Communication Complexity with Defective Randomness

Marshall Ball* Oded Goldreich† Tal Malkin‡

June 21, 2021

Abstract

Starting with the two standard model of randomized communication complexity, we study the communication complexity of functions when the protocol has access to a defective source of randomness. Specifically, we consider both the public-randomness and private-randomness cases, while replacing the commonly postulated perfect randomness with distributions over ℓ bit strings that have min-entropy at least $k \leq \ell$. We present general upper and lower bounds on the communication complexity in these cases, where the bounds are typically linear in $\ell - k$ and also depend on the size of the fooling set for the function being computed and on its standard randomized complexity.

Contents

Keywords: Randomized Communication Complexity, Randomness Extraction, Min-Entropy

*Computer Science Department, Columbia University, New York. E-mail: marshall@cs.columbia.edu

†Faculty of Mathematics and Computer Science, Weizmann Institute of Science, Rehovot, Israel. Email: oded.goldreich@weizmann.ac.il

‡Computer Science Department, Columbia University, New York. E-mail: tal@cs.columbia.edu

1 Introduction

While communication complexity is typically viewed as a tool for establishing lower bound on other models of computation, one may also view it as a study of (two-party) collaborations that can be carried out using a small amount of communication. The (two) parties participating in such a typical collaboration have a common goal, which is modeled as the computation of a function of their private inputs, and they wish to achieve it efficiently, which means using a small amount of communication (i.e., much smaller than required for communicating their entire input).

Given this perspective, one can ask whether randomness is helpful, and it is well-known that it is extremely helpful. For example, computing the equality function requires deterministic protocols that use a linear amount of communication (i.e., are not significantly better than the straightforward one), but can be performed by randomized protocols that use a constant amount of communication. The question addressed in this work is *what happens when the parties have at their disposal only defective sources of randomness?*

1.1 The Models

Our starting point is the two standard models of randomized communication complexity, which are closely related in the standard setting but may not be so in the current setting. In the standard public randomness model one postulates that the parties have access to a common (i.e., public) source of perfect randomness, whereas in the standard private randomness model one postulates that the each party has access to a private source of perfect randomness (which is uncorrelated to the other party's source). Indeed, in the standard setting, the public randomness model can easily emulate the private randomness model, and the opposite emulation is also possible at a very moderate cost [9].

We consider variants of these two models in which the postulated sources of perfect randomness are replaced by defective sources of randomness. In particular, we consider sources that output ℓ-bit long strings such that no string appears with probability exceeding 2^{-k}; that is, we consider sources of min-entropy k, with a focus on the case that $k \in [\Omega(\log \ell), \ell)$. A special case of interest is when the min-entropy rate (i.e., k/ℓ) is a constant smaller than 1 and the actual inputs are of length related to ℓ; yet, we shall consider the problem in almost full generality.

1.2 Our Results

We show that if the random sources available to the two parties are moderately defective in the sense that their min-entropy rate is a constant smaller than 1, then computing the equality function on strings of length comparable to the length of the random sources requires a linear amount of communication, just as in the case that one uses no randomness at all. More generally, we show that, when using defective sources of randomness, no improvement can be obtained over the lower bound on the communication complexity of deterministic protocol that follows by a "fooling set" argument (see Definition 2.2). The foregoing assertions refers to the case that $\min(m, \ell) = \Omega(n)$ and $k < (1 - \Omega(1)) \cdot \ell$.

Theorem 1.1 (general lower bounds): *Suppose that $f : \{0,1\}^n \times \{0,1\}^n \to \{0,1\}$ has a fooling set of size 2^m, and let $k \leq \ell$.*

- (public randomness version): *If f is computed by a protocol whose only source of randomness is a public random string of length ℓ that has min-entropy k, then the protocol uses at least $\min(m-1, \ell-k-1)/2$ bits of communication.*
- (private randomness version): *If f is computed by a protocol in which the only source of randomness is provided by two independent random strings of length ℓ, each seen by one of the parties and having min-entropy k, then the protocol uses at least $\min(m-1, \ell-k-1)/2$ bits of communication.*

We stress that, in the current context, the two models (i.e., public-randomness and private-randomness) are not easily reducible to one another.[1] Recall that lower-bounding the size of a fooling set is one of the King's Roads for proving lower bounds on the deterministic communication complexity of functions.[2] In particular, equality has a fooling set of size 2^n. In general, the logarithm of the size of the fooling set seems a reasonable proxy for the deterministic communication complexity, but it is indeed interesting to ask whether a result as Theorem 1.1 holds with m replaced by the deterministic communication complexity of f.

While Theorem 1.1 asserts that using a moderately defective random sources of length that is comparable to the input is useless, it does not rule out the benefit of sources that are either less defective or are shorter (i.e., have shorter length). It turns out that it is possible to benefit from the use of such sources.

Theorem 1.2 (generic upper bounds): *For $f : \{0,1\}^n \times \{0,1\}^n \to \{0,1\}$ and $k \geq 2\log_2 n + O(1)$, the following holds.*

- (public randomness version): *Suppose that the randomized communication complexity of f in the standard public-randomness model is C. Then, f can be computed by a protocol whose sole source of randomness is a public random string of length ℓ that has min-entropy k using $O(\ell - k + 1) \cdot C$ bits of communication.*

- (private randomness version): *Suppose that the randomized communication complexity of f in the standard private-randomness model is C. Then, f can be computed by a protocol whose sole source of randomness is provided by two independent random strings of length ℓ, each seen by one of the parties and having min-entropy k, using $2(\ell - k) + 3\log_2 n + O(C)$ bits of communication.*

The protocols use suitable methods of randomness extraction. Specifically, in the public-randomness case the two parties apply a seeded extractor to the only random string available to them, while using all possible seeds (of length $\log_2(\ell - k) + O(1)$). In the private-randomness case the parties apply a two-source extractor to the $2 \cdot (\ell - k + \log_2 n + O(1))$-bit long prefix of their sources, which requires them to only communicate this prefix.

Recall that in the case of perfect randomness, a common random source (i.e., public randomness) is preferable to private randomness, since the public randomness is known to both parties whereas uncorrelated private randomness require coordination (or communication). In contrast, in the context of defective randomness, two independent sources (even when each is only seen by one party) seem preferable to a single source of (defective) public randomness. We stress that our results only suggest that the communication complexity in the private (defective-randomness) case may be lower than in the public (defective-randomness) case, and establishing such a separation is left as an open problem. We mention that for some functions such separation does not exist (see Proposition 3.5).

We focus on the case of min-entropy that is at least logarithmic in the length of the input to the protocol (i.e., $k \geq \log_2 n$), because this is the minimal amount of perfect randomness that is required for constant-communication protocols for equality.[3] Still, one may study the case of sub-logarithmic min-entropy (and possibly integrate our results with those of [1]).

[1] In particular, the fact that a random source of logarithmic length suffices does not apply here: we are given defective random sources of certain length, and cannot easily transform them to significantly shorter length.

[2] However, as shown by Dietzfelbinger, Hromkovic, and Schnitger [4], the deterministic complexity may be exponentially larger than the lower bound provided by any fooling set.

[3] See [1, Thm. 3], which shows that computing the equality function when having access only to k bits of perfect public randomness requires communication complexity $\Omega(n/2^k)$.

1.3 Remotely Related Works

Goldwasser, Sudan, and Vaikuntanathan [6] raised the general question of which distributed computing tasks that require randomness can be performed also when having access to defective sources of randomness.[4] Specifically, they showed that (Byzantine) agreement tasks fall into this category; that is, they can be performed quite well also in the case that each party has access to a (single) defective source of randomness. We stress that since the parties do not trust each other, the fact that their sources are independent of one another does not mean that they can extract almost perfect randomness by using some adequate extractor.

The following works that refer to different models of communication complexity are more related to the current study.

- Canonne *et al.* [2] considered a model that lies between the standard public and private randomness models (when the amount of randomness is sublogarithmic in the length of the inputs). Specifically, they considered two parties that are each given access to a private source of perfect randomness such that the two sources are tightly correlated (i.e., for a parameter $\rho \in [\pm 1]$, for each i, the i^{th} bit in the first source is ρ-correlated with the i^{th} bit in the second source). We mention that their motivation is not to study the usefulness of defective sources of randomness but rather to study the effect on uncertainty (about "contents") in communication complexity.

- Canetti and Goldreich [1] studied trade-offs between randomness and communication complexity. In particular, they showed that a logarithm amount of (perfect) randomness is sufficient for any communication protocol and that in some cases this upper bound is tight.

- Chor and Goldreich [3] studied the "distributional communication complexity" of functions when the protocol is only required to be correct with a specified probability $p > 1/2$, where the probability is taken over input pairs that are each chosen according to some distribution of specified min-entropy bound (i.e., min-entropy at least k). We stress that their study is fundamentally different from ours; they study the average-case (on inputs) behavior of protocols, where the inputs are drawn from a defective source of randomness, whereas we study the worst-case (on inputs) behavior of protocols that employ defective sources of randomness.

2 Preliminaries

We consider two-party randomized protocols for computing functions of the form $f : \{0,1\}^n \times \{0,1\}^n \to \{0,1\}$, while using a defective source of randomness. Specifically, we consider sources of randomness that produce ℓ-bit long strings having min-entropy at least k; that is, each outcome occurs with probability at most 2^{-k}. Such sources are called (ℓ, k)-sources.

We consider both the public-randomness model in which the parties have access to common (public) randomness, and the private-randomness model in which each party has its private source of randomness, which is independent of the randomness of the other party. In our context (of defective random sources) it is important to stress that the postulated sources of randomness are the only ones available to the parties.

The results hold not only for "alternating protocols" (in which the parties alternatively exchange single bits), but directly for any protocol in which the sender of the next bit is determined by the

[4] In a somewhat related vein, a body of work has investigated whether defective randomness suffices for cryptographic security in a variety of settings. McInnes and Pinkas [8] initiated this line of work by showing that information theoretic symmetric key cryptography is impossible without pure randomness. Dodis *et al.* [5] later extended this result to rule out the feasibility of a variety of cryptographic tasks from defective randomness, including computationally-secure symmetric key cryptography.

communication so far; that is, no need to lose a factor of two in translation (from such general protocols to "alternating" ones).

2.1 Specific Background About Communication Complexity

We shall use the following basic result that refers to deterministic communication protocols.

Claim 2.1 (the "corners lemma" (cf., e.g., [7, Prop. 1.13–1.14] or [10, Lem. 1.3–1.4])): *Let Π' be a deterministic communication protocol and suppose that $\gamma \stackrel{\text{def}}{=} \Pi'(x_1, x_2) = \Pi'(y_1, y_2)$. Then, $\Pi'(x_1, y_2) = \Pi'(y_1, x_2) = \gamma$.*

In addition, a basic notion of communication complexity that underlies many of its lower bound proofs is that of a fooling set, defdined as follows.

Definition 2.2 (fooling set (cf., e.g., [7, Sec. 1.3] or [10, Chap. 1])): *We say that $S \subseteq \{0,1\}^{n+n}$ is a* fooling set for $f : \{0,1\}^{n+n} \to \{0,1\}$ *if every f-monochromatic rectangle contains at most one point in S, where an f-*monochromatic rectangle *is a set $X \times Y$ such that $X, Y \subseteq \{0,1\}^n$ and f is constant on $X \times Y$* (i.e., $f(x,y) = f(x',y')$ for every $(x,y),(x',y') \in X \times Y$).

Note that a fooling set cannot contain two pre-images of $f^{-1}(0)$ (resp., $f^{-1}(1)$) that differ only on one coordinate; that is, if (x,y) and (x',y') are in a fooling set for f and $f(x,y) = f(x',y')$, then $x \neq x'$ and $y \neq y'$ (because two points that differ on a single coordinate constitute an f-monochromatic rectangle).

2.2 Specific Background About Randomness Extraction

As stated above, an (ℓ, k)-source is a distribution over ℓ-bit long strings having min-entropy at least k, where the min-entropy of a random variable X is $\min_{v \in \mathrm{Supp}(X)}\{\log_2(1/\Pr[X=v])\}$. That is, X has min-entropy k if and only if for every v it holds that $\Pr[X=v] \leq 2^{-k}$.

We say that $\mathrm{EXT} : \{0,1\}^d \times \{0,1\}^\ell \to \{0,1\}^m$ is a (seeded) (k,ϵ)-extractor if for every random variable X of min-entropy k the total variation distance between $\mathrm{EXT}(U_d, X)$ and U_m is at most ϵ, where U_n denotes the uniform distribution on $\{0,1\}^n$. In this case ϵ is called the error of EXT, and d is its seed length.

We say that $\mathrm{EXT} : \{0,1\}^\ell \times \{0,1\}^\ell \to \{0,1\}^m$ is a two-source extractor for independent (ℓ,k)-sources if for every two independent random variables X and Y of min-entropy k the total variation distance between $\mathrm{EXT}(X,Y)$ and U_m is at most ϵ, called its error. This definition is readily extended to independent sources of parameters (ℓ_1, k_1) and (ℓ_2, k_2) respectively.

3 The Public-Randomness Model

For a protocol Π in the public-randomness model, we denote by $\Pi(x,y;r)$ the transcript of the communication on input $(x,y) \in \{0,1\}^{n+n}$ and randomness $r \in \{0,1\}^\ell$. The output of such a protocol is determined by its transcript (e.g., it may be its last bit), and is denoted $\overline{\Pi}(x,y;r)$.

Definition 3.1 (communication complexity with a weak public source): *An (ℓ,k)-*public-randomness protocol for computing a function *$f : \{0,1\}^n \times \{0,1\}^n \to \{0,1\}$ is a protocol that satisfies $\Pr[\overline{\Pi}(x,y;\Xi) = f(x,y)] \geq 2/3$, for every $(x,y) \in \{0,1\}^{n+n}$ and every (ℓ,k)-source Ξ.*

Theorem 3.2 (a general lower bound): *Suppose that $f : \{0,1\}^{n+n} \to \{0,1\}$ has a fooling get of size 2^m. Then, any (ℓ,k)-public-randomness protocol for computing f has communication complexity at least $\min(m-1, \ell-k-1)/2$.*

Proof: Suppose that f has a (ℓ, k)-public-randomness protocol, denoted Π, of communication complexity $t \leq (n-1)/2$. We first observe that there exists a dense set of possible source-outcomes R and two input pairs (x_1, y_1) and (x_2, y_2) that reside in the fooling set such that Π in constant on all triples (x_i, y_i, r), where $r \in R$ and $i \in \{1, 2\}$. The theorem will follow by using the standard "corners lemma" (in a non-standard way) and defining a source that is uniform over R. Details follow.

The following technical claim has nothing to do with communication complexity; it holds for any function $F : [2^m] \times \{0,1\}^\ell \to \{0,1\}^t$, where in the current case $[2^m]$ represents the indices of the strings in the fooling set (for f), $\{0,1\}^\ell$ represents possible outcomes of the public source, and $\{0,1\}^t$ represents possible transcripts of Π.

Claim 3.2.1 (a simple combinatorial claim): *Let $F : [2^m] \times \{0,1\}^\ell \to \{0,1\}^t$. Then, for any $S = \{(x_1, y_1), ..., (x_{2^m}, y_{2^m})\}$ and $t \leq (m-1)/2$, there exist distinct $i, j \in [2^m]$, a string $\gamma \in \{0,1\}^t$, and a set $R \subseteq \{0,1\}^\ell$ of density at least 2^{-2t-1} such that for every $r \in R$ it holds that $F(i, r) = F(j, r) = \gamma$.*

We will apply Claim 3.2.1 to the hitting set S and to the function $F(i, r) \stackrel{\text{def}}{=} \Pi(x_i, y_i; r)$. But let us prove the claim first.

Proof: A simple counting implies that, for every $i \in [2^m]$, there exist $\gamma_i \in \{0,1\}^t$ and a set $R_i \subseteq \{0,1\}^\ell$ of density 2^{-t} such that for every $r \in R_i$ it holds that $F(i, r) = \gamma_i$. Similarly, there exist $\gamma \in \{0,1\}^t$ and $G \subseteq [2^m]$ of density 2^{-t} such that $\gamma_i = \gamma$ for every $i \in G$.

The key observation is that if $t \leq (n-1)/2$, then there exist distinct $i, j \in G$ such that $|R_i \cap R_j| \geq 2^{\ell-2t-1}$. This is shown by fixing an arbitrary $G' \subseteq G$ of size 2^{t+1}, which is possible since $2^{t+1} \leq 2^{n-t}$, and assuming towards the contradiction that, for every distinct $i, j \in G'$, it holds that $|R_i \cap R_i| < 2^{\ell-2t-1}$. Then, we get

$$\begin{aligned}
\left|\sum_{i \in G'} R_i\right| &\geq \sum_{i \in G'} |R_i| - \sum_{i \neq j \in G'} |R_i \cap R_j| \\
&> 2^{t+1} \cdot 2^{\ell-t} - \binom{2^{t+1}}{2} \cdot 2^{\ell-2t-1} \\
&> 2^{\ell+1} - 2^{2t+1} \cdot 2^{\ell-2t-1} \\
&= 2^\ell
\end{aligned}$$

which is impossible. The claim follows by fixing $i \neq j$ such that $|R_i \cap R_j| \geq 2^{\ell-2t-1}$, and defining $R = R_i \cap R_j$. ■

Applying Claim 3.2.1 to the hitting set $S = \{(x_1, y_1), ..., (x_{2^m}, y_{2^m})\}$ of the hypothesis, while letting $F(i, r) \stackrel{\text{def}}{=} \Pi(x_i, y_i; r)$ and using $t \leq (m-1)/2$, we infer that the fooling set contains two points (x_i, y_i) and (x_j, y_j) such that $\Pi(x_i, y_i; r) = \Pi(x_j, y_j; r) = \gamma$ holds for any $r \in R$, where $R \subseteq \{0,1\}^\ell$ has density at least 2^{-2t-1}.

Next, applying the "corners lemma" (i.e., Claim 2.1), we infer that $\Pi(x_i, y_i; r) = \Pi(x_i, y_j; r) = \Pi(x_i, y_j; r) = \Pi(x_j, y_j; r)$ for every $r \in R$. Note that this application of the "corners lemma" refers to the residual deterministic protocols $\Pi'_r(x, y) = \Pi(x, y; r)$, for all $r \in R$, and it implies that $\Pi'_r(x_i, y_j) = \Pi'_r(x_j, y_i) = \gamma$ for each $r \in R$.

Lastly, picking an $(\ell, \ell-2t-1)$-source that is uniform on R, we infer that, when fed with randomness from this source, the execution of Π does not distinguish these four input-pairs (i.e., (x_i, y_i), (x_i, y_j), (x_i, y_j) and (x_j, y_j)). On the other hand, by hypothesis that (x_i, y_i) and (x_j, y_j) belong to a fooling set, these four input-pairs cannot have the same f-value (i.e., it cannot be that $f(x_i, y_i) = f(x_i, y_j) = f(x_j, y_i) = f(x_j, y_j)$, since this would mean that (x_i, y_i) and (x_j, y_j) reside in the f-monochromatic rectangle $\{x_i, x_j\} \times \{y_i, y_j\}$). Hence, the hypothesis that Π is a (ℓ, k)-public-randomness protocol for f implies that the foregoing source has min-entropy below k; that is, $\ell - 2t - 1 < k$. The theorem follows, since we established $t > (\ell - k - 1)/2$, under the hypothesis $t \leq (m-1)/2$. ■

An archetypical corollary. Recalling that equality has a fooling set of size 2^n and applying Theorem 3.2, we get

Corollary 3.3 (lower bound for equality): *Any (ℓ,k)-public-randomness protocol for computing equality of n-bit strings has communication complexity at least* $\min(n-1,\ell-k-1)/2$.

This lower bound is tight up to a constant factor, since equality has a constant communication protocol in the standard public-randomness model and the following generic result definitely applies to it.

Theorem 3.4 (a generic upper bound): *Suppose that $f:\{0,1\}^{n+n}\to\{0,1\}$ has communication complexity $\mathcal{C}^{\mathtt{pub}}(f)$ in the standard public-randomness model. Then, for every $k\leq\ell$ such that $k>\log_2 n+O(1)$, there exists an (ℓ,k)-public-randomness protocol for computing f with communication complexity $O(\ell-k)\cdot\mathcal{C}^{\mathtt{pub}}(f)$.*

Recall that equality has constant communication complexity in the standard public-randomness model.

Proof: Recall that the *randomness complexity* of any protocol for computing f can be reduced to $m\stackrel{\text{def}}{=}\log_2 n+O(1)$ (while possibly increasing its communication complexity by a constant factor).[5] The key observation is that the parties can *emulate the extraction* of m almost-random bits from the public (ℓ,k)-source, by trying all possible seeds for an adequate randomness extractor, and use the extracted bit to emulate the original randomized protocol. Specifically, for $k\geq m$, such extraction is possible using a (perfectly random) seed of length $d\stackrel{\text{def}}{=}\log(\ell-k)+O(1)$ (see, e.g., [11, Sec. 3.1]). Hence, the parties can emulate the randomized protocol by invoking it 2^d times using as randomness the "extracted outputs" under all possible seeds. Details follow.

Let $\mathrm{EXT}(s,r)$ denote the output of the extractor $\mathrm{EXT}:\{0,1\}^d\times\{0,1\}^\ell\to\{0,1\}^m$ on seed s and source outcome r. Then, given (defective) public-randomness $r\in\{0,1\}^\ell$, the parties emulate 2^d invocations of the standard randomized protocol such that in the i^{th} invocation they use public-randomness $\mathrm{EXT}(i,r)$, where $i\in[2^d]\equiv\{0,1\}^d$, and rule by majority. Actually, we use a randomized protocol for the standard model that has error probability at most 0.1 (rather than at most 1/3), which can be obtained by a constant number of repetitions.

We claim that if EXT has error 0.05 on any (ℓ,k)-source R, then, for every fixed input pair, with probability at least 2/3 over the outcome of R, the majority of the extracted values (over all possible seeds) yield protocol executions with the correct output. This is the case because otherwise the statistical difference between $\mathrm{EXT}(U_d,R)$ and U_m is at least $\frac{1}{3}\cdot\frac{1}{2}-0.1>0.05$, where the first (resp., second) term represents a lower bound (resp., upper bound) on the probability that the protocol yields a wrong answer when run with randomness $\mathrm{EXT}(U_d,R)$ (resp., U_m). This yields an (ℓ,k)-public-randomness protocol of communication complexity $2^d\cdot O(\mathcal{C}^{\mathtt{pub}}(f))=O(\ell-k)\cdot\mathcal{C}^{\mathtt{pub}}(f)$. ■

On the gap between the lower and upper bound. The bounds provided by Theorems 3.3 and 3.4 leave a gap of a factor $\Theta(\mathcal{C}^{\mathtt{pub}}(f))$ in the non-trivial case (i.e., $\Omega(\ell-k)$) versus $O(\ell-k)\cdot\mathcal{C}^{\mathtt{pub}}(f))$. The following example implies that the gap cannot be closed by increasing the lower bound.

Proposition 3.5 (improved upper bound): *For every $m<n$, there exists a function $f:\{0,1\}^{n+n}\to\{0,1\}$ that satisfies the following two conditions:*

1. *The function f has communication complexity $\mathcal{C}^{\mathtt{pub}}(f)=\Theta(m)$ in the standard public-randomness model;*

[5] See, e.g., [1, Thm. 5] and [9]. The basic argument leaves the communication complexity intact, but increases the error probability by an arbitrary small constant, where this constant effects the additive constant in m. To regain the original error bound, three repetitions suffice.

2. *For every $k \leq \ell$ such that $k > \log_2 n + O(1)$, there exists an (ℓ, k)-public-randomness protocol for computing f with communication complexity $O(\ell - k) + O(\mathcal{C}^{\mathtt{pub}}(f))$.*

Proof: Consider the function $f(x'x'', y'y'') = \mathtt{EQ}(x', y') \oplus \mathtt{IP}_2(x'', y'')$, where $|x''| = m = n - |x'|$, $\mathtt{EQ}$ denotes the equality function, and $\mathtt{IP}_2$ denotes inner-product mod 2. Then, $\mathcal{C}^{\mathtt{pub}}(f) \geq \mathcal{C}^{\mathtt{pub}}(\mathtt{IP}_2) = \Omega(m)$, where the first inequality follows by a straightforward reduction and the lower bound is proved in [3]. We obtain an (ℓ, k)-public-randomness protocol for computing f with communication complexity $O(\ell - k) \cdot \mathcal{C}^{\mathtt{pub}}(\mathtt{EQ}) + m + 1 = O(\ell - k) + O(\mathcal{C}^{\mathtt{pub}}(f))$, by combining the generic protocol for $\mathtt{EQ}$ (see Theorem 3.4) with the straightforward deterministic protocol for $\mathtt{IP}_2$. ■

4 The Private-Randomness Model

For a protocol Π in the private-randomness model, we denote by $\Pi((x, r), (y, s))$ the transcript of the communication on input $(x, y) \in \{0,1\}^{n+n}$ with private randomness $r, s \in \{0,1\}^{\ell}$; that is, the first (resp., second) party gets input x (resp., y) and private randomness r (resp., s). The output of such a protocol is determined by its transcript (e.g., it may be its last bit), and is denoted $\overline{\Pi}((x, r), (y, s))$.

Definition 4.1 (communication complexity with weak private sources): *An (ℓ, k)-*private-randomness protocol for computing a function *$f : \{0,1\}^n \times \{0,1\}^n \to \{0,1\}$ is a protocol that satisfies $\Pr[\overline{\Pi}((x, \Xi'), (y, \Xi'')) = f(x, y)] \geq 2/3$, for every $(x, y) \in \{0,1\}^{n+n}$ and every pair of independent (ℓ, k)-sources Ξ' and Ξ''.*

Theorem 4.2 (a general lower bound): *Suppose that $f : \{0,1\}^{n+n} \to \{0,1\}$ has a fooling get of size 2^m. Then, any (ℓ, k)-private-randomness protocol for computing f has communication complexity at least $\min(m - 1, \ell - k - 1)/2$.*

Proof: The proof is analogous to the proof of Theorem 3.2. We start with a hypothetical (ℓ, k)-private-randomness protocol, denoted Π, that computes f with communication complexity $t \leq (n-1)/2$. Then, we apply Claim 3.2.1 to the (somewhat less natural) function $F : [2^m] \times \{0,1\}^{\ell} \to \{0,1\}^t$ defined by $F(i, r) \stackrel{\text{def}}{=} \Pi((x_i, r), (y_i, r))$, where $S = \{(x_1, y_1), ..., (x_{2^m}, y_{2^m})\}$ is a fooling set for f. Hence, we infer that *there exist distinct $i, j \in [2^m]$, a string $\gamma \in \{0,1\}^t$, and a set $R \subseteq \{0,1\}^{\ell}$ of density at least 2^{-2t-1} such that for every $r \in R$ it holds that $\Pi((x_i, r), (y_i, r)) = \Pi((x_j, r), (y_j, r)) = \gamma$.*

Now, applying Claim 2.1 thrice, we infer that $\Pi((x_a, r), (y_b, s)) = \gamma$ for every $r, s \in R$ and $a, b \in \{i, j\}$. Specifically, for both $a \in \{i, j\}$ and every $r, s \in R$, considering the residual protocol $\Pi'_a(r, s) = \Pi((x_a, r), (y_a, s))$ and using $\Pi((x_a, r), (y_a, r)) = \Pi((x_a, s), (y_a, s)) = \gamma$, we infer that $\Pi((x_a, r), (y_a, s)) = \gamma$. Hence, $\Pi((x_i, r), (y_i, s)) = \gamma = \Pi((x_j, r), (y_j, s))$. Now, considering the residual protocol $\Pi'_{r,s}(x, y) = \Pi((x, r), (y, s))$ and using $\Pi((x_i, r), (y_i, s)) = \Pi((x_j, r), (y_j, s))$, we get that $\Pi((x_i, r), (y_j, s)) = \gamma = \Pi((x_j, r), (y_i, s))$.

Picking a pair of independent $(\ell, \ell - 2t - 1)$-sources that are each uniform on R, we infer that the execution of Π does not distinguish the four input-pairs (x_i, y_i), (x_i, y_j), (x_i, y_j) and (x_j, y_j). On the other hand, by hypothesis that (x_i, y_i) and (x_j, y_j) belong to a fooling set, and so these four input-pairs cannot have the same f-value. Hence, the hypothesis that Π is a (ℓ, k)-private-randomness protocol for f implies that $\ell - 2t - 1 < k$. The theorem follows, since we established $t > (\ell - k - 1)/2$, under the hypothesis $t \leq (m - 1)/2$. ■

Theorem 4.3 (a generic upper bound): *Suppose that $f : \{0,1\}^{n+n} \to \{0,1\}$ has communication complexity $\mathcal{C}^{\mathtt{priv}}(f)$ in the standard private-randomness model. Then, for every $k \leq \ell$ such that $k > 2\log_2 n + 2\log_2 \ell + O(1)$, there exists an (ℓ, k)-private-randomness protocol for computing f with communication complexity $\min(2(\ell - k) + 3\log_2 n + O(\mathcal{C}^{\mathtt{priv}}(f)), \ell + \log_2 n + O(\mathcal{C}^{\mathtt{priv}}(f))))$.*

Proof: The bounds follow by having one party send a $\min(2 \cdot (\ell - k + \log_2 n + O(1)), \ell)$-bit long prefix of its private randomness to the second party, who extracts almost perfect randomness from the two outcomes (using a two-source extractor), sends one half of it back, and then both parties execute the standard protocol. Details follow.

First, recall that the randomness complexity of any protocol for computing f can be reduced to $m \stackrel{\text{def}}{=} \log_2 n + O(1)$ (while possibly increasing its communication complexity by a constant factor). Second, recall that a seedless (two-source) randomness extractor can extract $2m$ almost random bits from an (ℓ, k)-source and an independent (ℓ', k')-source, provided that $2m \leq k + k' - \max(\ell, \ell') - O(1)$ (see [3, Thm. 7(2)]).[6] Now, if $\ell' \stackrel{\text{def}}{=} 2 \cdot (\ell - k + \log_2 n) + O(1) \leq \ell$, then an ℓ'-bit prefix of an (ℓ, k)-source has min-entropy $k' \stackrel{\text{def}}{=} \ell' - (\ell - k) = (\ell - k) + 2\log_2 n + O(1)$, and so $k + k' - \max(\ell, \ell') - O(1) = 2\log_2 n + O(1)$. Hence, sending the prefix of the first source sent to the second party, allows it to extract $2\log_2 n + O(1)$ bits that are almost random. Sending half of these bits to the first party allows the two parties to emulate the original protocol. The communication complexity of the proposed protocol is at most $\ell' + \log_2 n + O(1) + O(\mathcal{C}^{\mathtt{priv}}(f))$, which equals $2(\ell - k) + 3\log_2 n + O(\mathcal{C}^{\mathtt{priv}}(f))$.

As for the case of $\ell' > \ell$, recall that a seedless (two-source) randomness extractor can extract $2m$ almost random bits from a pair of independent (ℓ, k)-source, provided that $2m \leq k - 2\log_2 \ell - O(1)$ (see [3, Thm. 7(1)]). In this case, sending the outcome of the first source to the second party allows for the foregoing emulation, at a total communication cost of $\ell + \log_2 n + O(\mathcal{C}^{\mathtt{priv}}(f))$. ■

Acknowledgments

Marshall Ball was partially supported by an IBM Research PhD Fellowship. Oded Goldreich was partially supported by an ISF grant number (Nr. 1146/18), and the research was conducted while he enjoyed the hospitality of the computer science department at Columbia University.

This work was supported in part by the Office of the Director of National Intelligence (ODNI), Intelligence Advanced Research Projects Activity (IARPA) via Contract No. 2019-1902070006. The views and conclusions contained herein are those of the authors and should not be interpreted as necessarily representing the official policies, either express or implied, of ODNI, IARPA, or the U.S. Government. The U.S. Government is authorized to reproduce and distribute reprints for governmental purposes notwithstanding any copyright annotation therein.

References

[1] Ran Canetti and Oded Goldreich. Bounds on Tradeoffs between Randomness and Communication Complexity. In *31st FOCS*, pages 766–775, 1990.

[2] Clement Canonne, Venkatesan Guruswami, Raghu Meka, and Madhu Sudan. Communication with Imperfectly Shared Randomness. *ECCC*, TR14-153, 2014.

[3] Benny Chor and Oded Goldreich. Unbiased Bits from Sources of Weak Randomness and Probabilistic Communication Complexity. *SICOMP*, Vol. 17 (2), pages 230–261, 1988.

[4] Martin Dietzfelbinger, Juraj Hromkovic, and Georg Schnitger. A Comparison of Two Lower-Bound Methods for Communication Complexity. *Theoretical Computer Science*, Vol. 168 (1), pages 39–51, 1996.

[5] Yevgeniy Dodis, Shien Jin Ong, Manoj Prabhakaran, and Amit Sahai. On the (Im)possibility of Cryptography with Imperfect Randomness. In *45th FOCS*, pages pages 196–205, 2004.

[6] Note that for every $d \geq 0$, a (ℓ, k)-source may be viewed as an $(\ell + d, k)$-source.

[6] Shafi Goldwasser, Madhu Sudan, and Vinod Vaikuntanathan. Distributed Computing with Imperfect Randomness. *19th DISC*, pages 288–302, 2005.

[7] Eyal Kushilevitz and Noam Nisan. *Communication Complexity*. Cambridge University Press, 1996.

[8] James L. McInnes and Benny Pinkas. On the Impossibility of Private Key Cryptography with Weakly Random Keys. In *Proc. of CRYPTO'90*, pages 421–436, 1991.

[9] Ilan Newman. Private vs. Common Random Bits in Communication Complexity. *IPL*, Vol. 39 (2), pages 67–71, 1991.

[10] Anup Rao and Amir Yehudayoff. *Communication Complexity*. Cambridge University Press, 2020.

[11] Ronen Shaltiel. An introduction to randomness extractors. *38th ICALP*, Lecture Notes in Computer Science, Vol. 6756, pages 21-41, 2011.