Certifying Equality With Limited Interaction*

Joshua Brody[1], Amit Chakrabarti[2], Ranganath Kondapally[2], David P. Woodruff[3], and Grigory Yaroslavtsev[4]

1 **Swarthmore College, Swarthmore, PA, USA**
`joshua.e.brody@gmail.com`
2 **Dartmouth College, Dartmouth, NH, USA**
`ac@cs.dartmouth.edu, ranganathk@gmail.com`
3 **IBM Almaden, Almaden, CA, USA**
`dpwoodru@us.ibm.com`
4 **Brown University, ICERM, Providence, RI, USA**
`grigory@grigory.us`

Abstract

The EQUALITY problem is usually one's first encounter with communication complexity and is one of the most fundamental problems in the field. Although its deterministic and randomized communication complexity were settled decades ago, we find several new things to say about the problem by focusing on three subtle aspects. The first is to consider the *expected* communication cost (at a worst-case input) for a protocol that uses limited interaction—i. e., a bounded number of rounds of communication—and whose error probability is zero or close to it. The second is to treat the *false negative* error rate separately from the *false positive* error rate. The third is to consider the *information cost* of such protocols. We obtain asymptotically optimal rounds-versus-cost tradeoffs for EQUALITY: both expected communication cost and information cost scale as $\Theta(\log\log\cdots\log n)$, with $r-1$ logs, where r is the number of rounds. These bounds hold even when the false negative rate approaches 1. For the case of zero-error communication cost, we obtain essentially matching bounds, up to a tiny additive constant. We also provide some applications.

1998 ACM Subject Classification F.1.2 Modes of Computation

Keywords and phrases equality, communication complexity, information complexity

Digital Object Identifier 10.4230/LIPIcs.APPROX-RANDOM.2014.545

1 Introduction

1.1 Context

Over the last three decades, communication complexity [44] has proved itself to be among the most useful of abstractions in computer science. A number of basic problems in communication complexity have found a wide range of applications throughout the theory of computing, with EQUALITY, INDEX, and DISJOINTNESS being notable superstars.

Revisiting these basic problems and asking more nuanced questions or studying natural variants has extended their range of application. We highlight two examples. Our first example is DISJOINTNESS. The optimal $\Omega(n)$ lower bound for this problem [29, 41] was already considered one of the major results in communication complexity before DISJOINTNESS was revisited in the *multi-party* number-in-hand model to obtain a number of data stream lower

[*] Grigory Yaroslavtsev was supported by an Institute Postdoctoral Fellowship at Brown ICERM.

17th Int'l Workshop on Approximation Algorithms for Combinatorial Optimization Problems (APPROX'14) /
18th Int'l Workshop on Randomization and Computation (RANDOM'14).
Editors: Klaus Jansen, José Rolim, Nikhil Devanur, and Cristopher Moore; pp. 545–581
Leibniz International Proceedings in Informatics
LIPICS Schloss Dagstuhl – Leibniz-Zentrum für Informatik, Dagstuhl Publishing, Germany

bounds [3, 4, 13, 23] culminating in optimal space bounds for the (higher) frequency moments. Later, DISJOINTNESS was revisited in an *asymmetric* communication setting [40] yielding an impressive array of lower bounds for data structures in the cell-probe model. Very recently, DISJOINTNESS was revisited yet again in a *high-error* setting, yielding deep insights into extended formulations for the MAX-CLIQUE problem [8]. Our second example is INDEX. The straightforward $\Omega(n)$ lower bound on its one-way communication complexity [1] is already an important starting point for numerous other lower bounds. Revisiting INDEX in an *interactive* communication setting and considering communication tradeoffs has led to new classes of data stream lower bounds for memory-checking problems [34, 12, 14]. Separately, lower bounding the *quantum* communication complexity of INDEX [39] has led, among other things, to strong lower bounds for locally decodable codes [30, 16].

1.2 Our Results

In this work we revisit the EQUALITY problem: Alice and Bob each hold an n-bit string, and their task is to decide whether these strings are equal. This is arguably the most basic communication problem that admits a nontrivial protocol: using randomization and allowing a constant error rate, the problem can be solved with just $O(1)$ communication (this becomes $O(\log n)$ if one insists on private coins only); see, e. g., Kushilevitz and Nisan [32, Example 3.13] and Freivalds [22]. This is why a student's first encounter with communication complexity is usually through the EQUALITY problem. Such a fundamental problem deserves the most thorough of studies.

At first glance, EQUALITY might appear "solved": its deterministic communication complexity is at least n, whereas its randomized complexity is $O(1)$ as noted above, as is its *information complexity* [6] (for more on this, see Section 1.3). However, one can ask the following more nuanced question. What happens if Alice and Bob want to be *certain* (or nearly certain) that their inputs are indeed equal when the protocol directs them to say so? And what happens if Alice and Bob want to run a protocol with limited interaction, i. e., a bounded number of back-and-forth rounds of communication?

Formally, let $\text{EQ}_n : \{0,1\}^n \times \{0,1\}^n \to \{0,1\}$ denote the Boolean function that underlies this communication problem, defined by $\text{EQ}_n(x,y) = 1 \iff x = y$. Consider the zero-error case: the players must always correctly output $\text{EQ}_n(x,y)$ on every input (x,y). However, the players may use a randomized protocol and their goal is to minimize the *expected* number of bits they exchange. If their protocol is required to use only one round—this means that Alice sends a message to Bob, who then outputs the answer—then it is easy to see that Alice's message must uniquely identify her input to Bob. From this it is easy to show that on some input, x, Alice must send at least n bits to Bob, even in expectation.

Things improve a lot if one allows two rounds of communication—Alice sends a message to Bob, who replies to Alice, who then outputs the answer. Using standard techniques, Alice can send Bob a $\lceil \log n \rceil$-bit[1] *fingerprint* of x. When $x \neq y$, this fingerprint demonstrates with probability at least $1 - 1/n$ that $\text{EQ}_n(x,y) = 0$. If necessary, Bob responds to this failure by sending y to Alice, which costs only 1 bit in expectation. The net result is an expected communication cost of $O(\log n)$ on unequal inputs, and $O(n)$ on equal inputs. Generalizing this idea, we obtain an r-round protocol where the expected cost drops to $O(\text{ilog}^{r-1} n)$ on unequal inputs, where $\text{ilog}^j n := \log\log\cdots\log n$ (with j logs).

Our main high-level message in this work is that *the above tradeoff between the number of*

[1] Throughout this paper we use "log" to denote the logarithm to the base 2.

rounds and the communication cost is optimal, and that this remains the case even allowing for some false positives, even allowing for a false negative rate of $1 - o(1)$, and *even if we consider information cost.* We shall get precise about information cost measures in Section 2, but for now we remark that an information cost lower bound is stronger than a communication cost bound, even in our expected-cost model.

While our main focus is on EQUALITY, our rounds-versus-information tradeoff can be applied to three other problems: OR-EQUALITY, DISJOINTNESS, and PRIVATE-INTERSECTION. It is well known that information cost has clean direct-sum properties [15, 4, 5]. Together with our results for EQUALITY, this gives us lower bounds for the bounded-round randomized communication complexity of the OR-EQUALITY problem, whose underlying function is $\text{OREQ}_{n,k} : \{0,1\}^{nk} \times \{0,1\}^{nk} \to \{0,1\}$, defined by $\text{OREQ}_{n,k}(x_1, \ldots, x_k, y_1, \ldots, y_k) = \bigvee_{i=1}^{k} \text{EQ}_n(x_i, y_i)$: Alice holds each $x_i \in \{0,1\}^n$ and Bob holds each $y_i \in \{0,1\}^n$. The OREQ problem is closely related to DISJOINTNESS, especially the variant called *small set disjointness* or k-DISJ$_N$. Here, Alice and Bob are given sets $A, B \subseteq \{1, 2, \ldots, N\}$ respectively, with the promise that $|A| \leq k$ and $|B| \leq k$, where $1 \leq k \leq N$. Their goal is to output 1 iff $A \cap B = \varnothing$. Using this close relation (see Lemma 54 for a formal treatment), we obtain bounded-round lower bounds for k-DISJ as well. For low-error protocols, our bounds asymptotically match those recently given by Sağlam and Tardos [42]; our proof is quite different and should be of independent interest. Additionally, our lower bound also allows a false negative rate of $1 - o(1)$ — we show that if players only need to verify equal inputs with probability, say, 0.001, the problem remains difficult.

Another key property is that information cost is a measure of *privacy* of a protocol for a function f. Klauck [31] defines[2] the privacy of a protocol Π with respect to a distribution μ:

$$\text{PRIV}^{\mu}(\Pi) := \text{I}(X : \Pi(X,Y) \mid Y, f(X,Y)) + \text{I}(Y : \Pi(X,Y) \mid X, f(X,Y))\,.$$

This definition coincides with $\text{IC}_{\mu}(\Pi)$ up to the conditioning on $f(X,Y)$ in the mutual information expressions. However, in many cases, including this paper, this conditioning does not asymptotically affect the definition, and one has $\text{PRIV}^{\mu}(\Pi) = \Theta(\text{icost}^{\mu}(\Pi))$. One can then naturally define $\text{PRIV}_{\delta}(f) = \min_{\delta\text{-error } \Pi} \max_{\text{input dist } \mu} \text{PRIV}^{\mu}(\Pi)$, and one has that $\text{PRIV}_{\delta}(f) = \Theta(\text{IC}_{\delta}(f))$.

There is a large body of work on trying to solve EQUALITY privately. These are known as *private equality tests* in the cryptography and privacy literature [19, 38]. A harder problem is that of determining the intersection $A \cap B$ of sets A, B in some finite universe, where each of $|A|$ and $|B|$ is promised to be at most k. This is a fundamental problem studied in private datamining, see, e. g., the work by Freedman et al. [21]. We refer to the latter problem as the PRIVATE-INTERSECTION problem. It is worth noting that the PRIVATE-INTERSECTION problem is studied both under computational assumptions on the players, as in the work by Freedman et al. [21], and also using information-theoretic notions of privacy, such as $\text{PRIV}_{\delta}(f)$, as in the work by Ada et al. [2]. Note that for the PRIVATE-INTERSECTION problem, we are asking for a correct protocol which reveals as little information about A and B as possible, with no constraints on the communication.

While the information complexity of PRIVATE-INTERSECTION is known to be $\Theta(k)$, in certain applications the players can only exchange messages in a bounded number r of rounds, since, e. g., the number of rounds is related to the overall latency of the protocol. The number of rounds may in fact influence the latency drastically while the actual number

[2] We have replaced the max in Klauck's definition with a sum; this agrees with Klauck's original definition up to a factor of 2.

of bits communicated may not. This is because the more interactive protocols are, i. e., the larger the number of rounds, the more coordination is needed between the players, which may not be possible if, e. g., a player goes offline.

We apply our information complexity lower bound for EQUALITY to the PRIVATE-INTERSECTION problem in which each player should locally output the entire set intersection $S \cap T$. Our information complexity lower bound for EQUALITY can be combined with a recent direct sum theorem with *aborts*, which (roughly speaking) states that the information complexity of solving all k copies of a problem is k times the information cost of solving each copy with a protocol that is allowed to output "abort" with a constant $1/10$ probability but, conditioned on non-abortion, is correct with a very high $1 - 1/k$ probability [37].[3] By changing such a protocol for EQUALITY so that whenever it would have output "abort", it instead declares that $x \neq y$, we show how to obtain an $\Omega(k \operatorname{ilog}^r k)$ information cost bound for PRIVATE-INTERSECTION for any r-round protocol with constant success probability. As $\mathrm{I}(\Pi : S \mid T, S \cap T) + \mathrm{I}(\Pi : T \mid S, S \cap T) = \mathrm{I}(\Pi : S \mid T) + \mathrm{I}(\Pi : T \mid S) \pm O(k)$, it follows that $\mathrm{PRIV}_{1/3}(\text{PRIVATE-INTERSECTION}) = \Omega(k \operatorname{ilog}^r k)$.

For a concise (yet technically precise) listing of our results, please see Section 2.

1.3 Related Work

The study of the EQUALITY problem goes back to the original communication complexity paper of Yao [44], who showed that the deterministic communication complexity of EQ_n is at least n, using a fooling set argument. Mehlhorn and Schmidt [35] developed the *rank lower bound* technique, which can recover this result. They further examined OR-EQUALITY, giving a lower bound of nk bits for deterministic protocols that compute $\mathrm{OREQ}_{n,k}$ via the rank technique. They also gave $O(n + \log n)$ and $O(n \log n)$ bounds for the nondeterministic and co-nondeterministic communication complexities of $\mathrm{OREQ}_{n,n}$, respectively. Furthermore, they studied the "Las Vegas" communication complexity of $\mathrm{OREQ}_{n,n}$, which brought them close to some of the things we study here. Specifically, they gave a zero-error private-coin randomized protocol such that the expected communication cost on any inputs $(x_1, \ldots, x_n, y_1, \ldots, y_n)$ is at most $O(n(\log n)^2)$.

Feder et al. [20] studied the randomized communication complexity of EQUALITY in the direct-sum setting. Here, players have k strings each and must compute $(\mathrm{EQ}_n(x_1, y_1), \ldots, \mathrm{EQ}_n(x_k, y_k))$: thus, the output is a k-bit string. Feder et al. showed that $O(k)$ communication suffices to compute EQUALITY on all k instances, with error *exponentially* small in k. This shows that the "amortized" communication complexity of EQ_n is $O(1)$, even under tiny error. More recently, Braverman and Rao [9] showed that amortized communication complexity nearly equals *information* complexity. Furthermore, Braverman [6] gave a specific protocol for EQ_n that has zero error and achieves information cost $O(1)$ regardless of the input distribution.

[3] It is crucial for us to use a strong direct sum theorem of [37] in the lower bound for PRIVATE-INTERSECTION. Unlike generic direct sum and direct product theorems which apply to any function the strong direct sum of [37] only applies to EQUALITY-type functions but gives a much stronger guarantee in the constant error regime that we study here. This is in contrast with the bounded round direct product theorem of [26, 27] (and other similar results such as [28]), who show that for r-round public-coin randomized information complexity $\mathrm{IC}^{r,\mathrm{pub}}_{1-(1-\varepsilon/2)^{\Omega(k\varepsilon^2/r^2)}}(f^k) = \Omega\Big(\, (\varepsilon k/r) \cdot (IC^{r,\mathrm{pub}}_{\varepsilon}(f) - O(r^2/\varepsilon^2)) \,\Big)$, where $\varepsilon > 0$ is arbitrary (the results of [26, 27] are stated in terms of communication complexity but their techniques also imply an information cost lower bound). One cannot apply this theorem to our problem, as one would need to set $\varepsilon = \Theta(k^{-1/3})$ to obtain our results. Because $\mathrm{IC}^{r,\mathrm{pub}}_{1/k^{1/3}}(\text{EQUALITY}) = o(k^{2/3})$ this theorem gives a trivial bound.

The problem $\textsc{oreq}_{n,k}$ is potentially easier than the k-fold direct sum of $\textsc{eq}_n$, and has itself been studied a few times before. Chakrabarti et al. [15] showed that its simultaneous-message complexity is $\Omega(k\sqrt{n})$, which is k times the complexity of $\textsc{eq}_n$ in that model. More recently, Kushilevitz and Weinreb [33] studied the deterministic complexity of $\textsc{oreq}_{n,k}$ under the promise that $x_i = y_i$ for at most one $i \in [k]$. Computing $\textsc{oreq}_{n,k}$ under this "0/1 intersection" promise is closely related to the clique-vs.-independent set problem. In this problem, Alice is given a clique in a graph. Bob is given an independent set, and they must decide if their inputs intersect. Kushilevitz and Weinreb were able to show that computing $\textsc{oreq}_{n,k}$ under this promise still requires $\Omega(kn)$ communication whenever $k \leq n/\log n$. Extending this lower bound to the setting where $k = n$ is an important open problem with several implications.

For the k-DISJ problem, Håstad and Wigderson [25] gave an $O(k)$-bit randomized protocol; a matching lower bound follows easily from the $\Omega(n)$ lower bound for general DISJOINTNESS. The Håstad–Wigderson protocol is clever and crucially exploits both the public randomness and the interactive communication between players. Sağlam and Tardos [42] extend this protocol to interpolate between the one-round and unbounded-round situations, showing that to compute k-DISJ in r rounds, $\Theta(k \operatorname{ilog}^r k)$ bits are necessary and sufficient. This lower bound extends tight $\Omega(k \log k)$ lower bounds for one-round protocols recently given by Dasgupta, Kumar, and Sivakumar [18] and by Buhrman et al. [11]

Since initial announcement of this work [10], we have learned that the communication complexity lower bound of Sağlam and Tardos [42], together with work of Harsha et al. [24] also give lower bounds for information complexity of OR-EQUALITY and similarly DISJOINTNESS. Additionally, with the recent direct product theorem for bounded-round communication complexity of Jain et al. [27] and the existing result equating information and amortized communication of Braverman and Rao [9] these results also extend to give information complexity lower bounds for bounded-round protocols for EQUALITY. Still, EQUALITY is one of the most important communication complexity problems; as such, it deserves careful study. Our information cost lower bounds are more direct and shed more light on this important problem. Additionally, previous results do not differentiate between errors for false positives and false negatives and cannot therefore admit the high false negative rate our bounds apply to.

The recent work of Braverman et al. [7] is similar in spirit to some of our results. They consider zero-error communication protocols for the even more fundamental AND function, obtaining exact information cost bounds. From this they derive nearly exact communication bounds for low-error protocols for DISJOINTNESS and k-DISJ. They also consider rounds-vs.-information tradeoffs for AND, showing that the information complexity of r-round protocols decays as $\Theta(1/r^2)$. Our work shows that the information complexity of EQUALITY decays exponentially with each additional round.

1.4 Road Map

The rest of the paper is organized as follows. Section 2 gives careful definitions of our model of computation and error and cost measures, followed by a listing of all our results. The listing provides pointers to later sections of the paper where these results are proved. Section 3 provides a sketch of our main result, which gives an information cost lower bound for EQUALITY.

We include complete details of our results including full proofs in the Appendix. Section A gives basic definitions and lemmas relating to information theory. The next two sections provide some warm-up. Section B gives upper bounds for EQUALITY including the iterated-log upper bound described informally above. Section C gives matching lower bounds for expected communication cost, first under zero error and then under two-sided error. Though the proofs in Sections B and C are not too complex, the combined story they tell is important. Together, these results paint a nearly complete picture of the behavior of EQUALITY in a bounded-round expected-communication setting, and highlight the importance of studying YES and NO instances separately.

Section D contains the full proof of our Main Theorem, which gives an information cost lower bound for EQUALITY. Section E obtains lower bounds for OREQ and k-DISJ as an application of the Main Theorem. Finally, Section F obtains lower bounds for PRIVATE-INTERSECTION.

2 Definitions and Formal Statement of Results

Throughout this paper we reserve the symbols "n" for input length of EQUALITY instances, "k" for list length of OR-EQUALITY instances and set size of k-DISJ instances, and "N" for universe size of k-DISJ instances. Many definitions and results will be parametrized by these quantities but to keep the notation clean we shall not make this parametrization explicit. We tacitly assume that n, k and N are sufficiently large integers.

Unless otherwise stated, all communication protocols appearing in this paper are public-coin randomized protocols involving two players named Alice and Bob. Because our work concerns expected communication cost in a bounded-round setting, we make the following careful definition of what communication is allowed. In each round, the player whose turn it is to speak sends the other player a message from a *prefix-free* subset[4] of $\{0,1\}^*$. This subset can depend on the communication history. After the final round in the protocol, the player that receives the last message announces the output (which, for us, is always a single bit): this announcement does not count as a round.

Let $\mathcal{P}$ be a communication protocol that takes inputs $(x, y) \in \mathcal{X} \times \mathcal{Y}$. The *transcript* of $\mathcal{P}$ on input (x, y) is defined to be the concatenation of the messages sent by the players, in order, as they execute $\mathcal{P}$ on (x, y). We denote this transcript by $\mathcal{P}(x, y)$ and remark that it is, in general, a random variable. We include the output as the final "message" in the transcript. We denote the output of a transcript $\mathfrak{t}$ by $\operatorname{out}(\mathfrak{t})$. We denote the length of a binary string z by $|z|$. The *communication cost* and *worst-case communication cost* of $\mathcal{P}$ on input (x, y) are defined to be

$$\operatorname{cost}(\mathcal{P}; x, y) := \mathbb{E}\left[|\mathcal{P}(x, y)|\right], \quad \text{and} \quad \operatorname{cost}^*(\mathcal{P}; x, y) := \max |\mathcal{P}(x, y)|,$$

where the expectation and the max are taken over the protocol's random coin tosses.

We now define complexity measures based on this notion of communication cost. Ordinarily we would just define the communication complexity of a function f as the minimum over protocols for f of the worst-case (over all inputs) cost of the protocol. When $f = \text{EQ}_n$, such a measure turns out to be too punishing, and hides the subtleties that we seek to study. Notice that the r-round protocol outlined in Section 1.2 achieves its cost savings only on unequal inputs, i. e., on $f^{-1}(0)$. On inputs in $f^{-1}(1)$, the protocol ends up costing at least n bits. The intuition is that it is much cheaper for Alice and Bob to *refute* the purported equality of their inputs than to *verify* it. Indeed, verification is so hard that interaction has no effect on the verification cost, whereas each additional round of communication decreases refutation cost exponentially.

[4] A set of strings is said to be prefix-free if no string in the set is a proper prefix of any other.

In fact, this intuition can be turned into precise theorems, both in zero-error and positive-error settings, as we shall see. To formalize things, we now define a family of complexity measures.

▶ **Definition 1** (Cost, Error, and Complexity Measures). Let $\mathcal{P}$ be a protocol that is supposed to compute a Boolean function $f : \mathcal{X} \times \mathcal{Y} \to \{0,1\}$. We define its *refutation cost*, *verification cost*, *overall cost*, *refutation error* (or false positive rate, or soundness error), and *verification error* (or false negative rate, or completeness error) as follows, respectively:

$$\begin{aligned}
\mathrm{rcost}(\mathcal{P}) &:= \max\nolimits_{(x,y)\in f^{-1}(0)} \mathrm{cost}(\mathcal{P}; x, y)\,,\\
\mathrm{vcost}(\mathcal{P}) &:= \max\nolimits_{(x,y)\in f^{-1}(1)} \mathrm{cost}(\mathcal{P}; x, y)\,,\\
\mathrm{cost}(\mathcal{P}) &:= \max\nolimits_{(x,y)\in \mathcal{X}\times\mathcal{Y}} \mathrm{cost}(\mathcal{P}; x, y)\,,\\
\mathrm{rerr}(\mathcal{P}) &:= \max\nolimits_{(x,y)\in f^{-1}(0)} \Pr[\mathrm{out}(\mathcal{P}(x,y)) = 1]\,,\\
\mathrm{verr}(\mathcal{P}) &:= \max\nolimits_{(x,y)\in f^{-1}(1)} \Pr[\mathrm{out}(\mathcal{P}(x,y)) = 0]\,.
\end{aligned}$$

Let λ be a probability distribution on the input space $\mathcal{X} \times \mathcal{Y}$. We then define the λ-distributional error $\mathrm{err}^\lambda(\mathcal{P})$ as well as the λ-distributional refutation cost, etc., as follows:

$$\begin{aligned}
\mathrm{rcost}^\lambda(\mathcal{P}) &:= \mathbb{E}_{(X,Y)\sim\lambda}[\mathrm{cost}(\mathcal{P}; X, Y) \mid f(X,Y) = 0]\,,\\
\mathrm{vcost}^\lambda(\mathcal{P}) &:= \mathbb{E}_{(X,Y)\sim\lambda}[\mathrm{cost}(\mathcal{P}; X, Y) \mid f(X,Y) = 1]\,,\\
\mathrm{cost}^\lambda(\mathcal{P}) &:= \mathbb{E}_{(X,Y)\sim\lambda}[\mathrm{cost}(\mathcal{P}; X, Y)]\,,\\
\mathrm{rerr}^\lambda(\mathcal{P}) &:= \mathbb{E}_{(X,Y)\sim\lambda}[\Pr[\mathrm{out}(\mathcal{P}(X,Y)) = 1 \mid f(X,Y) = 0]]\,,\\
\mathrm{verr}^\lambda(\mathcal{P}) &:= \mathbb{E}_{(X,Y)\sim\lambda}[\Pr[\mathrm{out}(\mathcal{P}(X,Y)) = 0 \mid f(X,Y) = 1]]\,,\\
\mathrm{err}^\lambda(\mathcal{P}) &:= \mathbb{E}_{(X,Y)\sim\lambda}[\Pr[\mathrm{out}(\mathcal{P}(X,Y)) \neq f(X,Y)]\,.
\end{aligned}$$

We shall usually restrict $\mathcal{P}$ to be deterministic when considering these distributional measures. Although these measures depend on both $\mathcal{P}$ and f, we do not indicate f in our notation to keep things simple.

Let $r \geq 1$ be an integer and let $\varepsilon, \delta \in [0,1]$ be reals. We define the r-round randomized *refutation complexity* and r-round λ-distributional refutation complexity of f as follows, respectively:

$$\begin{aligned}
\mathrm{R}^{(r),\mathrm{ref}}_{\varepsilon,\delta}(f) &:= \min\{\mathrm{rcost}(\mathcal{P}) : \mathcal{P} \text{ uses } r \text{ rounds, } \mathrm{rerr}(\mathcal{P}) \leq \varepsilon, \mathrm{verr}(\mathcal{P}) \leq \delta\}\,,\\
\mathrm{D}^{\lambda,(r),\mathrm{ref}}_{\varepsilon,\delta}(f) &:= \min\{\mathrm{rcost}^\lambda(\mathcal{P}) : \mathcal{P} \text{ is deterministic and uses } r \text{ rounds, } \mathrm{rerr}^\lambda(\mathcal{P}) \leq \varepsilon,\\
&\qquad \mathrm{verr}^\lambda(\mathcal{P}) \leq \delta\}\,.
\end{aligned}$$

We also define measures of *verification complexity* and *overall complexity* analogously, replacing "rcost" above with "vcost" and "cost" respectively, and denote them by

$$\mathrm{R}^{(r),\mathrm{ver}}_{\varepsilon,\delta}(f)\,,\ \mathrm{D}^{\lambda,(r),\mathrm{ver}}_{\varepsilon,\delta}(f)\,,\ \mathrm{R}^{(r)}_{\varepsilon,\delta}(f)\,,\ \text{and}\ \mathrm{D}^{\lambda,(r)}_{\varepsilon,\delta}(f)\,,$$

respectively. We define the *total complexity* of f as follows:

$$\begin{aligned}
&\mathrm{R}^{*,(r)}_{\varepsilon,\delta}(f) := \min\{\mathrm{cost}^*(\mathcal{P}) : \mathcal{P} \text{ uses } r \text{ rounds, } \mathrm{rerr}(\mathcal{P}) \leq \varepsilon, \mathrm{verr}(\mathcal{P}) \leq \delta\}\,, \text{ where}\\
&\mathrm{cost}^*(\mathcal{P}) := \max\nolimits_{(x,y)\in\mathcal{X}\times\mathcal{Y}} \mathrm{cost}^*(\mathcal{P}; x, y)\,.
\end{aligned}$$

Notice that refutation, verification, and overall complexities use (expected) communication cost as the underlying measure, whereas total complexity uses the (more standard) worst-case communication cost.

▶ **Definition 2** (Information Cost and Complexity)**.** Let $\mathcal{P}, f$, and λ be as above, and suppose the players in $\mathcal{P}$ are allowed to use private coins in addition to a public random string $\mathfrak{R}$. The λ-*information cost* of $\mathcal{P}$ and the r-round λ-*information complexity* of f are defined as follows, respectively:

$$\begin{aligned}&\mathrm{icost}^{\lambda}(\mathcal{P}) := \mathrm{I}(XY : \mathcal{P}(X,Y) \mid \mathfrak{R})\,,\\ &\mathrm{IC}^{\lambda,(r)}_{\varepsilon,\delta}(f) := \inf\{\mathrm{icost}^{\lambda}(\mathcal{P}) : \mathcal{P} \text{ uses } r \text{ rounds, } \mathrm{rerr}(\mathcal{P}) \le \varepsilon, \mathrm{verr}(\mathcal{P}) \le \delta\}\,.\end{aligned}$$

where $\mathrm{I}(\_ : \_ \mid \_)$ denotes conditional mutual information. For readers familiar with recent literature on information complexity [5, 6], we note that this is technically the "external" information cost rather than the "internal" one. However, we shall study information costs mostly with respect to a uniform input distribution, and in this setting there is no difference between external and internal information cost.

It has long been known that information complexity lower bounds standard worst-case communication complexity: this was the main reason for defining the notion [15]. The simple proof boils down to

$$\mathrm{I}(XY : \mathcal{P}(X,Y) \mid \mathfrak{R}) \le \mathrm{H}(\mathcal{P}(X,Y)) \le \max|\mathcal{P}(X,Y)|\,.$$

In our setting, with communication cost defined in the expected sense, it is still the case that

$$\mathrm{IC}^{\lambda,(r)}_{\varepsilon,\delta}(f) \le \mathrm{R}^{(r)}_{\varepsilon,\delta}(f) \tag{1}$$

This time the proof boils down to the inequality $\mathrm{H}(\mathcal{P}(X,Y)) \le \mathbb{E}[\,|\mathcal{P}(X,Y)|\,]$, which follows from Shannon's source coding theorem (see Fact 29 in appendix).

2.1 Summary of Results: Equality

The functions $\textsc{eq}_n$ and $\textsc{oreq}_{n,k}$ have been defined in Section 1 already. To formalize our bounds for these problems, we introduce the iterated logarithm functions $\mathrm{ilog}^k : \mathbb{R}_+ \to \mathbb{R}_+$, which are defined as follows.

$$\begin{aligned}&\mathrm{ilog}^0 z := \max\{1, z\}\,, \quad \forall\, z \in \mathbb{R}_+\,,\\ &\mathrm{ilog}^k z := \max\{1, \log(\mathrm{ilog}^{k-1} z)\}\,, \quad \forall\, k \in \mathbb{N}, z \in \mathbb{R}_+\,.\end{aligned}$$

For all practical purposes, we may pretend that $\mathrm{ilog}^0 = \mathrm{id}$, and $\mathrm{ilog}^k = \log \circ \mathrm{ilog}^{k-1}$, for $k \in \mathbb{N}$.

We use ξ to denote the uniform distribution on $\{0,1\}^n$, and put $\mu := \xi \otimes \xi$. Thus μ is the uniform distribution on inputs to $\textsc{eq}_n$. Strictly speaking these should be denoted ξ_n and μ_n, but we choose to let n be understood from the context. In all our complexity bounds, we tacitly assume that n is sufficiently large. The various parts of the summary theorems below are proved later in the paper, and we indicate on the right where these detailed proofs can be found.

▶ **Theorem 3** (Zero-Error Bounds)**.** *The complexity of* EQUALITY *satisfies the following bounds:*

1. $\mathrm{R}^{(r),\mathrm{ref}}_{0,0}(\textsc{eq}_n) \le \mathrm{ilog}^{r-1} n + 3$.
2. $\mathrm{R}^{(r),\mathrm{ver}}_{0,0}(\textsc{eq}_n) \le n$.
3. $\mathrm{R}^{(r),\mathrm{ref}}_{0,0}(\textsc{eq}_n) = \mathrm{D}^{\mu,(r),\mathrm{ref}}_{0,0}(\textsc{eq}_n) \ge \mathrm{ilog}^{r-1} n - 1$. [Theorem 32]
4. $\mathrm{R}^{(r),\mathrm{ver}}_{0,0}(\textsc{eq}_n) = \mathrm{D}^{\mu,(r),\mathrm{ver}}_{0,0}(\textsc{eq}_n) \ge n$. [Theorem 35]

Notice that these bounds are almost completely tight, differing at most by the tiny additive constant 4. Next, we allow our protocols some error. We continue to have bounds tight up to an additive constant for the verification cost (the case of one-sided error is especially interesting: just set $\delta = 0$ in the results below), and we have bounds tight up to a multiplicative constant in the other cases. To better appreciate the next several bounds, let us first consider the "trivial" one-round protocol for $\textsc{eq}_n$ that achieves ε refutation error. This protocol communicates $\min\{n, \log(1/\varepsilon)\}$ bits: it's as though the instance size drops from n to $\min\{n, \log(1/\varepsilon)\}$ when we allow this refutation error. This motivates the following definition.

▶ **Definition 4** (Effective Instance Size). When considering protocols for $\textsc{eq}_n$ with refutation and verification errors bounded by ε and δ, respectively, we define the effective instance size to be

$$\hat{n} := \min\{n + \log(1-\delta), \log((1-\delta)^2/\varepsilon)\}\,.$$

▶ **Theorem 5** (Two-Sided-Error Bounds). $\textsc{equality}$ *satisfies the following:*

5. $\mathrm{R}^{(r),\mathrm{ref}}_{\varepsilon,\delta}(\textsc{eq}_n) \le (1-\delta)\,\mathrm{ilog}^{r-1}\hat{n} + 5.$ [Corollary 26]

6. $\mathrm{R}^{(r),\mathrm{ver}}_{\varepsilon,\delta}(\textsc{eq}_n) \le (1-\delta)\hat{n} + 3.$ [Corollary 27]

7. $\mathrm{D}^{\mu,(r),\mathrm{ver}}_{\varepsilon,\delta}(\textsc{eq}_n) \ge (1-\delta)(\hat{n}-1)\ .$ [Theorem 43]

8. $\mathrm{R}^{(r),\mathrm{ver}}_{\varepsilon,\delta}(\textsc{eq}_n) \ge \frac{1}{8}(1-\delta)^2(\hat{n} + \log(1-\delta) - 5).$ [Theorem 44]

9. $\mathrm{D}^{\mu,(r),\mathrm{ref}}_{\varepsilon,\delta}(\textsc{eq}_n) = \Omega((1-\delta)^2\,\mathrm{ilog}^{r-1}\hat{n})$. *This bound holds for all* ε, δ *such that* $\delta \le 1 - 2^{-n/2}$ *and* $\varepsilon/(1-\delta)^2 < 1/8$. [Theorem 41]

10. $\mathrm{R}^{(r),\mathrm{ref}}_{\varepsilon,\delta}(\textsc{eq}_n) = \Omega((1-\delta)^3\,\mathrm{ilog}^{r-1}\hat{n})$. *This bound holds for all* ε, δ *such that* $\delta \le 1-2^{-n/2}$ *and* $\varepsilon/(1-\delta)^3 \le 1/64$. [Theorem 42]

Observe that the "constant refutation error" setting $\varepsilon = O(1)$ is not very interesting, as it makes these complexities constant. But observe also that the situation is very different for the verification error, δ: we continue to obtain strong lower bounds even when δ is very close to 1. This is in accordance with our intuition that verification (of equality) is much harder than refutation.

Finally, we turn to information complexity and arrive at the most important result of this paper.

▶ **Theorem 6** (Main Theorem: Information Complexity Bound). *Suppose* $\delta \le 1 - 8(\mathrm{ilog}^{r-2}\hat{n})^{-1/8}$. *Then:*

11. $\mathrm{IC}^{\mu,(r)}_{\varepsilon,\delta}(\textsc{eq}_n) = \Omega((1-\delta)^3\,\mathrm{ilog}^{r-1}\hat{n})$. [Theorem 51]

Applications. As applications, we can recover weaker lower bounds for $\textsc{or-equality}$, $\textsc{disjointness}$, and $\textsc{private-intersection}$. We emphasize that our main result is our thorough study of $\textsc{equality}$, including the direct development of information cost bounds for bounded-round protocols and the analysis of verification vs. refutation error.

2.2 On Yao's Minimax Lemma

Distributional lower bounds imply worst-case randomized ones by an averaging argument that constitutes the "easy" direction of Yao's minimax lemma [43]. Yet, in Theorem 5 we claim somewhat weaker randomized bounds than the corresponding distributional ones. The reason is that in our setting, the averaging argument will need to fix the random coins of a protocol so as to preserve multiple measures (e. g., refutation error as well as cost).

Though this is easily accomplished, we pay a penalty of small constant factor increase in our measures.

Ironically, the "hard" direction of Yao's minimax lemma is particularly easy in the case of EQ_n, because EQUALITY is in a sense *uniform self-reducible.* See Theorem 25, where we show how to turn a protocol designed for the uniform distribution into a randomized one with worst-case guarantees. In this way, *the uniform distribution is provably the hardest distribution for* EQUALITY.

3 Main Theorem: Bounded-Round Information Complexity of Equality

In this section we prove Theorem 6, which we think of as the most important result of this paper. We wish to lower bound the bounded-round information complexity of EQUALITY with respect to the uniform distribution. Recall that we are concerned chiefly with protocols that achieve very low refutation error, though they may have rather high verification error. We will prove our lower bound by proving a round elimination lemma for EQ_n that targets *information* cost, and then applying this lemma repeatedly.

This proof has much more technical complexity than our other lower bound proofs. Let us see why. There are two main technical difficulties and they arise, ultimately, from the same source: the inability to use (the easy direction of) Yao's minimax lemma. When proving a lower bound on *communication* cost, Yao's lemma allows us to fix the random string used by any purported protocol, which immediately moves us into the clean world of deterministic protocols. This hammer is unavailable to us when working with *information* cost. The most we can do is to "average away" the public randomness. We then have to deal with (private coin) randomized protocols the entire way through the round elimination argument. As a result, our intermediate protocols, obtained by eliminating some rounds of our original protocol, do not obey straightforward cost and error guarantees. This is the first technical difficulty, and our solution to it leads us to the concept of a "kernel" in Definition 7 below.

The second technical difficulty is that we are unable to switch to the simpler case of zero verification error like we did in the proof of Theorem 5, Parts (9) and (10). Therefore, all our intermediate protocols continue to have verification error. Since errors scale up with each round elimination, and the verification error starts out high, we cannot afford even a constant-factor scaling. We must play very delicately with our error parameters, which leads us to the somewhat complicated parametrization seen in Definition 8 below.

3.1 The Round Elimination Argument

A standard round elimination argument works by showing that if there is a "good" r-round protocol, then there exists a "good" $(r-1)$-round protocol. What it means to be a "good" protocol is typically parameterized, with the parameters degrading each time a round is eliminated. The trick is to carefully control how the parameters degrade, so that after all communication has been eliminated, a nontrival problem instance remains.

We follow the same approach for our round elimination argument. Central to our parameterization of EQUALITY protocols is the notion of a *kernel.* Roughly speaking, we start by assuming there is an r-round protocol for inputs that are nearly uniformly distributed over some set S, and we show that after elminating the first message, we can construct a protocol for inputs nearly uniformly distributed over a set $S' \subseteq S$. The sets S, S' are our *kernels*, and they capture where the remaining "action" is.

▶ **Definition 7** (Kernel). Let p and q be probability distributions on $\{0,1\}^n$, let $S \subseteq \{0,1\}^n$, and let $\ell \geq 0$ be a real number. The triple (p,q,S) is defined to be an *ℓ-kernel* if the following properties hold.

[K1] $\mathrm{H}(p) \geq n - \ell$ and $\mathrm{H}(q) \geq n - \ell$.
[K2] $p(S) \geq 2^{-\ell}$ and $q(S) \geq \frac{1}{2}$.
[K3] For all $x \in S$ we have $q(x) \geq 2^{-n-\ell}$.

▶ **Definition 8** (Parametrized Protocols). Suppose we have an integer $r \geq 1$, and nonnegative reals ℓ, a, b, and c. A protocol $\mathcal{P}$ for $\textsc{eq}_n$ is defined to be an $[r,\ell,a,b,c]$-protocol if there exists an ℓ-kernel (p,q,S) such that the following properties hold.

[P1] The protocol $\mathcal{P}$ is private-coin and uses r rounds, with Alice speaking in the first round.
[P2] We have $\mathrm{err}^{p\otimes q|S\times S}(\mathcal{P}) = \Pr_{(X,Y)\sim p\otimes q}[\mathrm{out}(\mathcal{P}(X,Y)) \neq \textsc{eq}_n(X,Y) \mid (X,Y) \in S \times S] \leq 2^{-a}$.
[P3] We have $\mathrm{verr}^{p\otimes\xi|S\times S}(\mathcal{P}) = \Pr_{X\sim p}[\mathrm{out}(\mathcal{P}(X,X)) = 0 \mid X \in S] \leq 1 - 2^{-b}$.
[P4] We have $\mathrm{icost}^{p\otimes q}(\mathcal{P}) \leq c$.

We alert the reader to the fact that [P2] considers overall error, and not refutation error. We encourage the reader to take a careful look at [P3] and verify the equality claimed therein. It is straightforward, once one revisits Definition 1 and recalls that ξ denotes the uniform distribution on $\{0,1\}^n$.

Since we have a number of parameters at play, it is worth recording the following simple observation.

▶ **Fact 9.** *Suppose that $\ell' \geq \ell, c' \geq c, a' \leq a$, and $b' \geq b$. Then every ℓ-kernel is also an ℓ'-kernel, and every $[r,\ell,a,b,c]$-protocol is also an $[r,\ell',a',b',c']$-protocol.* ◀

▶ **Theorem 10** (Information-Theoretic Round Elimination for EQUALITY). *If there exists an $[r,\ell,a,b,c]$-protocol with $r \geq 1$ and $c \geq 4$, then there exists an $[r-1,\ell',a',b',c']$-protocol, where*

$$\ell' := (c+\ell)2^{\ell+2b+7}\,, \qquad a' := a - (c+\ell)2^{\ell+2b+8}\,,$$
$$b' := b+2\,, \qquad c' := (c+2)2^{\ell+2b+6}\,.$$

Proof. Let $\mathcal{P}$ be an $[r,\ell,a,b,c]$-protocol, and let (p,q,S) be an ℓ-kernel satisfying the conditions in Definition 8. Assume WLOG that each message in $\mathcal{P}$ is generated using a fresh random string. Let $X \sim p$ and $Y \sim q$ be independent random variables denoting an input to $\mathcal{P}$. Let $M_1,\ldots,M_r$ be random variables denoting the messages sent in $\mathcal{P}$ on input (X,Y), with M_j being the jth message; note that these variables depend on X,Y, and the random strings used by the players. We then have

$$c \geq \mathrm{icost}^{p\otimes q}(\mathcal{P}) = \mathrm{I}(XY : M_1M_2\ldots M_r) = \mathrm{I}(X : M_1) + \mathrm{I}(XY : M_2\ldots M_r \mid M_1)\,, \tag{2}$$

where the final step uses the chain rule for mutual information, and the fact that M_1 and Y are independent. In particular, we have $\mathrm{I}(X : M_1) \leq c$, and so $\mathrm{H}(X \mid M_1) = \mathrm{H}(X) - \mathrm{I}(X : M_1) \geq n - \ell - c$. By Lemma 19,

$$\mathrm{H}(X \mid M_1, X \in S) \geq n - \frac{\ell+c+1}{p(S)} \geq n - (\ell+c+1)2^{\ell}\,. \tag{3}$$

Let $\mathcal{M}$ be the set of messages that Alice sends with positive probability as her first message in $\mathcal{P}$, given the random input X, i.e., $\mathcal{M} := \{\mathfrak{m} : \Pr[M_1 = \mathfrak{m}] > 0\}$. Consider a particular message $\mathfrak{m} \in \mathcal{M}$. Let $\mathcal{P}'_{\mathfrak{m}}$ denote the following protocol for EQ_n. The players simulate $\mathcal{P}$ on their input, except that Alice is assumed to have sent $\mathfrak{m}$ as her first message. As a result, $\mathcal{P}'_{\mathfrak{m}}$ has $r-1$ rounds and Bob is the player to send the first message in $\mathcal{P}'_{\mathfrak{m}}$. Let $\pi_{\mathfrak{m}}$ and q' be the distributions of $(X \mid M_1 = \mathfrak{m} \wedge X \in S)$ and $(Y \mid Y \in S)$, respectively.

Observe that $\text{icost}^{\pi_{\mathfrak{m}} \otimes q'}(\mathcal{P}'_{\mathfrak{m}}) = \mathrm{I}(XY : M_2 \dots M_r \mid M_1 = \mathfrak{m} \wedge (X,Y) \in S \times S)$. Letting L denote a random first message distributed identically to M_1, we now get

$$\begin{aligned}\mathbb{E}_L\left[\text{icost}^{\pi_L \otimes q'}(\mathcal{P}'_L)\right] &= \mathrm{I}(XY : M_2 \dots M_r \mid M_1, (X,Y) \in S \times S) \\ &\le \frac{\mathrm{I}(XY : M_2 \dots M_r \mid M_1) + 1}{p(S)q(S)} \le (c+1)2^{\ell+1}\,, \end{aligned} \tag{4}$$

where the first inequality uses Lemma 18 and the second inequality uses (2) and Property [K2]. Examining Properties [P2] and [P3], we obtain

$$\mathbb{E}_L\left[\text{err}^{\pi_L \otimes q'}(\mathcal{P}'_L)\right] = \text{err}^{p \otimes q | S \times S}(\mathcal{P}) \le 2^{-a}\,, \tag{5}$$

$$\mathbb{E}_L\left[\text{verr}^{\pi_L \otimes \xi}(\mathcal{P}'_L)\right] = \text{verr}^{p \otimes \xi | S \times S}(\mathcal{P}) \le 1 - 2^{-b}\,. \tag{6}$$

▶ **Definition 11** (Good Message). A message $\mathfrak{m} \in \mathcal{M}$ is said to be *good* if the following properties hold:

[G1] $\mathrm{H}(\pi_{\mathfrak{m}}) = \mathrm{H}(X \mid M_1 = \mathfrak{m} \wedge X \in S) \ge n - (\ell + c + 1)2^{\ell+b+3}$,
[G2] $\text{icost}^{\pi_{\mathfrak{m}} \otimes q'}(\mathcal{P}'_{\mathfrak{m}}) \le 2^{\ell+b+4}(c+1)$,
[G3] $\text{err}^{\pi_{\mathfrak{m}} \otimes q'}(\mathcal{P}'_{\mathfrak{m}}) \le 2^{-a+b+3}$,
[G4] $\text{verr}^{\pi_{\mathfrak{m}} \otimes \xi}(\mathcal{P}'_{\mathfrak{m}}) \le 1 - 2^{-b-1}$.

Notice that for all $\mathfrak{m} \in \mathcal{M}$ we have $\mathrm{H}(X \mid M_1 = \mathfrak{m}, X \in S) \le n$. Hence, viewing (3), (4), (5) and (6) as upper bounds on the expected values of certain nonnegative functions of L, we may apply Markov's inequality to these four conditions and conclude that

$$\Pr[L \text{ is good}] \ge 1 - 2^{-b-3} - 2^{-b-3} - 2^{-b-3} - \frac{1-2^{-b}}{1-2^{-b-1}} \ge 2^{-b-1} - 3 \cdot 2^{-b-3} > 0\,.$$

Thus, there exists a good message. *From now on, we fix* $\mathfrak{m}$ *to be such a good message.*

We may rewrite the left-hand side of [G4] as $\mathbb{E}_{Z \sim \pi_{\mathfrak{m}}}[\Pr[\text{out}(\mathcal{P}'_{\mathfrak{m}}(Z,Z)) = 0]]$. So if we define the set $T := \left\{x \in S : \Pr[\text{out}(\mathcal{P}'_{\mathfrak{m}}(x,x)) = 0] \le 1 - 2^{-b-2}\right\}$ and apply Markov's inequality again, we obtain

$$\pi_{\mathfrak{m}}(T) \ge 1 - \frac{1-2^{-b-1}}{1-2^{-b-2}} \ge 2^{-b-2}\,. \tag{7}$$

Defining the distribution $p' := \pi_{\mathfrak{m}} \mid T$ and the set $S' := \{x \in T : p'(x) \ge 2^{-n-\ell'}\}$, we now make two claims.

Claim 1: The triple (q', p', S') is an ℓ'-kernel.
Claim 2: We have $\text{err}^{p' \otimes q' | S' \times S'}(\mathcal{P}'_{\mathfrak{m}}) \le 2^{-a'}$, $\text{verr}^{q' \otimes \xi | S' \times S'}(\mathcal{P}'_{\mathfrak{m}}) \le 1 - 2^{-b'}$, and $\text{icost}^{p' \otimes q'}(\mathcal{P}'_{\mathfrak{m}}) \le c'$.

We prove these claims in Appendix D. Notice that these claims essentially say that $\mathcal{P}'_{\mathfrak{m}}$ has all the properties listed in Definition 8, except that Bob starts $\mathcal{P}'_{\mathfrak{m}}$. Interchanging the roles of Alice and Bob in $\mathcal{P}'_{\mathfrak{m}}$ gives us the desired $[r-1, \ell', a', b', c']$-protocol, which completes the proof of the theorem. ◀

The following easy corollary of Theorem 10 will be useful shortly; we defer its proof to Appendix D.

▶ **Corollary 12.** *Let $\tilde{n}, j, r \in \mathbb{N}$ and $a, b \in \mathbb{R}$ with $\tilde{n}$ sufficiently large, $j \geq 1$, $r \geq 1$, and $b \geq 0$. Suppose there exists an $[r, \ell, a-\ell, b, \ell]$-protocol, with $b \leq \ell = \frac{1}{8} \operatorname{ilog}^j \tilde{n}$. Then there exists an $[r-1, \ell', a-\ell', b+2, \ell']$-protocol with $b+2 \leq \ell' = (\operatorname{ilog}^{j-1} \tilde{n})^{1/2} \leq \frac{1}{8} \operatorname{ilog}^{j-1} \tilde{n}$.*

3.2 Finishing the Proof

We are now ready to state and prove the main lower bound on protocols with two-sided error.

▶ **Theorem 13** (Restatement of Main Theorem). *Let $\tilde{n} = \min\{n + \log(1-\delta), \log((1-\delta)/\varepsilon)\}$. Suppose $\delta \leq 1 - 8(\operatorname{ilog}^{r-2} \tilde{n})^{-1/8}$. Then we have $\mathrm{IC}^{\mu,(r)}_{\varepsilon,\delta}(\textsc{eq}_n) = \Omega((1-\delta)^3 \operatorname{ilog}^{r-1} \tilde{n})$.*

Proof. We may assume that $r \leq \log^* \tilde{n}$, for otherwise there is nothing to prove. The slight difference between $\tilde{n}$ above and $\hat{n}$, as in Definition 4, is insignificant and can be absorbed by the $\Omega(\cdot)$ notation.

Suppose, to the contrary, that there exists an r-round randomized protocol $\mathcal{P}^*$ for $\textsc{eq}_n$, with $\operatorname{rerr}^\mu(\mathcal{P}^*) \leq \varepsilon$, $\operatorname{verr}^\mu(\mathcal{P}^*) \leq \delta$ and $\operatorname{icost}^\mu(\mathcal{P}^*) \leq 2^{-16}(1-\delta)^3 \operatorname{ilog}^{r-1} \tilde{n}$. Recall that we denote the uniform distribution on $\{0,1\}^n$ by ξ and that $\mu = \xi \otimes \xi$. We have

$$\operatorname{err}^\mu(\mathcal{P}^*) = (1 - 2^{-n}) \operatorname{rerr}^\mu(\mathcal{P}^*) + 2^{-n} \operatorname{verr}^\mu(\mathcal{P}^*) \leq \varepsilon + 2^{-n}(\delta - \varepsilon) \leq \varepsilon + 2^{-n} \,.$$

Let $\mathcal{P}^*_s$ be the private-coin protocol for $\textsc{eq}_n$ obtained from $\mathcal{P}^*$ by fixing the public random string of $\mathcal{P}^*$ to be s. We have $\mathbb{E}_s[\operatorname{err}^\mu(\mathcal{P}^*_s)] \leq \varepsilon + 2^{-n}$, $\mathbb{E}_s[\operatorname{verr}^\mu(\mathcal{P}^*_s)] \leq \delta$, and $\mathbb{E}_s[\operatorname{icost}(\mathcal{P}^*_s)] \leq 2^{-16}(1-\delta)^3 \operatorname{ilog}^{r-1} \tilde{n}$. By Markov's inequality, there exists s such that $\mathcal{P}^*_s$ simultaneously has $\operatorname{err}^\mu(\mathcal{P}^*_s) \leq 4(\varepsilon + 2^{-n})/(1-\delta)$, $\operatorname{verr}^\mu(\mathcal{P}^*_s) \leq (1+\delta)/2$, and $\operatorname{icost}(\mathcal{P}^*_s) \leq 2^{-14}(1-\delta)^2 \operatorname{ilog}^{r-1} \tilde{n}$: this is because

$$1 - \frac{1-\delta}{4} - \frac{2\delta}{1+\delta} - \frac{1-\delta}{4} = \frac{(1-\delta)^2}{2(1+\delta)} > 0 \,.$$

Let $\mathcal{P} = \mathcal{P}^*_s$ for this s. Then $(\xi, \xi, \{0,1\}^n)$ is a 0-kernel and $\mathcal{P}$ is an $[r,\, 0,\, \log \frac{1-\delta}{4(\varepsilon+2^{-n})},\, \log \frac{2}{1-\delta},\, 2^{-14}(1-\delta)^2 \operatorname{ilog}^{r-1} \tilde{n}]$-protocol. Recalling Fact 9 and using $\log \frac{1-\delta}{\varepsilon+2^{-n}} \geq \tilde{n} - 1$, we see that

$$\mathcal{P} \text{ is an } \left[r,\, 0,\, \tilde{n} - 3,\, \log \tfrac{1}{1-\delta} + 1,\, 2^{-14}(1-\delta)^2 \operatorname{ilog}^{r-1} \tilde{n}\right]\text{-protocol.}$$

Put $\ell_j := \frac{1}{8} \operatorname{ilog}^j \tilde{n}$ for $j \in \mathbb{N}$. Applying round elimination (Theorem 10) to $\mathcal{P}$ and weakening the resulting parameters (using Fact 9) gives us an $[r-1, \ell_{r-1}, \tilde{n} - \ell_{r-1}, \log \frac{1}{1-\delta} + 3, \ell_{r-1}]$-protocol $\mathcal{P}'$.

The upper bound on δ gives us $\log \frac{1}{1-\delta} + 3 \leq \ell_{r-1}$, and so the conditions for Corollary 12 apply. Starting with $\mathcal{P}'$ and applying that corollary repeatedly, each time using the looser estimate on ℓ' in that corollary, we obtain a sequence of protocols with successively fewer rounds. Eventually we reach a $[1, \ell_1, \tilde{n} - \ell_1, \log \frac{1}{1-\delta} + 2(r-1) + 1, \ell_1]$-protocol. Applying Theorem 10 one more time, and using the tighter estimate on ℓ' this time, we get a $[0,\, \tilde{n}^{1/2},\, \tilde{n} - \tilde{n}^{1/2},\, \log \frac{1}{1-\delta} + 2r + 1,\, \tilde{n}^{1/2}]$-protocol $\mathcal{Q}$. Weakening parameters again, we see that $\mathcal{Q}$ is a $[0,\, \tilde{n}^{1/2},\, \frac{1}{2}\tilde{n},\, \frac{1}{3} \log \tilde{n},\, \tilde{n}^{1/2}]$-protocol. Let (p, q, S) be the $\tilde{n}^{1/2}$-kernel for $\mathcal{Q}$. By Property [K1], we have $\mathrm{H}(q) \geq n - \tilde{n}^{1/2}$. Using Lemma 19 and Property [K2], we then have

$$\mathrm{H}(q \mid S) \geq n - \frac{\tilde{n}^{1/2} + 1}{q(S)} \geq n - (2\tilde{n}^{1/2} + 2) \,. \tag{8}$$

Since $\mathcal{Q}$ involves no communication, it must behave identically on any two input distributions that have the same marginal on Alice's input. In particular, this gives us the following crucial equation:

$$\Pr_{X\sim p}[\mathrm{out}(\mathcal{Q}(X,X)) = 1 \mid X \in S] = \Pr_{(X,Y)\sim p\otimes q}[\mathrm{out}(\mathcal{Q}(X,Y)) = 1 \mid (X,Y) \in S \times S]\,. \tag{9}$$

Let α denote the above probability. Considering the left-hand side of (9), we have

$$\alpha = 1 - \mathrm{verr}^{p\otimes\xi|S\times S}(\mathcal{Q}) \geq 2^{-\frac{1}{3}\log\tilde{n}} = \tilde{n}^{-1/3}\,. \tag{10}$$

On the other hand, whenever $\mathcal{Q}$ outputs 1 on an input (x,y), then either $x = y$ or $\mathcal{Q}$ errs on (x,y). Therefore, considering the right-hand side of (9), we have

$$\begin{aligned}
\alpha &\leq \Pr_{(X,Y)\sim p\otimes q}[X = Y \mid (X,Y) \in S \times S] + \\
&\qquad\qquad \Pr_{(X,Y)\sim p\otimes q}[\mathrm{out}(\mathcal{P}(X,Y)) \neq \mathrm{EQ}_n(X,Y) \mid (X,Y) \in S \times S] \\
&\leq \max_{x\in S}\Pr_{Y\sim q|S}[Y = x] + \mathrm{err}^{p\otimes q|S\times S}(\mathcal{Q}) \\
&\leq \frac{2\tilde{n}^{1/2} + 3}{n} + 2^{-\frac{1}{2}\tilde{n}} && (11) \\
&\leq 2\tilde{n}^{-1/2} + 3\tilde{n}^{-1} + 2^{-\frac{1}{2}\tilde{n}}\,, && (12)
\end{aligned}$$

where (11) follows from (8) by applying Lemma 20, and (12) uses $\tilde{n} \leq n$.

The bounds (10) and (12) are in contradiction for sufficiently large $\tilde{n}$, which completes the proof. ◀

References

1 Farid Ablayev. Lower bounds for one-way probabilistic communication complexity and their application to space complexity. *Theoretical Computer Science*, 175(2):139–159, 1996.

2 Anil Ada, Arkadev Chattopadhyay, Stephen A. Cook, Lila Fontes, Michal Koucký, and Toniann Pitassi. The hardness of being private. In *IEEE Conference on Computational Complexity*, pages 192–202, 2012.

3 Noga Alon, Yossi Matias, and Mario Szegedy. The space complexity of approximating the frequency moments. *J. Comput. Syst. Sci.*, 58(1):137–147, 1999. Preliminary version in *Proc. 28th Annual ACM Symposium on the Theory of Computing*, pages 20–29, 1996.

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

5 Boaz Barak, Mark Braverman, Xi Chen, and Anup Rao. How to compress interactive communication. In *Proc. 41st Annual ACM Symposium on the Theory of Computing*, pages 67–76, 2010.

6 Mark Braverman. Interactive information complexity. In *Proc. 44th Annual ACM Symposium on the Theory of Computing*, pages 505–524, 2012.

7 Mark Braverman, Ankit Garg, Denis Pankratov, and Omri Weinstein. From information to exact communication. In *Proc. 45th Annual ACM Symposium on the Theory of Computing*, 2013. to appear.

8 Mark Braverman and Ankur Moitra. An information complexity approach to extended formulations. In *Proc. 45th Annual ACM Symposium on the Theory of Computing*, 2013. to appear.

9 Mark Braverman and Anup Rao. Information equals amortized communication. In *Proc. 52nd Annual IEEE Symposium on Foundations of Computer Science*, pages 748–757, 2011.

10 Joshua Brody, Amit Chakrabarti, and Ranganath Kondapally. Certifying equality with limited interaction. Technical Report TR12-153, ECCC, 2012.

11 Harry Buhrman, David García-Soriano, Arie Matsliah, and Ronald de Wolf. The non-adaptive query complexity of testing k-parities. *arXiv preprint arXiv:1209.3849*, 2012.

12 Amit Chakrabarti, Graham Cormode, Ranganath Kondapally, and Andrew McGregor. Information cost tradeoffs for augmented index and streaming language recognition. In *Proc. 51st Annual IEEE Symposium on Foundations of Computer Science*, pages 387–396, 2010.

13 Amit Chakrabarti, Subhash Khot, and Xiaodong Sun. Near-optimal lower bounds on the multi-party communication complexity of set disjointness. In *Proc. 18th Annual IEEE Conference on Computational Complexity*, pages 107–117, 2003.

14 Amit Chakrabarti and Ranganath Kondapally. Everywhere-tight information cost tradeoffs for augmented index. In *Proc. 15th International Workshop on Randomization and Approximation Techniques in Computer Science*, pages 448–459, 2011.

15 Amit Chakrabarti, Yaoyun Shi, Anthony Wirth, and Andrew C. Yao. Informational complexity and the direct sum problem for simultaneous message complexity. In *Proc. 42nd Annual IEEE Symposium on Foundations of Computer Science*, pages 270–278, 2001.

16 Amit Chakrabarti and Anna Shubina. Nearly private information retrieval. In *Proc. 32nd International Symposium on Mathematical Foundations of Computer Science*, volume 4708 of *Lecture Notes in Computer Science*, pages 383–393, 2007.

17 Thomas M. Cover and Joy A. Thomas. *Elements of Information Theory*. Wiley-Interscience [John Wiley & Sons], Hoboken, NJ, second edition, 2006.

18 Anirban Dasgupta, Ravi Kumar, and D. Sivakumar. Sparse and lopsided set disjointness via information theory. In *16th International workshop on Randomization*, volume 7409, pages 517–528, 2012.

19 Ronald Fagin, Moni Naor, and Peter Winkler. Comparing information without leaking it. *Commun. ACM*, 39(5):77–85, 1996.

20 Tomas Feder, Eyal Kushilevitz, Moni Naor, and Noam Nisan. Amortized communication complexity. *SIAM J. Comput.*, 24(4):736–750, 1995. Preliminary version in *Proc. 32nd Annual IEEE Symposium on Foundations of Computer Science*, pages 239–248, 1991.

21 Michael J. Freedman, Kobbi Nissim, and Benny Pinkas. Efficient private matching and set intersection. In *EUROCRYPT*, pages 1–19, 2004.

22 Rusins Freivalds. Probabilistic machines can use less running time. In *IFIP Congress*, pages 839–842, 1977.

23 Andre Gronemeier. Asymptotically optimal lower bounds on the NIH-multi-party information complexity of the AND-function and disjointness. In *Proc. 26th International Symposium on Theoretical Aspects of Computer Science*, pages 505–516, 2009.

24 Prahladh Harsha, Rahul Jain, David McAllester, and Jaikumar Radhakrishnan. The communication complexity of correlation. In *Proc. 22nd Annual IEEE Conference on Computational Complexity*, pages 10–23, 2007.

25 Johan Håstad and Avi Wigderson. The randomized communication complexity of set disjointness. *Theory of Computing*, pages 211–219, 2007.

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

27 Rahul Jain, Attila Pereszlényi, and Penghui Yao. A direct product theorem for the two-party bounded-round public-coin communication complexity. In *Proc. 53rd Annual IEEE Symposium on Foundations of Computer Science*, pages 167–176, 2012.

28 Rahul Jain, Pranab Sen, and Jaikumar Radhakrishnan. Optimal direct sum and privacy trade-off results for quantum and classical communication complexity. *CoRR*, abs/0807.1267, 2008.

29 Bala Kalyanasundaram and Georg Schnitger. The probabilistic communication complexity of set intersection. *SIAM J. Disc. Math.*, 5(4):547–557, 1992.

30 Iordanis Kerenidis and Ronald de Wolf. Exponential lower bound for 2-query locally decodable codes. *J. Comput. Syst. Sci.*, 69(3):395–420, 2004. Preliminary version in *Proc. 35th Annual ACM Symposium on the Theory of Computing*, pages 106–115, 2003.

31 Hartmut Klauck. On quantum and approximate privacy. In *STACS*, pages 335–346, 2002.

32 Eyal Kushilevitz and Noam Nisan. *Communication Complexity.* Cambridge University Press, Cambridge, 1997.

33 Eyal Kushilevitz and Enav Weinreb. The communication complexity of set-disjointness with small sets and 0-1 intersection. In *Proc. 50th Annual IEEE Symposium on Foundations of Computer Science*, pages 63–72, 2009.

34 Frédéric Magniez, Claire Mathieu, and Ashwin Nayak. Recognizing well-parenthesized expressions in the streaming model. In *Proc. 41st Annual ACM Symposium on the Theory of Computing*, pages 261–270, 2010.

35 Kurt Mehlhorn and Erik M. Schmidt. Las Vegas is better than determinism in VLSI and distributed computing (extended abstract). In *Proc. 14th Annual ACM Symposium on the Theory of Computing*, pages 330–337, 1982.

36 Peter Bro Miltersen, Noam Nisan, Shmuel Safra, and Avi Wigderson. On data structures and asymmetric communication complexity. *J. Comput. Syst. Sci.*, 57(1):37–49, 1998. Preliminary version in *Proc. 27th Annual ACM Symposium on the Theory of Computing*, pages 103–111, 1995.

37 M. Molinaro, D.P. Woodruff, and G. Yaroslavtsev. Beating the direct sum theorem in communication complexity with implications for sketching. In *Proc. 24th Annual ACM-SIAM Symposium on Discrete Algorithms*, 2013.

38 Moni Naor and Benny Pinkas. Oblivious polynomial evaluation. *SIAM J. Comput.*, 35(5):1254–1281, 2006.

39 Ashwin Nayak. Optimal lower bounds for quantum automata and random access codes. In *Proc. 40th Annual IEEE Symposium on Foundations of Computer Science*, pages 124–133, 1999.

40 Mihai Pătraşcu. Unifying the landscape of cell-probe lower bounds. *SIAM J. Comput.*, 40(3):827–847, 2011.

41 Alexander Razborov. On the distributional complexity of disjointness. *Theor. Comput. Sci.*, 106(2):385–390, 1992. Preliminary version in *Proc. 17th International Colloquium on Automata, Languages and Programming*, pages 249–253, 1990.

42 Mert Saglam and Gábor Tardos. On the communication complexity of sparse set disjointness and exists-equal problems. In *Proc. 54th Annual IEEE Symposium on Foundations of Computer Science*, pages 678–687, 2013.

43 Andrew C. Yao. Probabilistic computations: Towards a unified measure of complexity. In *Proc. 18th Annual IEEE Symposium on Foundations of Computer Science*, pages 222–227, 1977.

44 Andrew C. Yao. Some complexity questions related to distributive computing. In *Proc. 11th Annual ACM Symposium on the Theory of Computing*, pages 209–213, 1979.

A Information theory

A.1 Basic Probability, Properties of Entropy and Mutual Information

We will use the following fact about collision probability of a random function.

▶ **Fact 14.** *Given a subset $S \subseteq [n]$ for size $|S| \geq 2$, $i \geq 0$ and $t = \Theta(|S|^{i+2})$, a random function $h\colon [n] \to [t]$ has no collisions with probability at least $1 - 1/|S|^i$, namely for all $x, y \in S$ such that $x \neq y$ it holds that $h(x) \neq h(y)$. Moreover, a random hash function satisfying such guarantee can be constructed using only $O(\log n)$ random bits.*

▶ **Definition 15.** Let λ be a probability distribution on a finite set S and let $T \subseteq S$ be an event with $\lambda(T) \neq 0$. We write $\lambda \mid T$ to denote the distribution obtained by conditioning λ on T. To be explicit, $\lambda \mid T$ is given by

$$(\lambda \mid T)(x) = \begin{cases} 0\,, & \text{if } x \notin T\,, \\ \lambda(x)/\lambda(T)\,, & \text{if } x \in T\,. \end{cases}$$

Also, we write $\mathrm{H}(\lambda)$ to denote the entropy of a random variable distributed according to λ, i. e., $\mathrm{H}(\lambda) = \mathrm{H}(X)$, where $X \sim \lambda$.

▶ **Lemma 16** (Equivalent to Lemma 30). *With λ, S and T as above, let $f : S \to \mathbb{R}_+$ be a nonnegative function. Then $\mathbb{E}_{X\sim\lambda|T}[f(X)] \leq \mathbb{E}_{X\sim\lambda}[f(X)]/\lambda(T)$.* ◀

We give a summary of basic properties of the entropy of a discrete random variable X, denoted as $H(X)$, and the mutual information between two discrete random variables X and Y, denoted as $\mathrm{I}(X : Y) = H(X) - H(X \mid Y)$, below (see Chapter 2 in [17] for the proofs). We denote the support of a random variable X as $supp(X)$.

▶ **Proposition 17.** **1.** *Entropy span: $0 \leq H(X) \leq \log|supp(X)|$.*
2. *$\mathrm{I}(X : Y) \geq 0$ because $H(X \mid Y) \leq H(X)$.*
3. *Chain rule: $\mathrm{I}(X_1, X_2, \ldots, X_n : Y \mid Z) = \sum_{i=1}^{n} \mathrm{I}(X_i : Y \mid X_1, \ldots, X_{i-1}, Z)$.*
4. *Subadditivity: $H(X, Y \mid Z) \leq H(X \mid Z) + H(Y \mid Z)$, where the equality holds if and only if X and Y are independent conditioned on Z.*
5. *Fano's inequality: Let A be a random variable, which can be used as "predictor" of X, namely there exists a function g such that $\Pr[g(A) = X] \geq 1 - \delta$ for some $\delta < 1/2$. If $|supp(X)| \geq 2$ then $H(X \mid A) \leq \delta\log(|supp(X)| - 1) + h_2(\delta)$, where $h_2(\delta) = \delta\log(1/\delta) + (1-\delta)\log\frac{1}{1-\delta}$ is the binary entropy.*

▶ **Lemma 18.** *Let Z, W be jointly distributed random variables. Let $\mathcal{E}$ be an event. Then,*

$$\mathrm{I}(Z : W) \geq \Pr[\mathcal{E}]\,\mathrm{I}(Z : W \mid \mathcal{E}) - 1\,.$$

Proof. Let D be the indicator random variable for $\mathcal{E}$. Then we have

$$\mathrm{I}(Z : W \mid D) = \Pr[\mathcal{E}]\,\mathrm{I}(Z : W \mid \mathcal{E}) + \Pr[\neg\mathcal{E}]\,\mathrm{I}(Z : W \mid \neg\mathcal{E}) \geq \Pr[\mathcal{E}]\,\mathrm{I}(Z : W \mid \mathcal{E})\,. \tag{13}$$

Note that $\mathrm{I}(Z : D \mid W) \leq \mathrm{H}(D \mid W) \leq \mathrm{H}(D) \leq 1$. Using the chain rule for mutual information twice, we get

$$\mathrm{I}(Z : W \mid D) \leq \mathrm{I}(Z : WD) = \mathrm{I}(Z : W) + \mathrm{I}(Z : D \mid W) \leq \mathrm{I}(Z : W) + 1\,. \tag{14}$$

The lemma follows by combining inequalities (13) and (14). ◀

To appreciate the next two lemmas, it will help to imagine that $d \ll n$.

▶ **Lemma 19.** *Let Z, W be jointly distributed random variables, with Z taking values in $\{0,1\}^n$, and let $\mathcal{E}$ be an event. Then*

$$\mathrm{H}(Z \mid W) \geq n - d \quad \Longrightarrow \quad \mathrm{H}(Z \mid W, \mathcal{E}) \geq n - (d+1)/\Pr[\mathcal{E}]\,.$$

In particular, taking W to be a constant, we have $\mathrm{H}(Z) \geq n - d \implies \mathrm{H}(Z \mid \mathcal{E}) \geq n - (d+1)/\Pr[\mathcal{E}]$.

Proof. We use the fact that the entropy of Z can be at most n, even after arbitrary conditioning. This gives

$$\begin{aligned} n - d &\leq \mathrm{H}(Z \mid W) \\ &= \Pr[\mathcal{E}]\ \mathrm{H}(Z \mid W, \mathcal{E}) + (1 - \Pr[\mathcal{E}])\,\mathrm{H}(Z \mid W, \neg\mathcal{E}) + \mathrm{H}_b(\Pr[\mathcal{E}]) \\ &\leq \Pr[\mathcal{E}]\ \mathrm{H}(Z \mid W, \mathcal{E}) + (1 - \Pr[\mathcal{E}])n + 1\,, \end{aligned}$$

where $\mathrm{H}_b(x) := -x \log x - (1-x)\log(1-x)$. The lemma follows by rearranging the above inequality. ◀

▶ **Lemma 20.** *Let Z be a random variable taking values in $\{0,1\}^n$ and let $z \in \{0,1\}^n$. Then*

$$\mathrm{H}(Z) \geq n - d \quad \Longrightarrow \quad \Pr[Z = z] \leq (d+1)/n\,.$$

Proof. The lemma follows by rearranging the following inequality, which is a consequence of Lemma 19:

$$0 = \mathrm{H}(Z \mid Z = z) \geq n - \frac{d+1}{\Pr[Z = z]}\,. \qquad ◀$$

A.2 Protocols with Abortion

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

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

▶ **Definition 21** (Protocols with Abortion). Consider a communication problem given by $f : \mathcal{X} \times \mathcal{Y} \to \mathcal{Z}$ and a probability distribution μ over $\mathcal{X} \times \mathcal{Y}$. We say that a deterministic protocol $\mathcal{P}_D$ *(β, δ)-computes* f with respect to μ if it satisfies the following (where $(X, Y) \sim \mu$):

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

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

Using the definitions above we can now introduce the notion of information complexity for protocols with abortions formally.

▶ **Definition 22** (Information Complexity for Protocols with Abortion)**.** Let $\mathcal{P}$ be a protocol, which computes f. The *conditional information cost* of $\mathcal{P}$ under λ is defined as $\mathrm{I}(\mathcal{P}(X,Y) : X,Y \mid D)$, where $(X,Y,D) \sim \lambda$. The *conditional information complexity* of f with respect to λ, denoted by $\mathrm{IC}_{\mu,\delta}(f \mid \nu)$, is defined as the minimum conditional information cost of a δ-protocol for f. The *information complexity with aborts*, denoted by $\mathrm{IC}^{\mu}_{\alpha,\beta,\delta}(f \mid \nu)$, is the minimum conditional information cost of a protocol that (α,β,δ)-computes f. The analogous quantities $\mathrm{IC}^{\mu,(r)}_{\delta}(f \mid \nu)$ and $\mathrm{IC}^{\mu,(r)}_{\alpha,\beta,\delta}(f \mid \nu)$ are defined by taking the respective minimums over only r-round protocols.

B Upper Bounds

In this section, we provide deterministic and randomized protocols for EQ_n with low refutation cost and low verification cost. Recall Definition 4, which introduced the quantity $\hat{n} = \min\left\{n + \log(1-\delta), \log\frac{(1-\delta)^2}{\varepsilon}\right\}$ as the effective instance size. One can derive one-sided-error and zero-error versions of these results by setting δ and/or ε to zero as needed, and using the convention $\log(w/0) = +\infty$ for $w > 0$. One can in fact tighten the analysis for the case $\varepsilon = \delta = 0$ to obtain the bounds in Theorem 3.

▶ **Theorem 23.** *Suppose $n, r \in \mathbb{N}$ and $\varepsilon, \delta \in [0,1]$ are such that $\delta < 1 - 2^{-n/2}$ and $\operatorname{ilog}^{r-1} \hat{n} \geq 4$. Then*

$$\mathrm{D}^{\mu,(r),\mathrm{ref}}_{\varepsilon,\delta}(\mathrm{EQ}_n) \leq (1-\delta)\operatorname{ilog}^{r-1}\hat{n} + 5\,.$$

Proof. To gain intuition, we first consider $\delta = 0$, in which case we have $\hat{n} = \min\{n, \log(1/\varepsilon)\}$. The basic idea was already outlined in Section 1. Since we need only handle a random input, we do not need fingerprints. Instead, Alice and Bob take turns revealing increasingly longer prefixes of their inputs: in the jth round, the player to speak sends the next $\approx \operatorname{ilog}^{r-j}\hat{n}$ bits of her input. Whenever a player witnesses a mismatch in prefixes, she *aborts* (and the protocol outputs 0). If the protocol ends without an abortion, it outputs 1. The protocol described so far clearly has no false negatives, and after filling in some details (see below), we can show that it has the desired refutation cost and refutation error.

To achieve further savings for nonzero δ, we partition $\{0,1\}^n$ into sets $S, T \subseteq \{0,1\}^n$ such that $|S| \approx (1-\delta)2^n$. Each player aborts the protocol at her first opportunity if her input lies in T. Otherwise, they emulate the above protocol on the smaller input space $S \times S$.

We now make things precise. Set

$$\begin{aligned} n' &:= n + \lceil \log(1-\delta) \rceil\,, \\ n'' &:= \min\{n', 2 + \lceil \log((1-\delta)^2/\varepsilon) \rceil\}\,, \\ t_j &:= \begin{cases} \lceil \operatorname{ilog}^{r-j}\hat{n} \rceil\,, & \text{if } 1 \leq j < r\,, \\ n'' - \sum_{j=1}^{r-1} t_j\,, & \text{if } j = r\,. \end{cases} \end{aligned}$$

Choose an arbitrary partition of $\{0,1\}^n$ into subsets S and T such that $|S| = 2^{n'}$. Fix an arbitrary bijection $g : S \to \{0,1\}^{n'}$.

The protocol—which we call $\mathcal{P}$—works as follows on input $(x,y) \in \{0,1\}^n \times \{0,1\}^n$. We write $x[i_1 : i_2]$ to denote the substring $x_{i_1}x_{i_1+1}\ldots x_{i_2}$ of x. Each nonempty message in the protocol will be either the string "0", indicating abortion, or "1" followed by a *payload* string. Each player maintains a variable ℓ that records the length of the prefix that has been compared so far; initially they set $\ell \leftarrow 0$.

The players keep track of whether an abortion has occurred. Once an abortion occurs, all further messages in the protocol will be empty strings. Once r rounds have been completed, the appropriate player will output 0 if an abortion has occurred, and 1 otherwise.

Round j proceeds as follows. Let $P \in \{\text{Alice, Bob}\}$ be the player who speaks in this round, and let $z \in \{x, y\}$ be their input. If necessary, P aborts if $z \in T$. Now suppose that an abortion has not yet occurred. If $j = 1$, then P sends the substring $g(z)[1 : t_1]$, sets $\ell \leftarrow t_1$, and the round ends. Otherwise, suppose P receives a non-aborting message with payload w. If P finds that $w \neq g(z)[\ell + 1 : \ell + t_{j-1}]$ then she aborts, else if $j < r$, she continues the protocol by sending the next t_j bits of $g(z)$, i. e., she sends $g(z)[\ell + t_{j-1} + 1 : \ell + t_{j-1} + t_j]$, sets $\ell \leftarrow \ell + t_{j-1} + t_j$, and the round ends.

The protocol's logic is shown in pseudocode form below, for readers who prefer that presentation.

Algorithm 1: Round j of the protocol $\mathcal{P}$. Here $t_0 = 0$ and "Round $r + 1$" is the output announcement.

```
if $j \leq r$ then
    if aborted then send emptystring ;
    else
        if $z \in T$ then abort;
        $w \leftarrow$ payload of most recently received message ;
        if $w \neq g(z)[\ell + 1 : \ell + t_{j-1}]$ then abort;
        send "1" followed by $g(x)[\ell + t_{j-1} + 1 : \ell + t_{j-1} + t_j]$, and set $\ell \leftarrow \ell + t_{j-1} + t_j$ ;
else
    if aborted then output 0 ;
    else
        $w \leftarrow$ payload of most recently received message ;
        if $w \neq g(z)[\ell + 1 : \ell + t_{j-1}]$ then output 0 else output 1 ;
```

It is easy to see that $\text{verr}^{\mu}(\mathcal{P}) \leq \delta$, since players only abort an (x, x) input when $x \in T$. Next, note that a false positive occurs only when $(x, y) \in S \times S$ and $g(x)[1 : n''] = g(y)[1 : n'']$. When $n'' = n'$ (which corresponds, roughly, to $\varepsilon < (1 - \delta)2^{-n}$), Alice and Bob end up comparing all bits of $g(x)$ and $g(y)$, and we get $\text{rerr}^{\mu}(\mathcal{P}) = 0$. In the other case, we have $n'' = 2 + \lceil \log((1 - \delta)^2/\varepsilon) \rceil$. Letting $(X, Y) \sim \mu$, we have

$$\begin{aligned} \text{rerr}^{\mu}(\mathcal{P}) &= \Pr[(X, Y) \in S \times S \mid X \neq Y] \cdot \Pr\left[g(X)[1 : n''] = g(Y)[1 : n''] \mid g(X) \neq g(Y)\right] \\ &\leq \left(2^{n'-n}\right)^2 \cdot \frac{2^{n'-n''} - 1}{2^{n'} - 1} \leq 2^{2\lceil \log(1-\delta) \rceil} \cdot 2^{-n''} \leq 2^{2(1+\log(1-\delta))} \cdot \frac{\varepsilon}{4(1 - \delta)^2} = \varepsilon \, . \end{aligned}$$

Finally, we analyze the refutation cost. Let a_j denote the expected total communication in rounds $\geq j$, conditioned on not aborting before round j. For convenience, set $a_{r+1} = 0$. We claim that $a_j \leq 3$ for all $j > 2$ and prove so by induction from $r + 1 \rightsquigarrow 3$. The base case ($j = r + 1$) is trivial. Conditioned on not aborting before the jth round, the player whose turn it is to speak receives t_{j-1} bits to compare with her own input. Estimating as above, this will fail to cause an abortion with probability at most $2^{-t_{j-1}}$. Therefore, the player to speak will send at most 1 bit in this round to indicate abortion (or not) plus, with probability at most $2^{-t_{j-1}}$, will continue the communication, which will cost t_j bits in this round and a_{j+1} bits in expectation in subsequent rounds. The net result is that

$$a_j \le 1 + 2^{-t_{j-1}}(t_j + a_{j+1}) \le 1 + \frac{1}{\operatorname{ilog}^{r-j} d}\left(\lceil \operatorname{ilog}^{r-j} d\rceil + 3\right) \le 2 + \frac{4}{\operatorname{ilog}^{r-j} d} \le 3\,.$$

The first two rounds are slightly different, because each player summarily aborts when her input lies in T. In the first round, Alice aborts with probability at most δ. In the second round, conditioned on Alice not aborting, Bob aborts with probability all but $(1-\delta)2^{-t_1}$. The refutation cost of r-round protocols is therefore bounded by

$$\begin{aligned}\operatorname{rcost}^\mu(\mathcal{P}) = a_1 &\le 1 + (1-\delta)t_1 + (1-\delta)\left(1 + (1-\delta)2^{-t_1}(t_2 + a_3)\right)\\ &\le 1 + (1-\delta)(\lceil \operatorname{ilog}^{r-1}\hat{n}\rceil + 1) + (1-\delta)^2\frac{\lceil \operatorname{ilog}^{r-2}\hat{n}\rceil + 3}{\operatorname{ilog}^{r-2}\hat{n}}\\ &\le 1 + (1-\delta)\operatorname{ilog}^{r-1}\hat{n} + 2(1-\delta) + (1-\delta)^2\left(1 + \frac{4}{\operatorname{ilog}^{r-2}\hat{n}}\right)\\ &\le 1 + (1-\delta)\operatorname{ilog}^{r-1}\hat{n} + 2(1-\delta) + 2(1-\delta)^2\\ &\le 5 + (1-\delta)\operatorname{ilog}^{r-1}\hat{n}\,.\end{aligned}$$ ◀

▶ **Theorem 24.** *With $n, r, \varepsilon, \delta$ as above, we have $\mathrm{D}^{\mu,(r),\mathrm{ver}}_{\varepsilon,\delta}(\mathrm{EQ}_n) \le (1-\delta)\hat{n} + 3$.*

Proof. We construct a *one-round* protocol achieving the stated verification cost, using S, T, g as in Theorem 23. On input (x, y), Alice aborts if $x \in T$. Otherwise, she sends Bob a prefix of $g(x)$ of length $\min\{n + \lceil\log(1-\delta)\rceil, 2 + \lceil\log((1-\delta)^2/\varepsilon)\rceil$. Bob outputs 0 ("unequal") if (i) Alice aborted, (ii) $y \in T$, or (iii) Alice's prefix does not match that of $g(y)$.

As in the previous proof, this protocol—call it $\mathcal{Q}$—only produces false negatives when inputs lie in T, so that $\operatorname{verr}^\mu(\mathcal{Q}) \le \delta$. And as before, we get $\operatorname{rerr}^\mu(\mathcal{Q}) = 0$ for small ε and $\operatorname{rerr}^\mu(\mathcal{Q}) \le 2^{2\lceil\log(1-\delta)\rceil} \cdot \frac{\varepsilon}{4(1-\delta)^2} \le \varepsilon$ otherwise. As for verification cost, the protocol always sends a bit to indicate abortion (or not), and for all $(x, x) \in S \times S$ the protocol sends at most $\hat{n} + 2$ bits. Thus, $\operatorname{vcost}^\mu(\mathcal{Q}) \le 1 + (1-\delta)(\hat{n} + 2) \le (1-\delta)\hat{n} + 3$. ◀

▶ **Theorem 25.** *Let $\mathcal{P}$ be an r-round deterministic protocol for EQ_n. Then, there exists an r-round randomized protocol $\mathcal{Q}$ for EQ_n with $\operatorname{verr}(\mathcal{Q}) = \operatorname{verr}^\mu(\mathcal{P})$, $\operatorname{rerr}(\mathcal{Q}) = \operatorname{rerr}^\mu(\mathcal{P})$, $\operatorname{rcost}(\mathcal{Q}) = \operatorname{rcost}^\mu(\mathcal{P})$, and $\operatorname{vcost}(\mathcal{Q}) = \operatorname{vcost}^\mu(\mathcal{P})$.*

Proof. Construct $\mathcal{Q}$ as follows. Alice and Bob use public randomness to generate a uniform bijection $G : \{0,1\}^n \to \{0,1\}^n$. On input (x, y), they run $\mathcal{P}$ on $(G(x), G(y))$. Note that if $x = y$ then $(G(x), G(y))$ is uniform over $\mathrm{EQ}_n^{-1}(1)$, and if $x \ne y$ then $(G(x), G(y))$ is uniform over $\mathrm{EQ}_n^{-1}(0)$. Thus, distributional guarantees for $\mathcal{P}$ under the uniform distribution become worst-case guarantees for $\mathcal{Q}$. ◀

Together with Theorems 23 and 24, this gives upper bounds for randomized protocols.

▶ **Corollary 26.** $\mathrm{R}^{(r),\mathrm{ref}}_{\varepsilon,\delta}(\mathrm{EQ}_n) \le (1-\delta)\operatorname{ilog}^{r-1}\hat{n} + 5$.

▶ **Corollary 27.** $\mathrm{R}^{(r),\mathrm{ver}}_{\varepsilon,\delta}(\mathrm{EQ}_n) \le (1-\delta)\hat{n} + 3$.

C Bounded-Round Communication Lower Bounds for Equality

In this section, we prove all of our communication cost lower bounds on EQ_n. We deal with information cost in the next section. We think of these lower bounds as "combinatorial" (as opposed to "information theoretic"). An important ingredient in some of these combinatorial lower bounds is the *round elimination* technique, which dates back to the work of Miltersen et al. [36].

C.1 Preliminaries

We recall two well-known results from information theory (see, e. g., Cover and Thomas [17]), and state a convenient estimation lemma. The second fact below is one direction of Shannon's source coding theorem. It states that any prefix-free code must have expected length at least the entropy of the source.

▶ **Fact 28** (Kraft Inequality). *Let $S \subseteq \{0,1\}^*$ be a prefix-free set. Then*

$$\sum_{x \in S} 2^{-|x|} \leq 1\,.$$

▶ **Fact 29** (Source Coding Theorem). *Let X be a random variable taking values in a prefix-free set $S \subseteq \{0,1\}^*$. Then*

$$\mathbb{E}[\,|X|\,] \geq \mathrm{H}(X)\,.$$

▶ **Lemma 30.** *Let X, X' be uniformly distributed over sets $\mathcal{X}, \mathcal{X}'$, respectively, with $\mathcal{X}' \subseteq \mathcal{X}$. Let $f : \mathcal{X} \to \mathbb{R}_+$ be a nonnegative function. Then, we have $\mathbb{E}_{X'}[f(X')] \leq \big(|\mathcal{X}|/|\mathcal{X}'|\big)\ \mathbb{E}_X[f(X)]$.*

Proof. By the nonnegativity of f, we have

$$\mathbb{E}_X[f(X)] = \frac{1}{|\mathcal{X}|}\sum_{x\in\mathcal{X}} f(x) \geq \frac{1}{|\mathcal{X}|}\sum_{x\in\mathcal{X}'} f(x) = \left(\frac{|\mathcal{X}'|}{|\mathcal{X}|}\right)\frac{1}{|\mathcal{X}'|}\sum_{x\in\mathcal{X}'} f(x) = \frac{|\mathcal{X}'|}{|\mathcal{X}|}\,\mathbb{E}_{X'}[f(X')]\,. \blacktriangleleft$$

▶ **Lemma 31.** *For $a \leq 2^{n/2}$, $t \leq \log^* n - 2$, and $x \in \left[\frac{1}{a}, 1\right]$, we have $\operatorname{ilog}^{t-1} n \geq \operatorname{ilog}^t(2^n x) \geq \left(1 - \frac{\log a}{n}\right)\operatorname{ilog}^{t-1} n$.*

Proof. The upper bound is trivial. We prove the lower bound by induction on t. We have $\log(2^n x) = n + \log x \geq n - \log a > \left(1 - \frac{\log a}{n}\right)n$, and the claim holds for $t = 1$. For $t > 1$, we have

$$\begin{aligned}
\operatorname{ilog}^t(2^n x) &\geq \log\left(1 - \frac{\log a}{n}\right) + \log\left(\operatorname{ilog}^{t-2} n\right) && \text{[by induction hypothesis]}\\
&\geq -\frac{2\log a}{n} + \operatorname{ilog}^{t-1} n && \text{[using } 1 - w \geq 2^{-2w} \text{ for } 0 \leq w \leq 1/2\text{]}\\
&\geq \left(1 - \frac{\log a}{n}\right)\operatorname{ilog}^{t-1} n && \text{[using } \operatorname{ilog}^{t-1} n \geq 2\text{]}\ . \qquad \blacktriangleleft
\end{aligned}$$

C.2 Lower Bounds for Zero-Error Protocols

In this section, we provide nearly exact bounds for zero-error protocols.

▶ **Theorem 32.** *For all $r < \log^* n$ we have $\mathrm{D}_{0,0}^{\mu,(r),\mathrm{ref}}(\mathrm{EQ}_n) \geq \operatorname{ilog}^{r-1} n - 1$.*

To prove this theorem, we must analyze EQUALITY protocols on finite sets of arbitrary size. Given a finite set S, define EQ_S to be the EQUALITY problem, but when $x, y \in S$.

▶ **Theorem 33.** *For all integers $r > 0$, we have $\mathrm{D}_{0,0}^{\mu,(r),\mathrm{ref}}(\mathrm{EQ}_S) \geq \operatorname{ilog}^r |S| - 1$.*

Proof. Assume $\operatorname{ilog}^r |S| > 1$ as otherwise there is nothing to prove. Define m to be the unique real such that $m = \log|S|$. It might be helpful to think of m as an integer, but this is not necessary.

The proof proceeds by induction on r. When $r = 1$, Alice must send her entire input to achieve zero error in a single round. This costs $\lceil m \rceil > \operatorname{ilog}^1 m - 1$ bits, and the theorem holds. Now, assume $\mathrm{D}_{0,0}^{\mu,(\ell),\mathrm{ref}}(\mathrm{EQ}_T) \geq \operatorname{ilog}^{\ell}|T| - 1$ for all finite sets T, and let $\mathcal{P}$ be an optimal $(\ell+1)$-round deterministic protocol for EQ_S. We aim to show that $\mathrm{rcost}^{\mu}(\mathcal{P}) \geq \operatorname{ilog}^{\ell+1}|S| - 1 = \operatorname{ilog}^{\ell} m - 1$. Let $\mathfrak{m}_1, \ldots, \mathfrak{m}_t$ be the possible messages Alice sends in the first round of $\mathcal{P}$. For $1 \leq i \leq t$, Let A_i denote the set of inputs on which Alice sends $\mathfrak{m}_i$, and let ℓ_i denote the length of $\mathfrak{m}_i$. Assume without loss of generality that $\ell_1 \leq \ell_2 \leq \cdots \leq \ell_t$. Since $\mathcal{P}$ is optimal, we must have $|A_1| \geq |A_2| \geq \cdots \geq |A_t|$: otherwise, we can permute which messages are sent on which sets A_i and reduce the overall cost of the protocol.

We analyze the cost of $\mathcal{P}$ by conditioning on Alice's first message. Under the uniform distribution, Alice sends $\mathfrak{m}_i$ with probability $p_i := |A_i|/2^m$. If $y \notin A_i$, Bob refutes equality and the protocol aborts. Thus, over $x \neq y$ inputs, the probability that Bob aborts is $(|A_i|-1)/(2^m-1)$. Furthermore, conditioned on the events that (i) Alice's first message is $\mathfrak{m}_i$ and that (ii) Bob doesn't abort, Alice and Bob's inputs are each uniform over A_i. Thus, the remaining communication is at least $\mathrm{D}_{0,0}^{\mu,(\ell),\mathrm{ref}}(\mathrm{EQ}_{A_i})$.

Fix $\tau := 2/\operatorname{ilog}^{\ell-1} m$. Call the ith message *small* if $p_i \leq \tau$ and *large* otherwise. We bound

$$
\begin{aligned}
\mathrm{rcost}^{\mu}(\mathcal{P}) &= \sum_{1\leq i\leq t} p_i\Big(\ell_i + \frac{|A_i|-1}{2^m-1}\,\mathrm{D}_{0,0}^{\mu,(\ell),\mathrm{ref}}(\mathrm{EQ}_{A_i})\Big)\\
&\geq \sum_{1\leq i\leq t} p_i\Big(-\log p_i + (p_i - 2^{-m})\,\mathrm{D}_{0,0}^{\mu,(\ell),\mathrm{ref}}(\mathrm{EQ}_{A_i})\Big)\\
&\geq \sum_{\text{small } \mathfrak{m}_i} p_i(-\log p_i) + \sum_{\text{large } \mathfrak{m}_i} p_i\Big(-\log p_i + (p_i - 2^{-m})(\operatorname{ilog}^{\ell}|A_i| - 1)\Big)\\
&\geq \Pr[\text{small message}]\cdot(\operatorname{ilog}^{\ell}(m) - 1) + \\
&\qquad\qquad \sum_{\text{large } \mathfrak{m}_i} p_i\Big(-\log p_i + p_i \operatorname{ilog}^{\ell}|A_i| - p_i - 1\Big)\\
&= \Pr[\text{small message}]\cdot(\operatorname{ilog}^{\ell}(m) - 1) + \sum_{\text{large } \mathfrak{m}_i} p_i f(p_i)\,,
\end{aligned}
$$

where we define $f(x) := -\log x + x \operatorname{ilog}^{\ell}(2^m x) - x - 1$. The first inequality holds by the source coding theorem (Fact 29) and the third inequality holds because $p_i \leq \tau$ for all small messages.

We now claim that $f'(x) > 0$ for all $x \in [\tau, 1]$. We prove this by explicitly calculating the derivative of f. If $x \geq \tau$, then $-1/(x\ln 2) \geq -\operatorname{ilog}^{\ell-1}(m)/(2\ln 2)$. By Lemma 31, we have

$$
\begin{aligned}
f'(x) &= -\frac{1}{x\ln 2} + \operatorname{ilog}^{\ell}(2^m x) - \frac{1}{(\ln 2)(\ln x\cdot 2^m)\prod_{j=0}^{\ell-2}\ln(\operatorname{ilog}^{j} x\cdot 2^m)} - 1\\
&\geq -\frac{\operatorname{ilog}^{\ell-1} m}{2\ln 2} + \operatorname{ilog}^{\ell-1} m - \frac{(\operatorname{ilog}^{\ell-1} m)\operatorname{ilog}^{\ell} m}{m} - o(1) - 1\\
&= \left(\operatorname{ilog}^{\ell-1} m\right)\left(1 - \frac{1}{2\ln 2}\right) - 1 - o(1) = \Omega(\operatorname{ilog}^{\ell-1} m)\,,
\end{aligned}
$$

which proves the claim. It now follows that for large messages, $f(p_i)$ is minimized at $f(\tau)$. Note that

$$
\begin{aligned}
f(\tau) &= -\log\tau + \tau\operatorname{ilog}^{\ell}(2^m\tau) - \tau - 1\\
&\geq \operatorname{ilog}^{\ell} m - 1 + \frac{2}{\operatorname{ilog}^{\ell-1} m}\operatorname{ilog}^{\ell-1} m\Big(1 - \frac{\operatorname{ilog}^{\ell}(m) - 1}{m}\Big) - \frac{2}{\operatorname{ilog}^{\ell-1} m} - 1\\
&> \operatorname{ilog}^{\ell} m - 1\,.
\end{aligned}
$$

Plugging this back into our inequality for the cost of $\mathcal{P}$, we get

$$\mathrm{rcost}^{\mu}(\mathcal{P}) \geq \Pr[\text{small message}]\cdot(\mathrm{ilog}^{\ell} m-1)+\Pr[\text{large message}]\cdot(\mathrm{ilog}^{\ell} m-1) = \mathrm{ilog}^{\ell} m-1\,. \blacktriangleleft$$

▶ **Theorem 34.** $\mathrm{D}_{0,0}^{\mu,(r),\mathrm{ver}}(\mathrm{EQ}_n) \geq n$. *Note that this lower bound is independent of r.*

Proof. Let $\mathcal{P}$ be a deterministic zero-error protocol for EQ_n. As the protocol has no error, the communication matrix is partitioned into monochromatic rectangles. In particular, there are 2^n 1-rectangles, since each (x, x) input must map to a different rectangle.[5] Let R_x, T_x, and ℓ_x denote the rectangle consisting of the input pair (x, x), the protocol transcript corresponding to (x, x), and the length of this protocol transcript, respectively. Note that $\{T_x\}$ form a prefix-free coding of $\{0, 1\}^n$. By Kraft's inequality, we have $\sum_x 2^{-\ell_x} \leq 1$. Therefore, in expectation $\mathbb{E}[2^{-\ell_x}] \leq 2^{-n}$, and by Jensen's inequality, we get the following.

$$-n \geq \log \mathbb{E}[2^{-\ell_x}] \geq \mathbb{E}[\log(2^{-\ell_x})] = -\,\mathbb{E}[\ell_x]\,.$$

Multiplying each side of the inequality by -1, we have $\mathbb{E}_x[\ell_x] \geq n$. This is precisely $\mathrm{vcost}^{\mu}(\mathcal{P})$, thus the proof is complete. ◀

▶ **Theorem 35.** $\mathrm{R}_{0,0}^{(r),\mathrm{ver}}(\mathrm{EQ}_n) \geq n$. *As above, this lower bound is independent of r.*

Proof. Let $\mathcal{P}$ be a randomized zero-error protocol for EQ_n. Given any string s, let $\mathcal{P}_s$ denote the deterministic protocol obtained by fixing the public randomness to s. Proceeding along the same lines as in the proof of Theorem 34, we have $\mathbb{E}[\ell_{x,s}] \geq n$, where $\ell_{x,s}$ is the length of the protocol transcript in $\mathcal{P}_s$ on input (x, x). This holds for every $\mathcal{P}_s$, hence $\mathbb{E}_{x,s}[\ell_{x,s}] \geq n$. Therefore, there exists x such that $\mathbb{E}_s[\ell_{x,s}] \geq n$. Recalling the definition of vcost, we have $\mathrm{vcost}(\mathcal{P}) \geq \mathrm{cost}(\mathcal{P}; x, x) = \mathbb{E}_s[\ell_{x,s}] \geq n$, completing the proof. ◀

C.3 Refutation Lower Bounds for Protocols with Two-Sided Error

In this section, we give combinatorial lower bounds on the refutation cost of EQUALITY protocols that admit error. All of the bounds in this section will be asymptotic rather than nearly exact. For this reason, we will strive for simplicity of the proofs at the possible expense of some technical accuracy. For instance, we will often drop ceilings or floors in the mathematical notation. We will also assume that players have the ability to instantly abort a protocol when equality has been refuted. This is easily implemented, as seen in Section C.2 at negligible communication cost. We prefer to avoid the technical machinery needed to express this explicitly.

▶ **Definition 36.** An $\langle n, r, \varepsilon, \delta, c\rangle$-EQUALITY protocol $\mathcal{P}$ is a r-round deterministic protocol with $\mathrm{rerr}^{\mu}(\mathcal{P}) \leq \varepsilon$, $\mathrm{verr}^{\mu}(\mathcal{P}) \leq \delta$, and $\mathrm{rcost}^{\mu}(\mathcal{P}) \leq c$.

For the sake of brevity, we often drop the "EQUALITY" and simply refer to an $\langle n, r, \varepsilon, \delta, c\rangle$-protocol. Our first lemma demonstrates that disallowing false negatives changes the communication complexity very little.

▶ **Lemma 37.** *If there exists a $\langle n, r, \varepsilon, \delta, c\rangle$-*EQUALITY *protocol, then there exists a $\langle n', r, \varepsilon', 0, c'\rangle$-*EQUALITY *protocol, where $n' = n + \log(1 - \delta)$, $\varepsilon' = 2\varepsilon/(1 - \delta)^2$, and $c' = 2c/(1 - \delta)^2$.*

[5] If (x, x) and (y, y) were in the same rectangle, then so would (x, y) and (y, x). Thus, the protocol would err on these inputs.

Proof. Let $S = \{x : \mathrm{out}(\mathcal{P}(x,x)) = 0\}$ be the set of inputs on which $\mathcal{P}$ gives a false negative, and let $T = \{0,1\}^n \setminus S$. Since $\mathcal{P}$ has false negative rate δ under the uniform distribution, we have $|T| \geq (1-\delta)2^n = 2^{n'}$.

First create a new $\textsc{eq}_n$ protocol $\mathcal{P}'$ which works as follows. On input (x,y), Alice aborts and outputs 0 if $x \in S$; otherwise, the players emulate $\mathcal{P}$ and output $\mathrm{out}(\mathcal{P}(x,y))$. Note that $\mathcal{P}'$ makes precisely the same false negatives as in $\mathcal{P}$, and aborting when $x \in S$ can only decrease the false positive rate and the expected communication on inputs in $\textsc{eq}_n^{-1}(0)$. Thus, $\mathcal{P}'$ is also a $\langle n, r, \varepsilon, \delta, c\rangle$-protocol.

Next, fix an arbitrary bijection $g : \{0,1\}^{n'} \to T$, and construct an $\textsc{eq}_{n'}$ protocol $\mathcal{Q}$ in the following way. On input (X,Y), players emulate $\mathcal{P}'$ on input $(g(X), g(Y))$ and output $\mathrm{out}(\mathcal{P}'(g(X), g(Y)))$. Note that $g(X), g(Y) \in T$, so there are no false negatives. There can be as many false positives as in $\mathcal{P}'$. However, the sample space is smaller ($2^{2n'} - 2^{n'}$ vs $2^{2n} - 2^n$), so the false positive rate can increase. By Lemma 30, the overall error is at most $2\varepsilon/(1-\delta)^2$. Similarly, the communication in $\mathcal{Q}$ on any input (X,Y) is the same as the communication in $\mathcal{P}'$ on input $(g(X), g(Y))$, but since the sample space is smaller (again $2^{2n'} - 2^{n'}$ vs. $2^{2n} - 2^n$), the expected communication can increase. However, the overall increase in communication is at most a factor of $2/(1-\delta)^2$ by Lemma 30. ◀

▶ **Lemma 38** (Combinatorial Round Elimination for $\textsc{equality}$). *If there is an $\langle n, r, \varepsilon, 0, c\rangle$-$\textsc{equality}$ protocol, then there is an $\langle n - 3c - 2, r - 1, 12\varepsilon 2^{3c}, 0, 12c2^{3c}\rangle$-$\textsc{equality}$ protocol.*

Proof. Let $\mathcal{P}$ be a $\langle n, r, \varepsilon, 0, c\rangle$-protocol. Let $Z(x,y) = 1$ if the protocol errs on input (x,y), and let $Z(x,y) = 0$ otherwise. Then we have

$$\mathbb{E}_x\left[\mathbb{E}_{y\neq x}[|\mathcal{P}(x,y)|]\right] \leq c, \quad \text{and} \quad \mathbb{E}_x\left[\mathbb{E}_{y\neq x}[Z(x,y)]\right] \leq \varepsilon\ .$$

Call x good if (1) $\mathbb{E}_{y\neq x}[\mathcal{P}(x,y)|] \leq 3c$, and (2) $\mathbb{E}_{y\neq x}[Z(x,y)] \leq 3\varepsilon$. By two applications of Markov's inequality and a union bound, at least $2^n/3$ x are good. Next, fix Alice's first message m so it is constant over the maximal number of good x. It follows that m is constant over a set A of good x of size $|A| \geq 2^{n-3c-2}$. This induces a $(r-1)$-round protocol $\mathcal{Q}$ for $\textsc{eq}_A$. It remains to bound the cost and error of $\mathcal{Q}$. Applying Lemma 30 twice, we have that the cost and error are bounded by (respectively)

$$\begin{aligned}
\mathrm{rcost}^\mu(\mathcal{Q}) &= \mathbb{E}_{x\in A}\left[\mathbb{E}_{y\in A, y\neq x}[|\mathcal{P}(x,y)|]\right] \leq \frac{2^n}{2^{n-3c-2}}\, \mathbb{E}_{x\in A}\left[\mathbb{E}_{y\in\{0,1\}^n, y\neq x}[|\mathcal{P}(x,y)|]\right] \\
&\leq 12c2^{3c}\,, \\
\mathrm{verr}^\mu(\mathcal{Q}) &= \mathbb{E}_{x\in A}\left[\mathbb{E}_{y\in A, y\neq x}[Z(x,y)]\right] \leq \frac{2^n}{2^{n-3c-2}}\, \mathbb{E}_{x\in A}\left[\mathbb{E}_{y\in\{0,1\}^n, y\neq x}[Z(x,y)]\right] \\
&\leq 12\varepsilon 2^{3c}\,.
\end{aligned}$$

◀

▶ **Corollary 39.** *Let n, j, r, d be integers with $n > d$, d sufficiently large, and $r \geq 1$. Suppose there exists an $\langle n, r, \varepsilon\ell, 0, \ell\rangle$-protocol, where $\ell = \frac{1}{6}\,\mathrm{ilog}^j d$. Then, there exists an $\langle n - 3\ell - 2, r - 1, \varepsilon\ell', 0, \ell'\rangle$-protocol with $\ell' = \frac{1}{6}\,\mathrm{ilog}^{j-1} d$.*

Proof. This boils down to the following estimations, which are valid for all sufficiently large d.

$$12\ell 2^{3\ell} = 2(\mathrm{ilog}^j d)2^{\frac{1}{2}\,\mathrm{ilog}^j d} = 2\,\mathrm{ilog}^j d\sqrt{\mathrm{ilog}^{j-1} d} < \frac{1}{6}\,\mathrm{ilog}^{j-1} d\,.$$

◀

▶ **Theorem 40** (Lower Bound for Protocols with False Negatives Disallowed). *Let n be a sufficiently large integer, $\varepsilon < 1/4$ a real, and $r \geq 1$. Fix $\tilde{n} := \min\{n, \log(1/\varepsilon)\}$. Then, $\mathrm{D}^{\mu,(r),\mathrm{ref}}_{\varepsilon,0}(\textsc{eq}_n) = \Omega(\mathrm{ilog}^{r-1}\tilde{n})$.*

Proof. In this proof we tacitly assume $\operatorname{ilog}^{r-1}\tilde{n} \ge 100$.

Suppose for the sake of a contradiction that there exists a $\langle n, r, \varepsilon, 0, \frac{1}{6}\operatorname{ilog}^{r-1}\tilde{n}\rangle$-protocol $\mathcal{P}$. Applying Lemma 38 gives an $\langle n - \frac{3}{5}\operatorname{ilog}^{r-1}\tilde{n},\ r-1,\ \frac{\varepsilon}{6}\operatorname{ilog}^{r-2}\tilde{n},\ 0,\ \frac{1}{6}\operatorname{ilog}^{r-2}\tilde{n}\rangle$-protocol $\mathcal{P}'$. Next, applying Corollary 39 repeatedly, a total of $r-2$ times, gives an $\langle n - \frac{3}{5}\sum_{j=1}^{r-1}\operatorname{ilog}^{j}\tilde{n}, 1, \frac{\varepsilon}{6}\tilde{n},\ 0,\ \frac{\tilde{n}}{6}\rangle$-protocol. Finally, applying Lemma 38 once more gives an $\langle n - \frac{3}{5}\sum_{j=0}^{r-1}\operatorname{ilog}^{j}\tilde{n},\ 0,\ 2\varepsilon\tilde{n}2^{\tilde{n}/2},\ 0,\ 2\tilde{n}2^{\tilde{n}/2}\rangle$-protocol $\mathcal{Q}$.

Note that since $\mathcal{Q}$ has false negative rate zero, $\mathcal{Q}$ must output 1 with certainty. Thus, $\mathcal{Q}$ errs on all $X \neq Y$ inputs; i. e., $\mathcal{Q}$ has false positive rate 1. On the other hand, $\tilde{n} \le \log(1/\varepsilon)$, so the false positive rate of $\mathcal{Q}$ is $2\varepsilon\tilde{n}2^{\tilde{n}/6} \le \sqrt{\varepsilon} < 1/2$. This is a contradiction as long as the problem remains nontrivial.

Since $\operatorname{ilog}^{j}\tilde{n} \ge 100$, we have $\sum_{j=t+1}^{r-1}\operatorname{ilog}^{j}\tilde{n} < \frac{1}{5}\operatorname{ilog}^{t}\tilde{n}$. Also notice that since $\tilde{n} \le n$, we have $n - \frac{3}{5}\sum_{j=0}^{r-1}\operatorname{ilog}^{j}\tilde{n} > n/5$. Thus, we have a zero-round protocol for $\textsc{eq}_{n'}$ for some $n' = \Omega(n)$ that has false positive rate $< 1/2$ but must output 1 with certainty, a contradiction. ◀

▶ **Theorem 41** (Lower Bound for Protocols with Two-Sided Error). *Let n be a sufficiently large integer, and let ε, δ be reals such that $\delta \le 1 - 2^{-n/2}$ and $\varepsilon/(1-\delta)^2 < 1/8$. Let $\hat{n}$ be as given in Definition 4. Then, $\mathrm{D}^{\mu,(r),\mathrm{ref}}_{\varepsilon,\delta}(\textsc{eq}_n) = \Omega((1-\delta)^2\operatorname{ilog}^{r-1}\hat{n})$.*

Proof. Fix $d = \min\{n/2, \log((1-\delta)^2/2\varepsilon)\}$, so that $\log d = \Theta(\log\hat{n})$. Suppose, to the contrary, that there exists an $\langle n, r, \varepsilon, \delta, \frac{1}{12}(1-\delta)^2\operatorname{ilog}^{r-1}d\rangle$-protocol $\mathcal{P}$. Since $n + \log(1-\delta) > n/2$, Lemma 37 gives an $\langle n/2, r, 2\varepsilon/(1-\delta)^2, 0, \frac{1}{6}\operatorname{ilog}^{r-1}d\rangle$-protocol. The rest of the proof echoes the proof of Theorem 40. ◀

Next, we prove a combinatorial lower bound for randomized communication complexity.

▶ **Theorem 42.** *Let n be a sufficiently large integer, ε and δ reals such that $\delta < 1 - 2^{1-n/2}$ and $64\varepsilon < (1-\delta)^3$. Then, $\mathrm{R}^{(r),\mathrm{ref}}_{\varepsilon,\delta}(\textsc{eq}_n) = \Omega((1-\delta)^3\operatorname{ilog}^{r-1}\hat{n})$, where $\hat{n}$ is as in Definition 4.*

Proof. Let $\mathcal{P}$ be an r-round randomized protocol with $\operatorname{rerr}(\mathcal{P}) = \varepsilon, \operatorname{verr}(\mathcal{P}) = \delta$, and $\operatorname{rcost}^{\mu}(\mathcal{P}) = c$. Define $z = 1-\delta, \hat{\varepsilon} = 4\varepsilon/(1-\delta)$, and $\hat{c} = 4c/(1-\delta)$. Let $\mathcal{P}_s$ denote the deterministic protocol obtained from $\mathcal{P}$ by setting its random string to s. Call a string s good if (i) $\operatorname{verr}^{\mu}(\mathcal{P}_s) \le 1 - z/2$, (ii) $\operatorname{rerr}^{\mu}(\mathcal{P}_s) \le \hat{\varepsilon}$, and (iii) $\operatorname{rcost}^{\mu}(\mathcal{P}_s) \le \hat{c}$. Applying a Markov argument to each of these three conditions, we see that

$$\Pr[s \text{ is bad}] < \frac{1-z}{1-z/2} + \frac{z}{4} + \frac{z}{4} < 1\,,$$

where we used $(1-z)/(1-z/2) < 1 - z/2$. Thus there exists a good string s. Note that $\mathcal{P}_s$ is a $[n, r, \hat{\varepsilon}, \hat{\delta}, \hat{c}]$-protocol, and by Theorem 41, $\hat{c} = \Omega((1-\delta)^2\operatorname{ilog}^{r-1}\hat{n})$. Therefore, $c = \Omega((1-\delta)^3\operatorname{ilog}^{r-1}\hat{n})$. ◀

C.4 Verification Lower Bounds for Protocols with Two-Sided Error

▶ **Theorem 43.** $\mathrm{D}^{\mu,(r),\mathrm{ver}}_{\varepsilon,\delta}(\textsc{eq}_n) \ge (1-\delta)(\hat{n}-1)$*, where $\hat{n}$ is as in Definition 4.*

Proof. Fix a deterministic protocol $\mathcal{P}$ achieving $\operatorname{rerr}^{\mu}(\mathcal{P}) = \varepsilon$ and $\operatorname{verr}^{\mu}(\mathcal{P}) = \delta$. This protocol naturally partitions the communication matrix for $\textsc{eq}_n$ into combinatorial rectangles. Let $R_1, \ldots, R_c$ be the rectangles on which $\mathcal{P}$ outputs 1. Let s_i denote the number of (x,x) inputs in R_i. Since $\mathcal{P}$ has false negative rate δ, we have $\sum_i s_i = 2^n(1-\delta)$. Let $p_i = s_i/2^n$ and $q_i = p_i/(1-\delta)$. Notice that p_i is the probability that $(x,x) \in R_i$ for a uniformly chosen x. Similarly, q_i is the probability that $(x,x) \in R_i$ conditioned on $\mathcal{P}$ verifying equality on

(x, x). We now analyze the false positive rate. Recall that there are $2^{2n} - 2^n$ total $x \neq y$ inputs. It is easy to see that R_i contains at least $s_i^2 - s_i$ false positives. Therefore, we have

$$\varepsilon \geq \frac{1}{2^{2n} - 2^n} \sum_{i=1}^{c} (s_i^2 - s_i) = \sum_{i=1}^{c} \frac{s_i(s_i - 1)}{2^n(2^n - 1)} \geq \sum_{i=1}^{c} p_i(p_i - 2^{-n}) = -2^{-n}(1 - \delta) + \sum_{i=1}^{c} p_i^2 \,.$$

Rearranging terms and noting that $q_i = p_i/(1 - \delta)$, we have

$$\mathbb{E}[q_i] = \sum_{i=1}^{c} q_i^2 = \frac{1}{(1-\delta)^2} \sum_{i=1}^{c} p_i^2 \leq \frac{1}{(1-\delta)^2} \left(\varepsilon + 2^{-n}(1 - \delta)\right) = \frac{\varepsilon}{(1-\delta)^2} + \frac{2^{-n}}{(1-\delta)} \leq 2 \cdot 2^{-\hat{n}} \,.$$

Next, we analyze the verification cost of $\mathcal{P}$. Let ℓ_i denote the length of the protocol transcript for inputs in the rectangle R_i. Observe that the transcripts $\mathcal{P}(x, x)$ with $\mathrm{out}(\mathcal{P}(x, x)) = 1$ give a prefix-free encoding of the set of rectangles $\{R_1, \ldots, R_c\}$. Therefore,

$$\begin{aligned} \mathrm{vcost}^{\mu}(\mathcal{P}) &= \sum_{x \in \{0,1\}^n} \frac{|\mathcal{P}(x, x)|}{2^n} \geq \sum_{i=1}^{c} p_i \ell_i = (1 - \delta) \sum_{i=1}^{c} q_i \ell_i \geq (1 - \delta) \sum_{i=1}^{c} q_i(-\log q_i) \\ &= -(1 - \delta)\, \mathbb{E}[\log q_i] \geq -(1 - \delta) \log \mathbb{E}[q_i] \geq -(1 - \delta)(-\hat{n} + 1) = (1 - \delta)(\hat{n} - 1) \,, \end{aligned}$$

where the second inequality is from the source coding theorem (Fact 29) and the third is from Jensen's inequality. ◀

▶ **Theorem 44.** $\mathrm{R}^{(r),\mathrm{ver}}_{\varepsilon,\delta}(\mathrm{EQ}_n) > \frac{1}{8}(1 - \delta)^2(\hat{n} + \log(1 - \delta) - 5)$.

Proof. Suppose there exists a randomized protocol $\mathcal{P}$ with $\mathrm{rerr}(\mathcal{P}) \leq \varepsilon$, $\mathrm{verr}(\mathcal{P}) \leq \delta$, and $\mathrm{vcost}(\mathcal{P}) \leq m$. For a string s, let $\mathcal{P}_s$ denote the deterministic protocol obtained from $\mathcal{P}$ by fixing the public randomness to s. By the cost and error guarantees of $\mathcal{P}$, for all $(x, y) \in \mathrm{EQ}_n^{-1}(1)$ we have $\mathbb{E}_s\left[\mathrm{cost}(\mathcal{P}_s; x, y)\right] \leq m$ and $\mathbb{E}_s\left[\Pr[\mathrm{out}(\mathcal{P}_s(x, y)) = 0]\right] \leq \delta$, while for $(x, y) \in \mathrm{EQ}^{-1}(0)$ we have $\mathbb{E}_s\left[\Pr[\mathrm{out}(\mathcal{P}_s(x, y)) = 1]\right] \leq \varepsilon$. In particular, letting $(X, Y) \sim \mu$, we have

$$\begin{aligned} \mathbb{E}_{s,X,Y}\left[\Pr[\mathrm{out}(\mathcal{P}_s(X, Y)) = 1 \mid X \neq Y]\right] &\leq \varepsilon \,, \\ \mathbb{E}_{s,X,Y}\left[\Pr[\mathrm{out}(\mathcal{P}_s(X, Y) = 0 \mid X = Y]\right] &\leq \delta \,, \\ \mathbb{E}_{s,X,Y}\left[\mathrm{cost}(\mathcal{P}_s; X, Y) \mid X = Y\right] &\leq m \,. \end{aligned}$$

Define $z = 1 - \delta$, $\hat{\varepsilon} = 4\varepsilon/(1 - \delta)$, $\hat{\delta} = 1 - z/2$, and $\hat{m} = 4m/(1 - \delta)$. Call a string s good if (i) $\mathrm{verr}(\mathcal{P}_s) \leq 1 - z/2$, (ii) $\mathrm{rerr}(\mathcal{P}_s) \leq \hat{\varepsilon}$, and (iii) $\mathrm{vcost}^{\mu}(\mathcal{P}) \leq \hat{m}$. Applying a Markov argument to each condition,

$$\Pr[s \text{ is bad}] < \frac{1 - z}{1 - z/2} + \frac{z}{4} + \frac{z}{4} < 1 \,,$$

where we used $(1 - z)/(1 - z/2) < 1 - z/2$. Thus, there exists a good string s. Note that $\mathcal{P}_s$ is a deterministic $(\hat{\varepsilon}, \hat{\delta})$-error EQ_n protocol. Using Definition 4 to figure the new effective instance size and applying Theorem 43, we obtain

$$\frac{4m}{1 - \delta} \geq \mathrm{vcost}^{\mu}(\mathcal{P}_s) \geq \frac{z}{2}\left(\min\left\{n + \log(z/2),\, \log \frac{z(z/2)^2}{4\varepsilon}\right\} - 1\right) \geq \frac{z}{2}(\hat{n} + \log z - 5) \,.$$

The proof is completed by rearranging the above inequality and substituting $z = 1 - \delta$. ◀

The analysis in the above proof is very loose when δ is bounded away from 1. In particular, when there are no false negatives (i. e., when $\delta = 0$), we are able to show that $\mathrm{R}^{(r),\mathrm{ver}}_{\varepsilon,0} \geq c\hat{n}$ for every constant $c < 1$.

D Main Theorem: Bounded-Round Information Complexity of Equality

In this section we prove Theorem 6, which we think of as the most important result of this paper. We wish to lower bound the bounded-round information complexity of EQUALITY with respect to the uniform distribution. Recall that we are concerned chiefly with protocols that achieve very low refutation error, though they may have rather high verification error. We will prove our lower bound by proving a round elimination lemma for EQ_n that targets *information* cost, and then applying this lemma repeatedly.

This proof has much more technical complexity than our earlier lower bound proofs. Let us see why. There are two main technical difficulties and they arise, ultimately, from the same source: the inability to use (the easy direction of) Yao's minimax lemma. When proving a lower bound on *communication* cost, Yao's lemma allows us to fix the random string used by any purported protocol, which immediately moves us into the clean world of deterministic protocols. This hammer is unavailable to us when working with *information* cost. The most we can do is to "average away" the public randomness. We then have to deal with (private coin) randomized protocols the entire way through the round elimination argument. As a result, our intermediate protocols, obtained by eliminating some rounds of our original protocol, do not obey straightforward cost and error guarantees. This is the first technical difficulty, and our solution to it leads us to the concept of a "kernel" in Definition 45 below.

The second technical difficulty is that we are unable to switch to the simpler case of zero verification error like we did in the proof of Theorem 5, Parts (9) and (10). Therefore, all our intermediate protocols continue to have verification error. Since errors scale up with each round elimination, and the verification error starts out high, we cannot afford even a constant-factor scaling. We must play very delicately with our error parameters, which leads us to the somewhat complicated parametrization seen in Definition 46 below.

D.1 The Round Elimination Argument

▶ **Definition 45** (Kernel)**.** Let p and q be probability distributions on $\{0,1\}^n$, let $S \subseteq \{0,1\}^n$, and let $\ell \geq 0$ be a real number. The triple (p, q, S) is defined to be an *ℓ-kernel* if the following properties hold.

[K1] $\text{H}(p) \geq n - \ell$ and $\text{H}(q) \geq n - \ell$.
[K2] $p(S) \geq 2^{-\ell}$ and $q(S) \geq \frac{1}{2}$.
[K3] For all $x \in S$ we have $q(x) \geq 2^{-n-\ell}$.

▶ **Definition 46** (Parametrized Protocols)**.** Suppose we have an integer $r \geq 1$, and nonnegative reals ℓ, a, b, and c. A protocol $\mathcal{P}$ for EQ_n is defined to be an $[r, \ell, a, b, c]$-protocol if there exists an ℓ-kernel (p, q, S) such that the following properties hold.

[P1] The protocol $\mathcal{P}$ is private-coin and uses r rounds, with Alice speaking in the first round.
[P2] We have $\text{err}^{p\otimes q|S\times S}(\mathcal{P}) = \Pr_{(X,Y)\sim p\otimes q}[\text{out}(\mathcal{P}(X,Y)) \neq \text{EQ}_n(X,Y) \mid (X,Y) \in S \times S] \leq 2^{-a}$.
[P3] We have $\text{verr}^{p\otimes\xi|S\times S}(\mathcal{P}) = \Pr_{X\sim p}[\text{out}(\mathcal{P}(X,X)) = 0 \mid X \in S] \leq 1 - 2^{-b}$.
[P4] We have $\text{icost}^{p\otimes q}(\mathcal{P}) \leq c$.

We alert the reader to the fact that [P2] considers overall error, and not refutation error. We encourage the reader to take a careful look at [P3] and verify the equality claimed therein.

It is straightforward, once one revisits Definition 1 and recalls that ξ denotes the uniform distribution on $\{0,1\}^n$.

Since we have a number of parameters at play, it is worth recording the following simple observation.

▶ **Fact 47.** *Suppose that $\ell' \geq \ell, c' \geq c, a' \leq a$, and $b' \geq b$. Then every ℓ-kernel is also an ℓ'-kernel, and every $[r,\ell,a,b,c]$-protocol is also an $[r,\ell',a',b',c']$-protocol.* ◀

▶ **Theorem 48** (Information-Theoretic Round Elimination for EQUALITY). *If there exists an $[r,\ell,a,b,c]$-protocol with $r \geq 1$ and $c \geq 4$, then there exists an $[r-1,\ell',a',b',c']$-protocol, where*

$$\ell' := (c+\ell)2^{\ell+2b+7}\,, \qquad a' := a-(c+\ell)2^{\ell+2b+8}\,,$$
$$b' := b+2\,, \qquad c' := (c+2)2^{\ell+2b+6}\,.$$

Proof. Let $\mathcal{P}$ be an $[r,\ell,a,b,c]$-protocol, and let (p,q,S) be an ℓ-kernel satisfying the conditions in Definition 46. Assume WLOG that the each message in $\mathcal{P}$ is generated using a fresh random string. Let $X \sim p$ and $Y \sim q$ be independent random variables denoting an input to $\mathcal{P}$. Let $M_1,\ldots,M_r$ be random variables denoting the messages sent in $\mathcal{P}$ on input (X,Y), with M_j being the jth message; note that these variables depend on X, Y, and the random strings used by the players. We then have

$$c \geq \mathrm{icost}^{p\otimes q}(\mathcal{P}) = \mathrm{I}(XY : M_1M_2\ldots M_r) = \mathrm{I}(X : M_1) + \mathrm{I}(XY : M_2\ldots M_r \mid M_1)\,, \tag{15}$$

where the final step uses the chain rule for mutual information, and the fact that M_1 and Y are independent. In particular, we have $\mathrm{I}(X : M_1) \leq c$, and so $\mathrm{H}(X \mid M_1) = \mathrm{H}(X) - \mathrm{I}(X : M_1) \geq n - \ell - c$. By Lemma 19,

$$\mathrm{H}(X \mid M_1, X \in S) \geq n - \frac{\ell+c+1}{p(S)} \geq n - (\ell+c+1)2^{\ell}\,. \tag{16}$$

Let $\mathcal{M}$ be the set of messages that Alice sends with positive probability as her first message in $\mathcal{P}$, given the random input X, i.e., $\mathcal{M} := \{\mathfrak{m} : \Pr[M_1 = \mathfrak{m}] > 0\}$. Consider a particular message $\mathfrak{m} \in \mathcal{M}$. Let $\mathcal{P}'_{\mathfrak{m}}$ denote the following protocol for EQ_n. The players simulate $\mathcal{P}$ on their input, except that Alice is assumed to have sent $\mathfrak{m}$ as her first message. As a result, $\mathcal{P}'_{\mathfrak{m}}$ has $r-1$ rounds and Bob is the player to send the first message in $\mathcal{P}'_{\mathfrak{m}}$. Let $\pi_{\mathfrak{m}}$ and q' be the distributions of $(X \mid M_1 = \mathfrak{m} \wedge X \in S)$ and $(Y \mid Y \in S)$, respectively.

Observe that $\mathrm{icost}^{\pi_{\mathfrak{m}}\otimes q'}(\mathcal{P}'_{\mathfrak{m}}) = \mathrm{I}(XY : M_2\ldots M_r \mid M_1 = \mathfrak{m} \wedge (X,Y) \in S \times S)$. Letting L denote a random first message distributed identically to M_1, we now get

$$\begin{aligned}\mathbb{E}_L\left[\mathrm{icost}^{\pi_L\otimes q'}(\mathcal{P}'_L)\right] &= \mathrm{I}(XY : M_2\ldots M_r \mid M_1, (X,Y) \in S\times S)\\ &\leq \frac{\mathrm{I}(XY : M_2\ldots M_r \mid M_1) + 1}{p(S)q(S)} \leq (c+1)2^{\ell+1}\,,\end{aligned} \tag{17}$$

where the first inequality uses Lemma 18 and the second inequality uses (15) and Property [K2]. Examining Properties [P2] and [P3], we obtain

$$\mathbb{E}_L\left[\mathrm{err}^{\pi_L\otimes q'}(\mathcal{P}'_L)\right] = \mathrm{err}^{p\otimes q|S\times S}(\mathcal{P}) \leq 2^{-a}\,, \tag{18}$$

$$\mathbb{E}_L\left[\mathrm{verr}^{\pi_L\otimes \xi}(\mathcal{P}'_L)\right] = \mathrm{verr}^{p\otimes \xi|S\times S}(\mathcal{P}) \leq 1-2^{-b}\,. \tag{19}$$

▶ **Definition 49** (Good Message). A message $\mathfrak{m} \in \mathcal{M}$ is said to be *good* if the following properties hold:

[G1] $\mathrm{H}(\pi_{\mathfrak{m}}) = \mathrm{H}(X \mid M_1 = \mathfrak{m} \wedge X \in S) \geq n - (\ell + c + 1)2^{\ell+b+3}$,
[G2] $\mathrm{icost}^{\pi_{\mathfrak{m}} \otimes q'}(\mathcal{P}'_{\mathfrak{m}}) \leq 2^{\ell+b+4}(c+1)$,
[G3] $\mathrm{err}^{\pi_{\mathfrak{m}} \otimes q'}(\mathcal{P}'_{\mathfrak{m}}) \leq 2^{-a+b+3}$,
[G4] $\mathrm{verr}^{\pi_{\mathfrak{m}} \otimes \xi}(\mathcal{P}'_{\mathfrak{m}}) \leq 1 - 2^{-b-1}$.

Notice that for all $\mathfrak{m} \in \mathcal{M}$ we have $\mathrm{H}(X \mid M_1 = \mathfrak{m}, X \in S) \leq n$. Hence, viewing (16), (17), (18) and (19) as upper bounds on the expected values of certain nonnegative functions of L, we may apply Markov's inequality to these four conditions and conclude that

$$\Pr[L \text{ is good}] \geq 1 - 2^{-b-3} - 2^{-b-3} - 2^{-b-3} - \frac{1 - 2^{-b}}{1 - 2^{-b-1}} \geq 2^{-b-1} - 3 \cdot 2^{-b-3} > 0 \,.$$

Thus, there exists a good message. *From now on, we fix $\mathfrak{m}$ to be such a good message.*

We may rewrite the left-hand side of [G4] as $\mathbb{E}_{Z \sim \pi_{\mathfrak{m}}}[\Pr[\mathrm{out}(\mathcal{P}'_{\mathfrak{m}}(Z, Z)) = 0]]$. So if we define the set $T := \{x \in S : \Pr[\mathrm{out}(\mathcal{P}'_{\mathfrak{m}}(x, x)) = 0] \leq 1 - 2^{-b-2}\}$ and apply Markov's inequality again, we obtain

$$\pi_{\mathfrak{m}}(T) \geq 1 - \frac{1 - 2^{-b-1}}{1 - 2^{-b-2}} \geq 2^{-b-2} \,. \tag{20}$$

Defining the distribution $p' := \pi_{\mathfrak{m}} \mid T$ and the set $S' := \{x \in T : p'(x) \geq 2^{-n-\ell'}\}$, we now make two claims.

Claim 1: The triple (q', p', S') is an ℓ'-kernel.
Claim 2: We have $\mathrm{err}^{p' \otimes q' \mid S' \times S'}(\mathcal{P}'_{\mathfrak{m}}) \leq 2^{-a'}$, $\mathrm{verr}^{q' \otimes \xi \mid S' \times S'}(\mathcal{P}'_{\mathfrak{m}}) \leq 1 - 2^{-b'}$, and $\mathrm{icost}^{p' \otimes q'}(\mathcal{P}'_{\mathfrak{m}}) \leq c'$.

Notice that these claims essentially say that $\mathcal{P}'_{\mathfrak{m}}$ has all the properties listed in Definition 46, except that Bob starts $\mathcal{P}'_{\mathfrak{m}}$. Interchanging the roles of Alice and Bob in $\mathcal{P}'_{\mathfrak{m}}$ gives us the desired $[r-1, \ell', a', b', c']$-protocol, which completes the proof of the theorem.

It remains to prove the above claims. We start with Claim 1. Starting with the lower bound on $\mathrm{H}(\pi_{\mathfrak{m}})$ given by Property [G1] of the good message $\mathfrak{m}$, and using Lemma 19 followed by (20), we obtain

$$\mathrm{H}(p') = \mathrm{H}(\pi_{\mathfrak{m}} \mid T) \geq n - \frac{(c + \ell + 1)2^{\ell+b+3} + 1}{\pi_{\mathfrak{m}}(T)} \geq n - (c + \ell + 2)2^{\ell+2b+5} \geq n - \ell' \,. \tag{21}$$

We may lower bound $\mathrm{H}(q')$ using Properties [K1] and [K2] for (p, q, S) and applying Lemma 19. We have

$$\mathrm{H}(q') = \mathrm{H}(Y \mid Y \in S) \geq n - \frac{\ell + 1}{q(S)} \geq n - 2(\ell + 1) \geq n - \ell' \,.$$

Thus, (q', p', S') satisfies Property [K1] for an ℓ'-kernel. It is immediate that it also satisfies Property [K3]: by definition, for all $x \in S'$, we have $p'(x) \geq 2^{-n-\ell'}$.

It remains to verify Property [K2], which entails showing that $p'(S') \geq \frac{1}{2}$ and that $q'(S') \geq 2^{-\ell'}$. We can lower bound $p'(S')$ as follows:

$$p'(S') = 1 - \sum_{x \in \{0,1\}^n \setminus S'} p'(x) = 1 - \sum_{\substack{x \in \{0,1\}^n \\ p'(x) < 2^{-n-\ell'}}} p'(x) \geq 1 - 2^{-\ell'} \geq \frac{1}{2} \,. \tag{22}$$

To prove the second inequality, we first derive a lower bound on $\mathrm{H}(p' \mid S')$, thence on $|S'|$, and finally on $q'(S')$. We already showed that $\mathrm{H}(p') \geq n - (c + \ell + 2)2^{\ell+2b+5}$, at (21). By Lemma 19 and (22), we get

$$\mathrm{H}(p' \mid S') \geq n - \frac{(c + \ell + 2)2^{\ell+2b+5} + 1}{p'(S')} \geq n - \big((c + \ell + 2)2^{\ell+2b+6} + 2\big) \geq n - (c + \ell + 4)2^{\ell+2b+6} \,,$$

and so $|S'| \geq 2^{n-(c+\ell+4)2^{\ell+2b+6}}$. Since $q' = q \mid S$ and $S' \subseteq S$, we have

$$q'(S') \geq q(S') \geq |S'| \min_{y \in S'} q(y) \geq |S'| \min_{y \in S} q(y) \geq 2^{n-(c+\ell+4)2^{\ell+2b+6}} 2^{-n-\ell} = 2^{-\ell-(c+\ell+4)2^{\ell+2b+6}} ,$$

where the final inequality uses Property [K3]. Recalling the definition of ℓ' and applying a crude estimate (using the bound $c \geq 4$), we get $q'(S') \geq 2^{-\ell'}$. This finishes the proof of Claim 1.

We now prove Claim 2. Of the three bounds we need to prove, the verification error bound is the easiest. Recalling how T was defined, and noting that $S' \subseteq T$, we immediately obtain

$$\operatorname{verr}^{q' \otimes \xi | S' \times S'}(\mathcal{P}'_{\mathfrak{m}}) = \mathbb{E}_{Y' \sim q'}[\Pr[\operatorname{out}(\mathcal{P}'_{\mathfrak{m}}(Y', Y')) = 0 \mid Y' \in S']] \leq 1 - 2^{-b-2} .$$

To establish the overall error bound, we use

$$\operatorname{err}^{p' \otimes q' | S' \times S'}(\mathcal{P}'_{\mathfrak{m}}) \leq \frac{\operatorname{err}^{p' \otimes q'}(\mathcal{P}'_{\mathfrak{m}})}{p'(S')q'(S')} \leq \frac{\operatorname{err}^{\pi_{\mathfrak{m}} \otimes q'}(\mathcal{P}'_{\mathfrak{m}})}{\pi_{\mathfrak{m}}(T)p'(S')q'(S')} \leq \frac{2^{-a+b+3}}{2^{-b-2} \cdot \frac{1}{2} \cdot 2^{-\ell'}} \tag{23}$$

$$= 2^{-a+2b+6+(c+\ell)2^{\ell+2b+7}} \leq 2^{-a+(c+\ell)2^{\ell+2b+8}} , \tag{24}$$

where the final inequality in (23) follows from Property [K2] for an ℓ'-kernel and Property [G3], and (24) just uses a crude estimate (this time $c \geq 1$ suffices). The last thing remaining is to establish the information cost bound in Claim 2. We do this as follows.

$$\begin{aligned}
\operatorname{icost}^{p' \otimes q'}(\mathcal{P}'_{\mathfrak{m}}) &= \mathrm{I}(XY : M_2 \ldots M_r \mid M_1 = \mathfrak{m} \wedge X \in T \wedge Y \in S) \\
&\leq \frac{\mathrm{I}(XY : M_2 \ldots M_r \mid M_1 = \mathfrak{m} \wedge (X,Y) \in S \times S) + 1}{\Pr[X \in T \mid M_1 = \mathfrak{m} \wedge (X,Y) \in S \times S]} && (25) \\
&= \frac{\operatorname{icost}^{\pi_{\mathfrak{m}} \otimes q'}(\mathcal{P}'_{\mathfrak{m}}) + 1}{\pi_{\mathfrak{m}}(T)} && (26) \\
&\leq \frac{2^{b+\ell+4}(c+1) + 1}{2^{-b-2}} \leq (c+2)2^{\ell+2b+6} , && (27)
\end{aligned}$$

where (25) uses Lemma 18, (26) uses the independence of X and Y and (27) uses Property [G2] and Eq. (20).

This completes the proof of Claim 2 and, with it, the proof of the theorem. ◀

The following easy corollary of Theorem 48 will be useful shortly.

▶ **Corollary 50.** *Let $\tilde{n}, j, r \in \mathbb{N}$ and $a, b \in \mathbb{R}$ with $\tilde{n}$ sufficiently large, $j \geq 1$, $r \geq 1$, and $b \geq 0$. Suppose there exists an $[r, \ell, a - \ell, b, \ell]$-protocol, with $b \leq \ell = \frac{1}{8} \operatorname{ilog}^j \tilde{n}$. Then there exists an $[r-1, \ell', a - \ell', b+2, \ell']$-protocol with $b + 2 \leq \ell' = (\operatorname{ilog}^{j-1} \tilde{n})^{1/2} \leq \frac{1}{8} \operatorname{ilog}^{j-1} \tilde{n}$.*

Proof. This simply boils down to the following estimation, which is valid for all sufficiently large $\tilde{n}$:

$$(\ell + \ell)2^{\ell+2b+8} = 2^7(\operatorname{ilog}^j \tilde{n})2^{(3/8)\operatorname{ilog}^j \tilde{n}} = 2^7(\operatorname{ilog}^{j-1} \tilde{n})^{3/8} \log(\operatorname{ilog}^{j-1} \tilde{n}) \leq (\operatorname{ilog}^{j-1} \tilde{n})^{1/2} . ◀$$

D.2 Finishing the Proof

We are now ready to state and prove the main lower bound on protocols with two-sided error.

▶ **Theorem 51** (Restatement of Main Theorem). *Let $\tilde{n} = \min\{n + \log(1-\delta), \log((1-\delta)/\varepsilon)\}$. Suppose $\delta \leq 1 - 8(\operatorname{ilog}^{r-2} \tilde{n})^{-1/8}$. Then we have $\mathrm{IC}^{\mu,(r)}_{\varepsilon,\delta}(\mathrm{EQ}_n) = \Omega((1-\delta)^3 \operatorname{ilog}^{r-1} \tilde{n})$.*

Proof. We may assume that $r \le \log^* \tilde{n}$, for otherwise there is nothing to prove. The slight difference between $\tilde{n}$ above and $\hat{n}$, as in Definition 4, is insignificant and can be absorbed by the $\Omega(\cdot)$ notation.

Suppose, to the contrary, that there exists an r-round randomized protocol $\mathcal{P}^*$ for $\textsc{eq}_n$, with $\operatorname{rerr}^\mu(\mathcal{P}^*) \le \varepsilon$, $\operatorname{verr}^\mu(\mathcal{P}^*) \le \delta$ and $\operatorname{icost}^\mu(\mathcal{P}^*) \le 2^{-16}(1-\delta)^3 \operatorname{ilog}^{r-1} \tilde{n}$. Recall that we denote the uniform distribution on $\{0,1\}^n$ by ξ and that $\mu = \xi \otimes \xi$. We have

$$\operatorname{err}^\mu(\mathcal{P}^*) = (1-2^{-n}) \operatorname{rerr}^\mu(\mathcal{P}^*) + 2^{-n} \operatorname{verr}^\mu(\mathcal{P}^*) \le \varepsilon + 2^{-n}(\delta - \varepsilon) \le \varepsilon + 2^{-n} \,.$$

Let $\mathcal{P}_s^*$ be the private-coin protocol for $\textsc{eq}_n$ obtained from $\mathcal{P}^*$ by fixing the public random string of $\mathcal{P}^*$ to be s. We have $\mathbb{E}_s[\operatorname{err}^\mu(\mathcal{P}_s^*)] \le \varepsilon + 2^{-n}$, $\mathbb{E}_s[\operatorname{verr}^\mu(\mathcal{P}_s^*)] \le \delta$, and $\mathbb{E}_s[\operatorname{icost}(\mathcal{P}_s^*)] \le 2^{-16}(1-\delta)^3 \operatorname{ilog}^{r-1} \tilde{n}$. By Markov's inequality, there exists s such that $\mathcal{P}_s^*$ simultaneously has $\operatorname{err}^\mu(\mathcal{P}_s^*) \le 4(\varepsilon + 2^{-n})/(1-\delta)$, $\operatorname{verr}^\mu(\mathcal{P}_s^*) \le (1+\delta)/2$, and $\operatorname{icost}(\mathcal{P}_s^*) \le 2^{-14}(1-\delta)^2 \operatorname{ilog}^{r-1} \tilde{n}$: this is because

$$1 - \frac{1-\delta}{4} - \frac{2\delta}{1+\delta} - \frac{1-\delta}{4} = \frac{(1-\delta)^2}{2(1+\delta)} > 0 \,.$$

Let $\mathcal{P} = \mathcal{P}_s^*$ for this s. Then $(\xi, \xi, \{0,1\}^n)$ is a 0-kernel and $\mathcal{P}$ is an $[r,\, 0,\, \log \frac{1-\delta}{4(\varepsilon + 2^{-n})},\, \log \frac{2}{1-\delta},\, 2^{-14}(1-\delta)^2 \operatorname{ilog}^{r-1} \tilde{n}]$-protocol. Recalling Fact 47 and using $\log \frac{1-\delta}{\varepsilon + 2^{-n}} \ge \tilde{n} - 1$, we see that

$$\mathcal{P} \text{ is an } \left[r,\, 0,\, \tilde{n} - 3,\, \log \tfrac{1}{1-\delta} + 1,\, 2^{-14}(1-\delta)^2 \operatorname{ilog}^{r-1} \tilde{n}\right]\text{-protocol.}$$

Put $\ell_j := \frac{1}{8} \operatorname{ilog}^j \tilde{n}$ for $j \in \mathbb{N}$. Applying round elimination (Theorem 48) to $\mathcal{P}$ and weakening the resulting parameters (using Fact 47) gives us an $[r-1, \ell_{r-1}, \tilde{n} - \ell_{r-1}, \log \frac{1}{1-\delta} + 3, \ell_{r-1}]$-protocol $\mathcal{P}'$.

The upper bound on δ gives us $\log \frac{1}{1-\delta} + 3 \le \ell_{r-1}$, and so the conditions for Corollary 50 apply. Starting with $\mathcal{P}'$ and applying that corollary repeatedly, each time using the looser estimate on ℓ' in that corollary, we obtain a sequence of protocols with successively fewer rounds. Eventually we reach a $[1, \ell_1, \tilde{n} - \ell_1, \log \frac{1}{1-\delta} + 2(r-1) + 1, \ell_1]$-protocol. Applying Theorem 48 one more time, and using the tighter estimate on ℓ' this time, we get a $[0,\, \tilde{n}^{1/2},\, \tilde{n} - \tilde{n}^{1/2},\, \log \frac{1}{1-\delta} + 2r + 1,\, \tilde{n}^{1/2}]$-protocol $\mathcal{Q}$. Weakening parameters again, we see that $\mathcal{Q}$ is a $[0,\, \tilde{n}^{1/2},\, \frac{1}{2}\tilde{n},\, \frac{1}{3}\log \tilde{n},\, \tilde{n}^{1/2}]$-protocol. Let (p, q, S) be the $\tilde{n}^{1/2}$-kernel for $\mathcal{Q}$. By Property [K1], we have $\operatorname{H}(q) \ge n - \tilde{n}^{1/2}$. Using Lemma 19 and Property [K2], we then have

$$\operatorname{H}(q \mid S) \ge n - \frac{\tilde{n}^{1/2} + 1}{q(S)} \ge n - (2\tilde{n}^{1/2} + 2) \,. \tag{28}$$

Since $\mathcal{Q}$ involves no communication, it must behave identically on any two input distributions that have the same marginal on Alice's input. In particular, this gives us the following crucial equation:

$$\Pr_{X \sim p}[\operatorname{out}(\mathcal{Q}(X,X)) = 1 \mid X \in S] = \Pr_{(X,Y) \sim p \otimes q}[\operatorname{out}(\mathcal{Q}(X,Y)) = 1 \mid (X,Y) \in S \times S] \,. \tag{29}$$

Let α denote the above probability. Considering the left-hand side of (29), we have

$$\alpha = 1 - \operatorname{verr}^{p \otimes \xi | S \times S}(\mathcal{Q}) \ge 2^{-\frac{1}{3}\log \tilde{n}} = \tilde{n}^{-1/3} \,. \tag{30}$$

On the other hand, whenever $\mathcal{Q}$ outputs 1 on an input (x, y), then either $x = y$ or $\mathcal{Q}$ errs on

(x, y). Therefore, considering the right-hand side of (29), we have

$$\begin{aligned}\alpha &\le \Pr_{(X,Y)\sim p\otimes q}[X = Y \mid (X,Y) \in S \times S] + \\ &\qquad\qquad \Pr_{(X,Y)\sim p\otimes q}[\mathrm{out}(\mathcal{P}(X,Y)) \ne \textsc{eq}_n(X,Y) \mid (X,Y) \in S \times S] &(31)\\ &\le \max_{x\in S} \Pr_{Y\sim q|S}[Y = x] + \mathrm{err}^{p\otimes q|S\times S}(\mathcal{Q}) \\ &\le \frac{2\tilde{n}^{1/2}+3}{n} + 2^{-\frac{1}{2}\tilde{n}} &(32)\\ &\le 2\tilde{n}^{-1/2} + 3\tilde{n}^{-1} + 2^{-\frac{1}{2}\tilde{n}}\,, &(33)\end{aligned}$$

where (32) follows from (28) by applying Lemma 20, and (33) uses $\tilde{n} \le n$.

The bounds (30) and (33) are in contradiction for sufficiently large $\tilde{n}$, which completes the proof. ◀

E Applications, Including Bounded-Round Small-Set Disjointness

E.1 Lower Bounds

In this section we apply our new understanding of the bounded-round information complexity of EQUALITY to obtain two new lower bounds: one for OR-EQUALITY, and the other for the much-studied DISJOINTNESS problem with small-sized sets. As we shall see, both lower bounds are arguably tight.

▶ **Theorem 52** (Lower Bound for Or-Equality). *Let $k, n, r \in \mathbb{N}$ and $\delta, \varepsilon \in [0, 1]$. Put $\varepsilon' = \varepsilon + k/2^n$ and $\tilde{n} = \log\frac{1-\delta}{\varepsilon'}$. For $\delta < 1 - 8(\mathrm{ilog}^{r-2}\,\tilde{n})^{-1/8}$, we have*

$$\mathrm{R}^{(r)}_{\varepsilon,\delta}(\textsc{oreq}_{n,k}) \ge k \cdot \mathrm{IC}^{\mu,(r)}_{\varepsilon',\delta}(\textsc{eq}_n) = \Omega(k(1-\delta)^3\,\mathrm{ilog}^{r-1}\,\tilde{n})\,.$$

Proof. We just need to show the first inequality and then apply Theorem 6. That inequality is proved via standard direct sum arguments for information complexity [15, 4, 5]. In fact, the old simultaneous-message lower bound for $\textsc{oreq}_{n,k}$ from Chakrabarti et al. [15] applies more-or-less unchanged. For completeness, we now give a self-contained proof.

Let $\mathcal{P}$ be an r-round protocol for $\textsc{oreq}_{n,k}$ with $\mathrm{rerr}(\mathcal{P}) \le \varepsilon$, $\mathrm{verr}(\mathcal{P}) \le \delta$, and $\mathrm{R}^{(r)}_{\varepsilon,\delta}(\textsc{oreq}_{n,k}) \ge \max\{\mathrm{rcost}(\mathcal{P}), \mathrm{vcost}(\mathcal{P})\}$. Alice and Bob solve $\textsc{eq}_n$ by the following protocol $\mathcal{Q}_j$, where j is some fixed index in $\{1, 2, \dots, k\}$. Given an input $(x, y) \in \{0,1\}^n \times \{0,1\}^n$, they generate $\mathbf{X} := (X_1, \dots, X_k) \sim \xi^{\otimes k}$ and $\mathbf{Y} := (Y_1, \dots, Y_k) \sim \xi^{\otimes k}$ respectively, using private coins. They "plug in" x and y into the jth coordinates of $\mathbf{X}$ and $\mathbf{Y}$ respectively, thereby creating

$$\mathbf{Z}_{j,x} := (X_1, \dots, X_{j-1}, x, X_{j+1}, \dots, X_k) \text{ and } \mathbf{W}_{j,y} := (Y_1, \dots, Y_{j-1}, y, Y_{j+1}, \dots, Y_k)\,,$$

respectively. Finally, they emulate $\mathcal{P}$ on input $(\mathbf{Z}_{j,x}, \mathbf{W}_{j,y})$. Observe that

$$\textsc{oreq}_{n,k}(\mathbf{Z}_{j,x}, \mathbf{W}_{j,y}) \ne \textsc{eq}_n(x, y) \implies (x \ne y) \wedge \big(\exists\, i \in [k] \setminus \{j\} : X_i = Y_i\big)\,.$$

Therefore, $\mathrm{verr}(\mathcal{Q}_j) \le \mathrm{verr}(\mathcal{P}) \le \delta$ and, by a union bound,

$$\mathrm{rerr}(\mathcal{Q}_j) \le \mathrm{rerr}(\mathcal{P}) + \sum_{i=1}^{n} \Pr[X_i = Y_i] \le \varepsilon + k/2^n = \varepsilon'\,.$$

Since $\mathcal{Q}_j$ solves $\textsc{eq}_n$ with these error guarantees, it follows that $\mathrm{icost}^{\mu}(\mathcal{Q}_j) \ge \mathrm{IC}^{\mu,(r)}_{\varepsilon',\delta}(\textsc{eq}_n)$.

Now, let $(X,Y) \sim \mu$ and let $\mathfrak{R}$ denote the public randomness used by $\mathcal{P}$. We can now lower bound $\mathrm{R}^{(r)}_{\varepsilon,\delta}(\textsc{oreq}_{n,k})$ as follows:

$$\begin{aligned}
\mathrm{R}^{(r)}_{\varepsilon,\delta}(\textsc{oreq}_{n,k}) &\geq \max\nolimits_{x_1,\ldots,x_k,y_1,\ldots,y_k\in\{0,1\}^{kn}\times\{0,1\}^{kn}} \operatorname{cost}(\mathcal{P};x_1,\ldots,x_k,y_1,\ldots,y_k)\\
&\geq \mathbb{E}[\operatorname{cost}(\mathcal{P};X_1,\ldots,X_k,Y_1,\ldots,Y_k)]\\
&\geq \mathrm{H}(\mathcal{P}(X_1,\ldots,X_k,Y_1,\ldots,Y_k)) &(34)\\
&\geq \mathrm{I}(\mathcal{P}(X_1,\ldots,X_k,Y_1,\ldots,Y_k) : X_1Y_1\ldots X_kY_k \mid \mathfrak{R})\\
&\geq \sum_{j=1}^{k} \mathrm{I}(\mathcal{P}(X_1,\ldots,X_k,Y_1,\ldots,Y_k) : X_i,Y_i \mid \mathfrak{R}) &(35)\\
&= \sum_{j=1}^{k} \mathrm{I}(\mathcal{Q}_j(X,Y) : XY \mid \mathfrak{R}) &(36)\\
&= \sum_{j=1}^{k} \operatorname{icost}^{\mu}(\mathcal{Q}_j) \geq k\cdot \mathrm{IC}^{\mu,(r)}_{\varepsilon',\delta}(\textsc{eq}_n)\,,
\end{aligned}$$

where (34) uses Fact 29 and (35) uses the independence of $\{X_1Y_1,\ldots,X_kY_k\}$ and the resulting subadditivity of mutual information, and (36) holds because, for all $j\in[k]$, the distributions of $(\mathcal{Q}_j(X,Y),X,Y,\mathfrak{R})$ and $(\mathcal{P}(X_1,\ldots,X_k,Y_1,\ldots,Y_k),X_j,Y_j,\mathfrak{R})$ are identical. This completes the proof. ◀

By plugging in $\varepsilon=0$, $\delta=0$ in Theorem 52 we obtain the following corollary.

▶ **Corollary 53.** $\mathrm{R}^{(r)}_{0,0}(\textsc{oreq}_{n,k}) = \Omega(k \operatorname{ilog}^{r-1}(n-\log k))$. ◀

Armed with the above lower bound, we now derive a lower bound for k-DISJ via a simple reduction, which is probably folklore. For completeness, we again give a formal proof. Note that the reduction interchanges verification and refutation errors.

▶ **Lemma 54** (Reduction from OREQ to k-DISJ). *Let k,N be integers such that $N\geq k^c$ for some constant $c>2$. Let $n=\left\lfloor \log\left(\frac{N}{k}\right)\right\rfloor$. If there exists a protocol $\mathcal{P}$ for k-DISJ$_N$ then there exists a protocol $\mathcal{Q}$ for OREQ$_{n,k}$ such that* $\operatorname{rerr}(\mathcal{Q})\leq\operatorname{verr}(\mathcal{P})$ *and* $\operatorname{verr}(\mathcal{Q})\leq\operatorname{rerr}(\mathcal{P})$ *and* $\operatorname{vcost}(\mathcal{Q})\leq\operatorname{rcost}(\mathcal{P})$ *and* $\operatorname{rcost}(\mathcal{Q})\leq\operatorname{vcost}(\mathcal{P})$.

Proof. Given an input instance $(x_1,\ldots,x_k,y_1,\ldots,y_k)$ of OREQ$_{n,k}$, we can transform it into an instance (A,B) of k-DISJ$_N$ as follows:

$$\begin{aligned}
A &= \{x_1,\, x_2+2^n,\, x_3+2\cdot 2^n,\, \ldots,\, x_k+(k-1)2^n\}\\
B &= \{y_1,\, y_2+2^n,\, y_3+2\cdot 2^n,\, \ldots,\, y_k+(k-1)2^n\}\,.
\end{aligned}$$

It is easy to observe that $A\cap B\neq\emptyset$ iff $\exists\, i\in[k]$ such that $x_i=y_i$ because $x_i\in\{0,1,\ldots,2^n-1\}$. Therefore, $\textsc{oreq}_{n,k}(x_1,\ldots,x_k,y_1,\ldots,y_k)=\neg\, k\text{-}\textsc{disj}_N(A,B)$, which completes the proof. ◀

▶ **Corollary 55.** *We have* $\mathrm{R}^{(r)}_{\delta,\varepsilon}(k\text{-}\textsc{disj}_N)\geq\mathrm{R}^{(r)}_{\varepsilon,\delta}(\textsc{oreq}_{\lfloor\log(N/k)\rfloor,k})$. ◀

Combining Corollary 55 with Theorem 52, we arrive at the following theorem.

▶ **Theorem 56** (Lower Bound for k-Disjointness). *Let $k,N,r\in\mathbb{N}$, $\varepsilon,\delta\in[0,1]$ and $c>2$ be such that $N\geq k^c$ and $\delta<1-8(\operatorname{ilog}^{r-2}\tilde{n})^{-1/8}$, where $\tilde{n}=\log\frac{1-\delta}{\varepsilon+k^2/N}$. Then*

$$\mathrm{R}^{(r)}_{\delta,\varepsilon}(k\text{-}\textsc{disj}_N)=\Omega(k(1-\delta)^3\operatorname{ilog}^{r-1}\tilde{n})\,.$$

In particular, with $\delta=1-\Omega(1)$ and $\varepsilon\leq k^{-\Theta(1)}$, we have $\mathrm{R}^{(r)}_{\delta,\varepsilon}(k\text{-}\textsc{disj}_N)=\Omega(k\operatorname{ilog}^r k)$. ◀

By plugging in $\varepsilon = \delta = 0$ above we arrive at a further special case that is worth highlighting.

▶ **Corollary 57.** *With $N \geq k^{2+\Omega(1)}$, we have $\mathrm{R}^{(r)}_{0,0}(k\text{-DISJ}_N) = \Omega(k \operatorname{ilog}^r k)$.* ◀

E.2 Tightness

Our lower bounds in Section E.1 have the weakness that they apply only in zero-error or small-error settings. However, they are still tight in the following sense. We can design protocols that give matching *upper* bounds under similarly small error settings. For OREQ, we give such a protocol below. For k-DISJ, a suitable analysis of a recent protocol of Sağlam and Tardos[42] gives similar results.

▶ **Theorem 58.** *For all $r < \log^* k$, there exists a r-round protocol $\mathcal{P}$ for $\text{OREQ}_{n,k}$ with worst-case communication cost $O(k \operatorname{ilog}^r k)$, $\operatorname{rerr}(\mathcal{P}) < 2^{-\prod_{j=1}^{r} \operatorname{ilog}^j k}$, and $\operatorname{verr}(\mathcal{P}) = 0$.*

Proof. For ease of presentation, we give the details for a slightly weaker result, with refutation error $< k^{-10}$.

We begin with a high-level sketch of the proof, before giving formal proof details. Alice begins the protocol by sending, in parallel, k different t-bit equality tests, one for each of her inputs. Note that for any i where $x_i \neq y_i$, Bob witnesses non-equality with probability $1 - 2^{-t}$. Assuming $\text{OREQ}_{n,k}(x, y) = 0$, there will be roughly $k/2^t$ coordinates i where $x_i \neq y_i$ has not yet been witnessed. Bob now tells Alice which of his coordinates remain "alive" and sends t'-bit equality tests for each of *these* coordinates, where $t' = 2^t$. Note that Bob's overall communication is roughly k bits, and that after receiving this message, Alice witnesses non-equality on all but a $2^{-t'}$-fraction of unequal pairs. In each round, players end up sending an exponentially longer equality test on an exponentially smaller number of coordinates. When communication ends, players output $\text{OREQ}(x_1, \ldots, x_k, y_1, \ldots, y_k) = 1$ unless $x_i \neq y_i$ has been witnessed for all i. One potential issue with the above protocol is that too many coordinates could remain, and players wouldn't be able to communicate exponentially more bits about the remaining coordinates. This could happen both when an unusually large number of equality tests fail, or just for the simple reason that $x_i = y_i$ for many coordinates. In either case, the players simply abort and output $\text{OREQ}_{n,k} = 1$. This will cause an increase in error, but the increase will be small, and it will only increase the false positive rate. A formal proof lies below.

The protocol proceeds in a number of rounds. Throughout, players maintain a vector $w \in \{0, 1\}^k$ (initialized to $w = 1^k$), where $w_i = 0$ iff $x_i \neq y_i$ has been witnessed. Coordinate i is deemed "live" if $w_i = 1$.

In the first round of communication, Alice sends a $(2 \operatorname{ilog}^r k)$-bit equality test for each of the k live coordinates, at a total cost of $O(k \operatorname{ilog}^r k)$ bits.

In the jth round of communication $(1 < j < r)$, the player to speak first updates her copy of w by considering the $(j - 1)$th message: for for each live i, she sets $w_i = 0$ if $x_i \neq y_i$ is witnessed. Now, if more than $2k/\operatorname{ilog}^{r+1-j} k$ coordinates remain live, she sends "1", signifying that the protocol should abort and output $\text{OREQ}_{n,k} = 1$. Otherwise, she sends "0", followed by her updated copy of w, followed by a $(2 \operatorname{ilog}^{r+1-j} k)$-bit equality test for each live coordinate. Thus the jth message is $O(k)$ bits long.

The final round of communication is similar, except that the equality tests are $(12 \log k)$-bits long rather than $2 \operatorname{ilog}^{r+1-r} k = 2 \log k$ bits. The receiver of the final message updates his copy of w, evaluates each equality test, and outputs $\text{OREQ}_{n,k} = 1$ if any coordinates remain live. Otherwise, he outputs $\text{OREQ}_{n,k} = 0$.

The overall communication is thus $O(k \operatorname{ilog}^r k)$ bits. Note also that the protocol outputs $\textsc{oreq}_{n,k} = 0$ only when $x_i \neq y_i$ was witnessed for every i. Thus, the protocol produces no false negatives.

A false positive can happen for one of two reasons: either the protocol aborts (outputting $\textsc{oreq}_{n,k} = 1$), or one or more coordinates remain live at the end of the protocol, despite having $x_i \neq y_i$ for all i.

In the former case, note that (conditioned on not aborting before round j) we have at most $2k/\operatorname{ilog}^{r+1-j} k$ live coordinates during round j. Players execute a $(2 \operatorname{ilog}^{r+1-j} k)$-bit equality test during this round. Thus, a coordinate remains live after this test with probability at most $2^{-2 \operatorname{ilog}^{r+1-j} k} < 1/\operatorname{ilog}^{r-j} k$. Therefore, we expect at most $k/\operatorname{ilog}^{r-j} k$ coordinates to be live in the next round. By a (crude) Chernoff bound argument, the probability of aborting during round $j+1$ (again, conditioned on not previously aborting) is less than k^{-20}, and the overall probability of aborting before the end of the protocol is less than k^{-12} (say).

In the latter case, note that the final equality test uses $12 \log k$ bits per coordinate. Therefore, players *fail* to witness $x_i \neq y_i$ with probability at most $2^{-12 \log k} = k^{-12}$. By a union bound, the overall false positive rate is at most k^{-10}. ◀

F Direct Sum for Equality with Constant Error

In this section we prove our results for PRIVATE-INTERSECTION. In the proof we will use the following modification of the strong direct sum theorem of [37] (Theorem 2.1), which uses protocols with abortion (see definitions in Appendix A.2). The simulation procedure used in the proof of this theorem in [37] preserves the number of rounds in the protocol, which allows us to state their theorem as:

▶ **Theorem 59** (Strong Direct Sum [37]). *Let $\delta \leq 1/3$. Then for every function $f\colon \mathcal{X} \times \mathcal{Y} \to \mathcal{Z}$ and distribution λ on $\mathcal{X} \times \mathcal{Y} \times \mathrm{D}$ with marginal μ_p on $\mathcal{X} \times \mathcal{Y}$ and marginal ν_p on D, such that μ_p is partitioned by ν_p, it holds that $\mathrm{IC}_{\delta}^{\mu_p^k,(r)}(f^k|\nu_p^k) \geq \Omega(k)\, \mathrm{IC}_{\frac{1}{20},\frac{1}{10},\frac{\delta}{k}}^{\mu_p,(r)}(f|\nu_p)$.*

Using the direct sum above it remains to show the following:

▶ **Lemma 60.** *There exists a distribution on $\mathcal{X} \times \mathcal{Y} \times \mathcal{D}$ with marginals μ_p on $\mathcal{X} \times \mathcal{Y}$ and ν_p on $\mathcal{D}$, such that ν_p partitions μ_p and $\mathrm{IC}_{1/20,1/10,\delta/k}^{\mu_p,(r)}(\mathrm{EQ}_{n/k}\,|\nu_p) = \Omega(\operatorname{ilog}^r k)$.*

Proof. In the proof we can use the same hard distribution as in [37]. Let $\ell = n/k$. To construct μ_p and ν_p, let D_0 be a random variable uniformly distributed on $\{0,1\}$ and let $\mathbf{D}$ be a random variable uniformly distributed on $\{0,1\}^\ell$. Let $(\mathbf{X}, \mathbf{Y})$ be a random variable supported on $\{0,1\}^\ell \times \{0,1\}^\ell$ such that, conditioned on $D_0 = 0$ we have $\mathbf{X}$ and $\mathbf{Y}$ distributed independently and uniformly on $\{0,1\}^\ell$, and conditioned on $D_0 = 1$ we have $\mathbf{X} = \mathbf{Y} = \mathbf{D}$. Let μ_p be the distribution of $(\mathbf{X}, \mathbf{Y})$ and let ν_p be the distribution of $(D_0\mathbf{D})$. Note that ν_p partitions μ_p. Also, this distribution satisfies that $\Pr[\mathbf{X} = \mathbf{Y}] \geq 1/3$ and $\Pr[\mathbf{X} \neq \mathbf{Y}] \geq 1/3$.

Let W be a random variable distributed according to ν_p. Let E be an indicator variable over the private randomness of $\mathcal{P}$ which is equal to 1 if and only if conditioned on this private randomness $\mathcal{P}$ satisfies that it aborts with probability at most $1/10$ and succeeds with probability at least $1 - \delta/k$ conditioned on non-abortion. Given such protocol with abortion $\mathcal{P}$ we transform it into a protocol $\mathcal{P}'$ which never aborts, has almost the same information complexity and gives correct output on non-equal instances with high probability, while being correct on equal instances with constant probability. This is done by constructing $\mathcal{P}'$ so that whenever $\mathcal{P}$ outputs "abort", the output of $\mathcal{P}'$ is $X \neq Y$, otherwise $\mathcal{P} = \mathcal{P}'$. Under the distribution μ_p conditioned on the event $E = 1$ the protocol $\mathcal{P}'$ has the property that if

$X \neq Y$, then it outputs $X = Y$ with probability at most $(1/k)/\Pr_{\mu_p}[X \neq Y] \leq 3/k$. However, if $X = Y$, then the protocol may output $X \neq Y$ with probability $1/10 + (1/k)/\Pr_{\mu'_p}[X = Y] \leq 1/10 + 3/k \leq 1/5$, where the latter follows for $k \geq 30$. Thus, conditioned on $E = 1$, the protocol $\mathcal{P}'$ has failure probability $\epsilon = 1/k$ on non-equal instances $X \neq Y$, and constant failure probability $\delta = 1/5$ on equal instances $X = Y$, as desired. In this regime we can use Theorem 6. We have:

$$\begin{aligned}\mathrm{IC}^{\mu_p,(r)}_{1/20,1/10,\delta/k}(\mathrm{EQ}_{n/k}\,|\nu_p) &\geq \mathrm{I}(\mathcal{P} : X, Y | W)\\ &= \Omega(\mathrm{I}(\mathcal{P} : X, Y | W, E = 1)) - 1\\ &= \Omega(\mathrm{I}(\mathcal{P}' : X, Y | W, E = 1)) - 2.\end{aligned}$$

Here the inequality is by definition of information compelxity and the equalities follows from Proposition 17 together with the fact that $H(E) \leq 1$, $\Pr[E = 1] = 19/20$, and the fact that the transcripts of the protocols $\mathcal{P}$ and $\mathcal{P}'$ only differ in a single bit. The right-hand side can be bounded as follows.

$$\mathrm{I}(\mathcal{P}' : X, Y | W, E = 1)) = \Omega(\mathrm{IC}^{\mu,(r)}_{1/k,1/5}(\mathrm{EQ}_{n/k})). \tag{37}$$

This follows from the construction of the distributions μ_p and ν_p that we use. If $D_0 = 0$ then $\mathbf{X} = \mathbf{Y}$ and the information revealed by $\mathcal{P}$ is equal to zero. Otherwise, if $D_0 = 1$ then the distribution of $(\mathbf{X}, \mathbf{Y})$ is uniform. Because the latter happens with probability $1/2$ we have $\mathrm{I}(\mathcal{P}' : X, Y | W, E = 1)) \geq 1/2 \cdot \mathrm{IC}^{\mu,(r)}_{1/k,1/5}(\mathrm{EQ}_{n/k}))$ as desired.

Using (37) we have $\mathrm{IC}^{\mu_p,(r)}_{1/20,1/10,\delta/k}(\mathrm{EQ}_{n/k}\,|\nu_p) = \Omega(\mathrm{IC}^{\mu,(r)}_{1/k,1/5}(\mathrm{EQ}_{n/k}))$. The proof is completed by noting that setting $\epsilon = 1/k$ and $\delta = 1/5$ in Theorem 6 gives $\mathrm{IC}^{\mu,(r)}_{1/k,1,5}(\mathrm{EQ}_{n/k}) = \Omega(\mathrm{ilog}^r k)$. ◀