in r rounds with precisely $O(Ek^{1/r})$ communication and error probability $2^{-E} < 2^{-k}$. Our algorithms match this lower bound only when $r = O(1)$ or $r = \Omega(\log k)$, or for any r when solving the easier ExistsEqual problem.

We developed some CONGEST algorithms for triangle enumeration that employ two-party SetIntersection protocols. It is known that this strategy is suboptimal when $\Delta \gg n^{1/3}$ [CPZ19, CS19]. However, for the *local* triangle enumeration problem[7], our $O(\Delta/\log n + \log\log\Delta)$ algo-

[7]Every triangle must be reported by one of its three constituent vertices.

rithm is optimal [IG17] for every $\Delta = \Omega(\log n \log\log\log n)$. Whether there are faster algorithms for triangle *detection*[8] is an intriguing open problem. It is known that 1-round LOCAL algorithms must send messages of $\Omega(\Delta \log n)$ bits deterministically [ACKL17] or $\Omega(\Delta)$ bits randomized [FGKO18]. Even for 2-round triangle detection algorithms, there are no nontrivial communication lower bounds known.

A Reductions and Near Equivalences

Brody et al. [BCK+16] proved that SetIntersection on sets of size k is reducible to EqualityTesting on vectors of length $O(k)$, at the cost of one round and $O(k)$ bits of communication. However, the reduction is *randomized* and fails with probability at least $\exp(-\tilde{O}(\sqrt{k}))$. This is the probability that when k balls are thrown uniformly at random into k bins, some bin contains $\omega(\sqrt{k})$ balls.

Recall the statement of Theorem 1.1:

$$\begin{aligned}
\mathsf{Eq}(k, r, p_{\mathrm{err}}) &\le \mathsf{SetInt}(k, r, p_{\mathrm{err}}),\\
\exists\mathsf{Eq}(k, r, p_{\mathrm{err}}) &\le \mathsf{SetDisj}(k, r, p_{\mathrm{err}}),\\
\mathsf{SetInt}(k, r+1, p_{\mathrm{err}}) &\le \mathsf{Eq}(k, r, p_{\mathrm{err}}) + \zeta,\\
\mathsf{SetDisj}(k, r+1, p_{\mathrm{err}}) &\le \exists\mathsf{Eq}(k, r, p_{\mathrm{err}}) + \zeta,
\end{aligned}$$

where $\zeta = O(k + \log\log p_{\mathrm{err}}^{-1})$. In other words, under any error regime p_{err}, the communication complexity of SetIntersection and EqualityTesting are the same, up to one round and $O(k + \log\log p_{\mathrm{err}}^{-1})$ bits of communication, and that the same relationship holds between SetDisjointness and ExistsEqual. The proof is inspired by the probabilistic reduction of Brody et al. [BCK+16], but uses succinct encodings of perfect hash functions rather than random hash functions.

Proof. [Proof of Theorem 1.1] The leftmost inequalities have been observed before [ST13, BCK+16]. Given inputs x, y to ExistsEqual or EqualityTesting, Alice and Bob generate sets $A = \{(1, x_1), \ldots, (k, x_k)\}$ and $B = \{(1, y_1), \ldots, (k, y_k)\}$ before the first round of communication and then proceed to solve SetIntersection or SetDisjointness on (A, B). Knowing $A \cap B$ or whether $A \cap B = \emptyset$ clearly allows them to determine the correct output of EqualityTesting or ExistsEqual on (x, y).

The reverse direction is slightly more complicated. Let (A, B) be the instance of SetIntersection or SetDisjointness over a universe U with size at most $|U| = O(k^2/p_{\mathrm{err}})$. Alice examines her set A, and picks a *perfect* hash function $h : U \mapsto [k]$ for A, i.e., h is injective on A. (This can be done in $O(k)$ time, in expectation, using only *private* randomness. In principle Alice could do this step deterministically, given sufficient time.) Most importantly, h can be described using $O(k + \log\log|U|) = O(k + \log\log p_{\mathrm{err}}^{-1})$ bits [SS90], using a variant of the Fredman-Komlós-Szemerédi [FKS84] 2-level perfect hashing scheme.[9] Alice sends the $O(k + \log\log p_{\mathrm{err}}^{-1})$-bit description of h to Bob. Bob calculates $B_j = B \cap h^{-1}(j)$ and responds to Alice with the distribution $|B_0|, |B_1|, \ldots, |B_{k-1}|$, which takes at most $2k$ bits. They can now generate an instance of Equality Testing where the k equality tests are the pairs $A_0 \times B_0, A_1 \times B_1, \ldots, A_{k-1} \times B_{k-1}$. By construction, $A_j = A \cap h^{-1}(j)$ is a 1-element set. There is clearly a 1-1 correspondence between equal pairs and elements in $A \cap B$. We have Bob speak first in the EqualityTesting/ExistsEqual protocol; thus, the overhead for this reduction is just 1 round of communication and $O(k + \log\log p_{\mathrm{err}}^{-1})$ bits. □

[8] At least one vertex must announce there is a triangle; there is no obligation to list them all.

[9] We sketch how the encoding of h works, for completeness. First, pick a function $h' : U \mapsto [O(k^2)]$ that is collision-free on A. Fredman et al. [FKS84] proved that a function of the form $h'(x) = (ax \mod p) \mod O(k^2)$ works with constant probability, where $p = \Omega(k^2 \log|U|)$ is prime and $a \in [0, p)$ is random. Pick another function $h_* : [O(k^2)] \mapsto [k]$ that has at most twice the expected number of collisions on A, namely $2 \cdot \binom{k}{2}/k < k$, and partition A into k buckets $A_j = A \cap h_*^{-1}(j)$. The sizes $|A_0|, |A_1|, \ldots, |A_{k-1}|$ can be encoded with $2k$ bits. We now pick $O(\log k)$ pairwise independent hash functions $h_1, h_2, \ldots, h_{O(\log k)} : [O(k^2)] \mapsto [O(k^2)]$. For each bucket A_j, we define $h_{(j)}$ to be the function with the minimum i for which $h_{(j)}(x) = h_i(x) \mod |A_j|^2$ is injective on A_j. In order to encode which function $h_{(j)}$ is (given that $h_1, \ldots, h_{O(\log k)}$ are fixed and that $|A_j|$ is known), we simply need to write i in unary, i.e., using the bit-string $0^{i-1}1$. This takes less than 2 bits per j in expectation since each h_i is collision-free on A_j with probability at least 1/2. Combining $h', h_*, |A_0|, \ldots, |A_{k-1}|$ and $h_{(0)}, \ldots, h_{(k-1)}$ into a single injective function from $U \mapsto [O(k)]$ is straightforward, and done exactly as in [FKS84]. By marking which elements in this range are actually used ($O(k)$ more bits), we can generate the perfect $h : U \mapsto [k]$ whose range has size precisely k. Encoding h' takes $O(\log k + \log\log|U|)$ bits and encoding h_* takes $O(\log k)$ bits. The distribution $|A_0|, \ldots, |A_{k-1}|$ can be encoded with $2k$ bits. The functions $h_1, \ldots, h_{O(\log k)}$ can be encoded in $O(\log^2 k)$ bits, and the functions $h_{(0)}, \ldots, h_{(k-1)}$ with less than $2k$ bits in expectation.

References

[ACKL17] A. Abboud, K. Censor-Hillel, S. Khoury, and C. Lenzen. Fooling views: A new lower bound technique for distributed computations under congestion. *CoRR*, abs/1711.01623, 2017.

[BCK+16] J. Brody, A. Chakrabarti, R. Kondapally, D. P. Woodruff, and G. Yaroslavtsev. Certifying equality with limited interaction. *Algorithmica*, 76(3):796–845, 2016.

[BE10] L. Barenboim and M. Elkin. Sublogarithmic distributed MIS algorithm for sparse graphs using Nash-Williams decomposition. *Distributed Computing*, 22(5-6):363–379, 2010.

[BFS86] L. Babai, P. Frankl, and J. Simon. Complexity classes in communication complexity theory. In *Proceedings of the 27th Annual IEEE Symposium on Foundations of Computer Science (FOCS)*, pages 337–347, 1986.

[BGMdW13] H. Buhrman, D. García-Soriano, A. Matsliah, and R. de Wolf. The non-adaptive query complexity of testing k-parities. *Chicago J. Theor. Comput. Sci.*, 2013, 2013.

[CK18] A. Czumaj and C. Konrad. Detecting cliques in CONGEST networks. In *Proceedings of the 32nd International Symposium on Distributed Computing (DISC)*, volume 121 of *Leibniz International Proceedings in Informatics (LIPIcs)*, pages 16:1–16:15, 2018.

[CN85] N. Chiba and T. Nishizeki. Arboricity and subgraph listing algorithms. *SIAM Journal on Computing*, 14(1):210–223, 1985.

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

[CPZ19] Y.-J. Chang, S. Pettie, and H. Zhang. Distributed triangle detection via expander decomposition. In *Proceedings of the 30th Annual ACM-SIAM Symposium on Discrete Algorithms (SODA)*, pages 821–840, 2019.

[CS19] Y.-J. Chang and T. Saranurak. Improved distributed expander decomposition and nearly optimal triangle enumeration. In *Proceedings of the 2019 ACM Symposium on Principles of Distributed Computing (PODC)*, pages 66–73, 2019.

[DKO14] A. Drucker, F. Kuhn, and R. Oshman. On the power of the congested clique model. In *Proceedings of the 33rd ACM Symposium on Principles of Distributed Computing (PODC)*, pages 367–376, 2014.

[DKS12] A. Dasgupta, R. Kumar, and D. Sivakumar. Sparse and lopsided set disjointness via information theory. In *Proceedings of the 15th International Workshop on Approximation, Randomization, and Combinatorial Optimization (APPROX)*, pages 517–528, 2012.

[DP09] D. P. Dubhashi and A. Panconesi. *Concentration of Measure for the Analysis of Randomized Algorithms.* Cambridge University Press, 2009.

[FGKO18] O. Fischer, T. Gonen, F. Kuhn, and R. Oshman. Possibilities and impossibilities for distributed subgraph detection. In *Proceedings of the 30th Symposium on Parallelism in Algorithms and Architectures (SPAA)*, pages 153–162, 2018.

[FKNN95] T. Feder, E. Kushilevitz, M. Naor, and N. Nisan. Amortized communication complexity. *SIAM J. Comput.*, 24(4):736–750, 1995.

[FKS84] M. L. Fredman, J. Komlós, and E. Szemerédi. Storing a sparse table with $O(1)$ worst case access time. *J. ACM*, 31(3):538–544, 1984.

[GO18] T. Gonen and R. Oshman. Lower bounds for subgraph detection in the CONGEST model. In *Proceedings of the 21st International Conference on Principles of Distributed Systems (OPODIS)*, volume 95 of *Leibniz International Proceedings in Informatics (LIPIcs)*, pages 6:1–6:16, 2018.

[HW07] J. Håstad and A. Wigderson. The randomized communication complexity of set disjointness. *Theory of Computing*, 3(1):211–219, 2007.

[HPZZ19] D. Huang, S. Pettie, Y. Zhang, and Z. Zhang. The communication complexity of set intersection and multiple equality testing. *ArXiv*, abs/1908.11825, 2019.

[IG17] T. Izumi and F. Le Gall. Triangle finding and listing in CONGEST networks. In *Proceedings of the 36th ACM Symposium on Principles of Distributed Computing (PODC)*, pages 381–389, 2017.

[KN97] E. Kushilevitz and N. Nisan. *Communication Complexity.* Cambridge University Press, 1997.

[KR18] J. H. Korhonen and J. Rybicki. Deterministic subgraph detection in broadcast CONGEST. In *Proceedings of the 21st International Conference on Principles of Distributed Systems (OPODIS)*, volume 95 of *Leibniz International Proceedings in Informatics (LIPIcs)*, pages 4:1–4:16, 2018.

[KS92] B. Kalyanasundaram and G. Schnitger. The probabilistic communication complexity of set intersection. *SIAM J. Discrete Math.*, 5(4):545–557, 1992.

[Lov89] L. Lovasz. Communication complexity: A survey. Technical Report TR-204-89, Computer Science Dept., Princeton University, 1989.

[Nik13] V. Nikishkin. Amortized communication complexity of an equality predicate. In *Proceedings 8th International Computer Science Symposium in Russia (CSR)*, pages 212–223, 2013.

[Raz92] A. A. Razborov. On the distributional complexity of disjointness. *Theor. Comput. Sci.*, 106(2):385–390, 1992.

[Rou16] T. Roughgarden. Communication complexity (for algorithm designers). *Foundations and Trends in Theoretical Computer Science*, 11(3-4):217–404, 2016.

[RY] A. Rao and A. Yehudayoff. Communication complexity. (unpublished manuscript; available from the authors' homepages).

[SS90] J. P. Schmidt and A. Siegel. The spatial complexity of oblivious k-probe hash functions. *SIAM J. Comput.*, 19(5):775–786, 1990.

[ST13] M. Sağlam and G. Tardos. On the communication complexity of sparse set disjointness and exists-equal problems. In *Proceedings of the 54th Annual IEEE Symposium on Foundations of Computer Science (FOCS)*, pages 678–687, 2013.

[Yao77] A. C.-C. Yao. Probabilistic computations: Toward a unified measure of complexity (extended abstract). In *Proceedings of the 18th Annual IEEE Symposium on Foundations of Computer Science (FOCS)*, pages 222–227, 1977.

[Yao79] A. C.-C. Yao. Some complexity questions related to distributive computing. In *Proceedings of the 11th Annual ACM Symposium on Theory of Computing (STOC)*, pages 209–213, 1979.