High Entropy Random Selection Protocols

Harry Buhrman [*] Matthias Christandl [†] Michal Koucký [‡] Zvi Lotker [§]
Boaz Patt-Shamir [¶] Nikolay Vereshchagin [‖]

May 17, 2021

Abstract

We study the two party problem of randomly selecting a common string among all the strings of length n. We want the protocol to have the property that the output distribution has high *Shannon entropy* or high *min entropy*, even when one of the two parties is dishonest and deviates from the protocol. We develop protocols that achieve high, close to n, Shannon entropy and simultaneously min entropy close to $n/2$.

In the literature the randomness guarantee is usually expressed in terms of "resilience". The notion of Shannon entropy is not directly comparable to that of resilience, but we establish a connection between the two that allows us to compare our protocols with the existing ones.

We construct an explicit protocol that yields Shannon entropy $n - O(1)$ and has $O(\log^* n)$ rounds, improving over the protocol of Goldreich et al. [15] that also achieves this entropy but needs $O(n)$ rounds. Both these protocols need $O(n^2)$ bits of communication.

Next we reduce the number of rounds and the length of communication in our protocols. We show the existence, non-explicitly, of a protocol that has 6 rounds, $O(n)$ bits of communication and yields Shannon entropy $n - O(\log n)$ and min entropy $n/2 - O(\log n)$. Our protocol achieves the same Shannon entropy bound as, also non-explicit, protocol of Gradwohl et al. [17], however achieves much higher min entropy: $n/2 - O(\log n)$ versus $O(\log n)$.

Finally we exhibit a very simple 3-round explicit "geometric" protocol with communication length $O(n)$. We connect the security parameter of this protocol with the well studied Kakeya problem motivated by Harmonic Analysis and Analytic Number Theory. We prove that this protocol has Shannon entropy $n - o(n)$. Its relation to the Kakeya problem follows a new and different approach to the random selection problem than any of the previously known protocols.

[*]CWI and University of Amsterdam. Email: buhrman@cwi.nl.

[†]Cambridge University. Email: mc380@cam.ac.uk, work done while visiting CWI.

[‡]Computer Science Institute of Charles University. Email:koucky@iuuk.mff.cuni.cz, work done while visiting CWI.

[§]Bar Ilan University, Ramat-Gan, Israel zvi.lotker@gmail.com, work done while visiting CWI.

[¶]Tel Aviv University. Email: boaz@eng.tau.ac.il

[‖]Moscow State University and HSE University, Russian Federation. Email: ver@mccme.ru, the work was partially done while visiting CWI.

1 Introduction

We study the following communication problem. Alice and Bob want to select a common random string. They are not at the same location so they do not see what the other player does. They communicate messages according to some protocol, and in the end they output a string of n bits which is a function of the messages communicated. This string should be as random as possible, and in our case we measure the amount of randomness by Shannon entropy or min entropy of the probability distribution that is generated by this protocol.

The messages they communicate may depend on random experiments the players perform and on messages sent so far. The outcome of an experiment is known only to the party which performs it, so the other party cannot verify the outcome of such an experiment or whether the experiment was carried out at all. One or both the parties may deviate from the protocol and try to influence the selected string (*cheat*). We are interested in the situation when a party honestly follows the protocol and wants to have some guarantee that the selected string is indeed as random as possible.

2 Previous work and our results

There is a large body of previous work which considers the problem of random string selection, and related problems such as a leader selection and fault-tolerant computation. In some work on random selection, such as Blum's "coin-tossing over the telephone" [4], the adversary is assumed to be computationally bounded (e.g., probabilistic polynomial time). Generally, in this setting one utilizes one-way functions and other cryptographic primitives to limit the adversary's ability to cheat, and thus the resulting protocols rely on complexity assumptions.

In this paper, we study *the information-theoretic setting* (also known as "full information model"), where the adversary is computationally unbounded. In addition to its stronger security guarantees, the information-theoretic setting has the advantage that protocols typically do not require complexity-theoretic assumptions (such as the existence of one-way functions). Various such random selection protocols have been used to construct perfectly hiding bit-commitment schemes [19], to convert honest-verifier zero-knowledge proofs into general zero-knowledge proofs [9, 10, 16], to construct oblivious transfer protocols in the bounded storage model [8, 11], and to perform general fault-tolerant computation [15]. There has also been substantial work in the k-party case for $k \geq 3$, where the goal is to tolerate coalitions of a minority of cheating players. This body of work includes the well-studied "collective coin-flipping" problem e.g., [5, 24, 1, 22, 23, 14] (closely related to the "leader election" problem), and again the use of random selection as a tool for general fault-tolerant computation [15].

Note that unlike Blum's "coin-tossing over the telephone", in the information-theoretic setting we have to assume that $n > 1$. Indeed, for $n = 1$ for any protocol either Alice has a strategy to force the outcome 0, or Bob has a strategy to force the outcome 1. This follows from Zermelo's theorem for finite two-person game of perfect information [28]. On the other hand, for $n = 1$ non-trivial protocols exist for quantum coin tossing, see [20] for an overview.

Several different measures for the randomness guarantee of the protocol are used in the literature. The most widely used is the (μ, ϵ)-resilience. A two-party protocol is (μ, ϵ)-*resilient* if for every set $S \subset \{0,1\}^n$ with density μ (cardinality $\mu 2^n$), the output of the

protocol is in S with probability at most ϵ, provided at least one party is honest.

In this paper, however, we study another very natural randomness guarantee, namely min entropy or Shannon entropy of the resulting output distribution. We are not aware of any applications of randomness sources with high Shannon or min entropy in two-party protocols, thus our interest in protocols to generate such distributions is purely philosophical.

There is a very simple 2-round selection protocol with linear communication to obtain a distribution over n-bit strings with min entropy at least $n/2$: Alice selects at random the first half of the output string, and then Bob selects its second half. If at least one party is honest, then the min entropy of the resulting distribution is at least $n/2$. It follows from Goldreich et al. [15] and from a bound on quantum coin-flipping due to Kitaev (see [3]) that no protocol can achieve min entropy larger than $n/2$, thus for min entropy the problem is easy.

For Shannon entropy the situation is more complicated. There is a certain relationship between Shannon entropy and resilience: using Lemma 2 below we can deduce Shannon entropy guarantees from resilience guarantees. In particular, we can show that a protocol from [15] generates a distribution with Shannon entropy $n - O(1)$, which is close to the maximum, since Shannon entropy of any distribution over n-bit strings is at most n. More specifically, Goldreich et al. [15] constructed a protocol running in $O(n)$ rounds and communicating $O(n^2)$ bits that is $(\mu, O(\sqrt{\mu}))$-resilient for all $\mu > 0$. By Lemma 2 below this protocol generates a distribution with Shannon entropy $n - O(1)$.

A natural question is whether one can reduce the number of rounds or communication length keeping the guarantee $n - O(1)$ for Shannon entropy of the resulting distribution.

Reducing the number of rounds to $o(n)$ rounds. Regarding the number of rounds, we answer this question in positive by designing an explicit protocol ("the main protocol") that runs in $O(\log^* n)$ rounds, communicates $O(n^2)$ bits and generates a distribution with Shannon entropy $n - O(1)$ (Theorem 7). Moreover, a version of this protocol also guarantees the min entropy at least $n/2 - O(\log n)$ (Corollary 9).

Sanghvi and Vadhan in [25] showed a lower bound $\Omega(\log^* n)$ on the number of rounds of any random selection protocol that achieves constant statistical distance from the uniform distribution (for some constant less than 1). In Lemma 10, we show that Shannon entropy $n - O(1)$ implies having constant statistical distance from the uniform distribution (for some constant less than 1), so their lower bound translates to our protocols: our upper bound $O(\log^* n)$ for the number of rounds is tight.

Constantly many rounds. The next question is what entropy guarantee can be provided by protocols running in constantly many rounds. Gradwohl et al. [17] showed that there is a *non-explicit* $(1/n^2, O(1/n))$-resilient linear-communication protocol running in constantly many rounds. By Lemma 2 this resilience guarantee implies that the Shannon entropy of the protocol is $n - O(\log n)$.

In Section 4 we present an *explicit* protocol P_0 (Lemma 5) with Shannon entropy $n - \log n$ that runs in just 3 rounds. Its communication length is $O(n^2)$, which is larger than that of Gradwohl et al. However, regarding the Shannon entropy and the number of rounds this protocol is optimal: by a result of Stepanov [26] no three-round protocol can generate a distribution with Shannon entropy larger than $n - \log n + O(\log \log n)$, and no two-round

protocol can generate a distribution with Shannon entropy larger than $n/2 + O(\log \log n)$. A version of protocol P_0 has 6 rounds, communication length $O(n^2)$ and it simultaneously achieves Shannon entropy $n - O(\log n)$ and min entropy $n/2 - O(\log n)$ (Corollary 6).

More generally, for each $i \geq 0$, we present an explicit protocol P_i (Lemma 8) running in $2i + 3$ rounds with communication length $O(n^2)$ and Shannon entropy $n - \log^{(i+1)} n$ where $\log^{(j)} n$ denotes the jth iteration of the $\lceil \log_2 n \rceil$ function.

Reducing communication length. As we have mentioned, our main protocol, as well as the protocol of [15], has communication length $O(n^2)$ and Shannon entropy $n - O(1)$. It remains open whether one can reduce communication length in those protocols: it is unknown whether there are protocols with communication length $o(n^2)$ and Shannon entropy $n - O(1)$. There exist protocols with subquadratic communication and some resilience guarantees, however those guarantees do not imply Shannon entropy $n - O(1)$. More specifically, we mean the following two protocols.

1. Gradwohl et al. [17] for each μ exhibit a non-explicit $O(\log^* n)$-round protocol that is $(\mu, O(\sqrt{\mu}))$-resilient and that uses linear communication. The resilience guarantee of that protocol holds only for one value of μ, which is not enough for having entropy $n - O(1)$.[1]

2. Sanghvi and Vadhan [25], for every constant $\delta > 0$, give a protocol with communication length $n \log^{O(1)} n$ that is $(\mu, \sqrt{\mu + \delta})$-resilient (for every μ) and that has constant statistical distance from the uniform distribution (for some constant less than 1). This type of resilience essentially guarantees security only for sets of constant density. Indeed, their protocol allows the cheating party to bias the output distribution so that a particular string x_0 has a constant positive probability ε of being the output. Hence, the output distribution of their protocol may have constant min entropy and Shannon entropy at most $(1 - \epsilon)n + 1$ for some $\epsilon > 0$.[2]

However, it turns out that it is possible to reduce communication length (from $O(n^2)$) in our constant round protocols with Shannon entropy $n - o(n)$. As we already mentioned, Gradwohl et al. [17] showed that there is a non-explicit protocol running in constantly many rounds, with linear communication and Shannon entropy $n - O(\log n)$. In this paper, we show, also non-constructively (Theorem 12), that there is a protocol with linear communication complexity that achieves Shannon entropy $n - O(\log n)$ in just 3 rounds. Moreover, we show non-constructively (Corollary 14) that there is a 6-round protocol with linear communication complexity that achieves Shannon entropy $n - O(\log n)$ and simultaneously min-entropy $n/2 - O(\log n)$.

Constantly many rounds and linear communication: explicit protocols. The above mentioned protocols with linear communication and constantly many rounds are non-explicit. A natural question is whether there are explicit constant-round and linear-communication protocols guaranteeing Shannon entropy $n - o(n)$. In this paper, we con-

[1] Assume for instance that a random variable X with the range $\{0,1\}^n$ is $(\mu, 2\sqrt{\mu})$ for some μ. If $\mu \geq 1/\sqrt{n}$ then X may have the following distribution: $\Pr[X = 00\ldots0] = \mu$ and the remaining probability $1 - \mu$ is uniformly distributed over the remaining strings. Then $H(X) \leq (1-\mu)n+1 \leq n-\sqrt{n}+1$ and X is $(\mu, 2\sqrt{\mu})$-resilient, as $\Pr[X \in S] < \Pr[X = 00\ldots0] + |S|/2^n \leq \mu + \mu \leq 2\sqrt{\mu}$ for any set S of density μ. Otherwise, if $\mu < 1/\sqrt{n}$, let X be uniformly distributed over some $\sqrt{\mu}2^n$ strings. Then $H(X) = (1/2)\log \mu + n \leq n - (1/4)\log n$ and X is $(\mu, 2\sqrt{\mu})$-resilient, as $\Pr[X \in S] \leq |S|/(\sqrt{\mu}2^n) = \mu 2^n/(\sqrt{\mu}2^n) = \sqrt{\mu}$ for any set S of density μ.

[2] Indeed, assume that $\Pr[X = x_0] = \varepsilon$ and let $Y = 1$, if $X = x_0$, and $Y = 0$ otherwise. Then $H(X) = H(X,Y) = H(X|Y) + H(Y) \leq \varepsilon \cdot 0 + (1-\varepsilon) \cdot n + H(Y) \leq (1-\epsilon)n + 1$.

Reference	Number of rounds	Communication length	Resilience	Entropy	Min entropy
[15]	$O(n)$	$O(n^2)$	$(\mu, O(\sqrt{\mu}))$ for any μ	$n - O(1)$	$n/2$
[25]	$O(\log^* n)$	$n \log^{O(1)} n$	$(\mu, O(\sqrt{\mu + \delta}))$ for any μ	$\leq n(1-\varepsilon)+1$ for some $\varepsilon > 0$ that depends on δ	constant
[17] (non-explicit)	$O(\log^* n)$	$O(n)$	$(\mu, O(\sqrt{\mu}))$		
Corollary 9	$O(\log^* n)$	$O(n^2)$		$n - O(1)$	$n/2 - O(\log n)$
Lemma 8	$2i+1$ for any constant $i \geq 1$	$O(n^2)$		$n - \log^{(i)} n$	
[17] (non-explicit)	constant	$O(n)$	$(1/n^2, O(1/n))$	$n - O(\log n)$	$O(\log n)$
Corollary 6	6	$O(n^2)$		$n - O(\log n)$	$n/2 - O(\log n)$
Theorem 12 (non-explicit)	3	$O(n)$		$n - O(\log n)$	
Corollary 14 (non-explicit)	6	$O(n)$		$n - O(\log n)$	$n/2 - O(\log n)$
Theorem 17 (geometric)	3	$O(n)$		$n - O(n^{3/4})$	
Corollary 18 (geometric)	6	$O(n)$		$n - O(n^{3/4})$	$n/2 - O(1)$

Table 1: A comparison of protocols. In the second line (the protocol of [25])), $\delta > 0$ is a parameter of the protocol and can be chosen arbitrarily small (the number of rounds and the length of communication depend on δ). In the third line (the protocol of [17]), μ is a parameter of the protocol and can be chosen arbitrarily in $[0, 1]$.

struct an explicit three-round linear-communication protocol with Shannon entropy guarantee $n - O(n^{3/4})$ (Theorem 17). That protocol is related to Kakeya problem for finite fields. Besides, we construct three constant-round linear-communication candidate protocols. We conjecture that all of them guarantee Shannon entropy $n - O(\log n)$.

A comparison of the mentioned random selection protocols. For reader's convenience we have collected all the above mentioned protocols and their guarantees in Table 1.

Our techniques. For our main protocol ($O(\log^* n)$-round, $O(n^2)$-communication and $n - O(1)$-entropy) we use the recursion technique similar to that used in [25]. The existence of a non-explicit protocol (3-round $O(n)$-communication $n - O(\log n)$-entropy) is proved by standard probabilistic arguments based on Chernov bounds. However our explicit linear communication protocols use novel techniques, especially the "geometric protocol" from Theorem 17.

A comparison of this paper with its conference version [7]. The present paper is the improved and extended version of the paper [7] by the same authors. Here we summarize the main novel things in this paper (compared to [7]).

- The analysis of the Geometric protocol has been improved using Z. Dvir's technique from [12]. Previously, using G. Mockenhaupt and T. Tao bounds for Kakeya problem, we were able to prove that Geometric protocol yields Shannon entropy $3n/4 - O(1)$ (Theorem 2 from[7]). Now we can show that it yields Shannon entropy $n - O(n^{3n/4})$ (Theorem 17).
- In Lemma 7 from [7] we established some guarantee for the protocol $P(\text{Alice}, \text{Bob}, f_{\text{rot}})$ for all *prime* n, and it was mentioned in [7] that the same holds for the protocols $P(\text{Alice}, \text{Bob}, f_{\text{lin}})$ and $P(\text{Alice}, \text{Bob}, f_{\text{mul}})$. Now, using a result from [6], we can drop the assumption that n is prime and prove the same guarantee for all three protocols for *all* n (Theorem 19).
- Now we explicitly state all the results obtained by "averaging technique" (Corollaries 6, 9, 14, and 18).

The paper is organized as follows. In the next section we review the notion of entropy and of other measures of randomness, we establish some relationships among them, and define selection protocols. Section 4 contains our main protocol that achieves entropy $n - O(1)$ and constant-round protocols P_i with entropy $n - \log^{(i+1)} n$. In Section 5 we address the problem of reducing the communication complexity of our protocols.

3 Preliminaries

3.1 Random variables and Shannon entropy

Let Y be a random variable with a finite range S. *Shannon entropy of* Y is defined by:

$$H(Y) = -\sum_{s \in S} \Pr[Y = s] \cdot \log \Pr[Y = s].$$

If for some $s \in S$, $\Pr[Y = s] = 0$ then the corresponding term in the sum is considered to be zero. All logarithms are based two.

Let X, Y be (possibly dependent) jointly distributed random variables with ranges T, S, respectively. *Shannon entropy of* Y *conditional to* X is defined by:

$$H(Y|X) = \sum_{t \in T} \Pr[X = t] H(Y|X = t),$$

where $Y|X = t$ stands for the random variable whose range is S and which takes outcome $s \in S$ with probability $\Pr[Y = s|X = t]$.

The following are basic facts about Shannon entropy:

$$\begin{aligned}
H(f(Y)) &\leq H(Y) \text{ for any function } f\ , && (1)\\
H(X|f(Y)) &\geq H(X|Y) \text{ for any function } f\ , && (2)\\
H(Y) &\leq \log |S|, && (3)\\
H(Y|X) &\leq H(Y), && (4)\\
H(\langle X, Y\rangle) &= H(X) + H(Y|X), && (5)
\end{aligned}$$

Here $\langle X, Y\rangle$ stands for the random variable with range $S \times T$, which takes the outcome $\langle s, t\rangle$ with probability $\Pr[X = t, Y = s]$. We will abbreviate $H(\langle X, Y\rangle)$ as $H(X, Y)$ in the sequel.

The following corollaries of these facts are used in the sequel.

Corollary 1.

1. $H(X) \le H(\langle X, Y\rangle)$, (6)
2. $H(\langle X, Y\rangle) \le H(X) + H(Y)$ (7)

Proof. Inequality (6) follows from (5). Inequality (7) follows from (4) and (5). □

The *min entropy* of a random variable X with a finite range S is

$$H_\infty(X) = \min\{-\log \Pr[X = s] : s \in S\}.$$

It is straightforward that Shannon entropy is always greater than or equal to min entropy:

$$H(X) \ge H_\infty(X).$$

The *statistical distance* between random variables X, Y with the same finite range S is defined as the maximum

$$|\Pr[X \in A] - \Pr[Y \in A]|$$

over all subsets A of S. It is easy to see that the maximum is attained for A consisting of all s with $\Pr[X = s] > \Pr[Y = s]$ (as well as for its complement). For every integer $n \ge 1$, we denote by U_n the uniform probability distribution over $\{0, 1\}^n$.

For $\mu, \epsilon > 0$, a random variable X in $\{0, 1\}^n$ is (μ, ϵ)-*resilient* if for any set S of density at most μ (that is, $|X| \le \mu 2^n$), the probability that X is in S is at most ϵ. In order to compare our results with previous work, and to prove some of our results, we need the following

Lemma 2. *For any random variable X in $\{0, 1\}^n$ the following holds.*

1. *If X is $(2^{-j}, \varepsilon)$-resilient, then $H(X) \ge (n - j)(1 - \varepsilon)$.*
2. *If X is $(2^{-j}, \varepsilon_j)$-resilient for all $j = 0, 1, \ldots, n - 1$, then $H(X) \ge n - \sum_{j=0}^{n-1}(j + 1)\varepsilon_j$.*

This lemma is proved in Appendix. By the second item, if the series $\sum_{j=0}^{\infty}(j + 1)\varepsilon_j$ converges to c, then $H(X) \ge n - c$. In this way we will derive entropy guarantees $n - O(1)$ from resilience guarantees. For instance, if a random variable X with the range $\{0, 1\}^n$ is $(\mu, O(\sqrt{\mu}))$-resilient for all μ, then

$$H(X) \ge n - \sum_{j=0}^{\infty}(j + 1)O(\sqrt{2^{-j}}).$$

As the series $\sum_{j=0}^{\infty}(j + 1)2^{-j/2}$ converges, we have $H(X) \ge n - O(1)$. Item 1 does not suffice to make such conclusion, as $(n - j)(1 - \varepsilon)$ is not $n - O(1)$ for any constant positive ε. In particular, the entropy of the outcome of protocols of [17] can be less than $n(1 - \varepsilon)$, if one of the parties cheats.

3.2 Random selection protocols

Definition 1. A random selection protocol $\Pi = (A, B, f)$ over $\{0,1\}^n$ consists of a pair of functions A and B from $\{0,1\}^\infty \times \{0,1,\#\}^*$ to $\{0,1\}^*$ and a function $f : \{0,1,\#\}^* \to \{0,1\}^n$. It works as follows:

- Both A (Alice) and B (Bob) alternately output strings ("messages") m_i of arbitrary length that are a function of the conversation thus far and their sequences of random coin tosses r_A and r_B (from $\{0,1,\#\}^\infty$), respectively. That is, $m_1 = A(r_A, \text{empty string})$, $m_2 = B(r_B, m_1)$, $m_3 = A(r_A, m_1\#m_2)$, etc.
- The conversation between Alice and Bob is the transcript $m_1\#m_2\#\ldots\#m_r$, where r is a parameter defining the number of messages (i.e., the number of rounds) of the protocol.
- The output of the protocol is $f(m_1\#m_2\#\ldots\#m_r)$, which is a binary string of length n.

We are interested in the behavior of the protocol when one of functions A, B is replaced by an arbitrary "cheating" function A^* or B^* from $\{0,1\}^\infty \times \{0,1,\#\}^*$ to $\{0,1\}^*$. We assume that cheating functions A^* and B^* are total and always output a binary string. Thus our protocols have guaranteed output delivery, which is the same string for both parties, even if one or both parties cheat.

We say that Alice *follows the protocol* (is *honest*) if she uses the function A. Similarly for Bob.

Definition 2. We say that a protocol P for random string selection is (k,l)-*Shannon good* if the following properties hold:

- If Alice follows the protocol (and Bob possibly deviates from it), then the outcome has Shannon entropy at least k.
- If Bob follows the protocol (and Alice possibly deviates from it), then the outcome has Shannon entropy at least l.

A (k,k)-Shannon good protocol is called just k-*Shannon good*.

In a similar way we define (k,l)-*ME good protocols* and k-*ME good* protocols, using minimal entropy in place of Shannon entropy.

Throughout the paper we use the following easy observation (proven in Appendix) that holds for every protocol (A, B, f):

Lemma 3. *Assume that Alice's strategy A guarantees that Shannon entropy of the outcome is at least k for all deterministic strategies of Bob. Then the same guarantee holds for all randomized strategies of Bob as well. A similar statement is true for min entropy in place of Shannon entropy, and for resiliency.*

A string selection protocol P is called (μ,ϵ)-*resilient* if its output is (μ,ϵ)-resilient provided at least one of the parties is honest.

Averaging the asymmetry. One of the interesting features of many of our protocols is the asymmetry of cheating power of the two parties. We use this asymmetry to build the protocol with entropy $n - O(1)$. One can also use this asymmetry for "*averaging*" their cheating powers in the following simple way. Given a protocol $Q_n(\text{Alice}, \text{Bob})$ for selecting an n bit string, Alice and Bob first select the first $n/2$ bits of the string by running the protocol $Q_{n/2}(\text{Alice}, \text{Bob})$, and then they select the other half of the string by running the protocol $Q_{n/2}(\text{Bob}, \text{Alice})$ (Alice and Bob exchange their roles).

Lemma 4. *If the protocol Q_n is $(k(n), l(n))$-Shannon good then the* averaging *protocol is $(k(n/2) + l(n/2))$-Shannon good. Similarly, if the protocol Q_n is $(k(n), l(n))$-ME good then the* averaging *protocol is $(k(n/2) + l(n/2))$-ME good*

This lemma is proved in Appendix.

4 The main protocol

In this section we construct a protocol that is $(n - O(1))$-Shannon good. We start with the following protocol.

Lemma 5. *There is a $(n-1, n-\log n)$-Shannon good protocol P_0 running in 3 rounds and communicating $n^2 + n + \log n$ bits. If Bob is honest then the outcome of P_0 has min entropy at least $n - \log n$.*

Proof. The protocol $P_0(\text{Alice}, \text{Bob})$ is as follows:

1. Alice picks $x_1, x_2, \ldots, x_n \in \{0,1\}^n$ uniformly at random and sends them to Bob.
2. Bob picks $y \in \{0,1\}^n$ uniformly at random and sends it to Alice.
3. Alice picks an index $j \in \{1, \ldots, n\}$ uniformly at random and sends it to Bob.
4. The outcome R of the protocol is $x_j \oplus y$, i.e., the bit-wise xor of x_j and y.

1) Assume that Alice follows the protocol and Bob is trying to cheat. Hence, Alice picks uniformly at random $x_1, \ldots, x_n \in \{0,1\}^n$. Bob picks y. Then Alice picks a random index $j \in \{1, \ldots n\}$ and they set $R = x_j \oplus y$. Clearly, $H(x_1, \ldots, x_n) = n^2$, thus

$$\begin{aligned} n^2 &= H(x_1, \ldots, x_n) \le H(x_1, \ldots, x_n, y) \le H(x_1 \oplus y, \ldots, x_n \oplus y, y) \\ &\le H(x_1 \oplus y, \ldots, x_n \oplus y) + H(y) \le H(x_1 \oplus y, \ldots, x_n \oplus y) + n. \end{aligned}$$

Here the first inequality holds by (6), the second one by (1), the third one by (7), and the last one by (3). Therefore,

$$(n^2 - n)/n \le H(x_1 \oplus y, \ldots, x_n \oplus y)/n \le \sum_{i=1}^{n} H(x_i \oplus y)/n = H(x_j \oplus y | j) \le H(x_j \oplus y).$$

Here the second inequality holds by (7), the equality holds, as Alice chooses j uniformly, and the last inequality is true by (4).

2) Assume that Bob follows the protocol and Alice is trying to cheat. As Shannon entropy is greater than or equal to the min entropy, it suffices to prove the lower bound on the min entropy. WLOG we can assume that Alice uses a deterministic strategy. Fix

a deterministic strategy of Alice, which picks a particular sequence $x_1, \ldots, x_n$ in the first round and then sends a $i = i(y)$ in the third round. For every n bit string s the probability of event $x_{i(y)} \oplus y = s$ does not exceed the probability of event $\exists i,\ x_i \oplus y = s$, which is at most $n2^{-n}$ by union bound over i's. □

Remark 1. Note that the entropy bounds for the outcome of P_0 are tight. Indeed, a cheating Bob can set $y = x_1$ in the protocol. Then, in the third round, with probability $1/n$ Alice chooses $j = 1$ and the outcome becomes the all-zero string. Thus R takes the all-zero string with probability close to $1/n$ and the remaining probability is uniformly distributed over other strings. Hence

$$\begin{aligned} H(R) &\approx -(1/n)\log(1/n) - (1-1/n)\log((1-1/n)2^{-n}) \\ &= (\log n)/n - (1-1/n)\log(1-1/n) + (n-1) \approx n-1 \end{aligned}$$

(all approximate equalities hold with accuracy $o(1)$). Similarly, a cheating Alice can enforce the first $\lfloor \log n \rfloor$ bits of the outcome to be all zero bits. To this end she chooses $x_1, x_2, \ldots, x_n$ so that all $\lfloor \log n \rfloor$-bit strings occur among prefixes of $x_1, x_2, \ldots, x_n$. Thus in the third round she can choose j so that x_j has the same length-$\lfloor \log n \rfloor$ prefix as y. Hence $H(R) \le n - \lfloor \log n \rfloor$ in that case.

From Lemmas 5 and 4 we obtain the following

Corollary 6. *There is a $(n - O(\log n))$-Shannon good protocol running in 6 rounds[3] and communicating $O(n^2)$ bits. If either Alice or Bob is honest, then the min entropy of the protocol is at least $n/2 - O(\log n)$.*

Our protocol P_0 achieves our goal of having entropy of the outcome close to n if Alice is honest. However if she is dishonest she can fix up to $\log n$ bits of the outcome to her will. Clearly, Alice's cheating power comes from the fact that she can choose up to $\log n$ bits in the last round of the protocol. If we would reduce the number of strings x_j she can choose from in the last round, her cheating ability would decrease as well. Unfortunately, that would increase cheating ability of Bob. Hence, there is a trade-off between cheating ability of Alice and Bob. To overcome this, we will reduce the number of strings Alice can choose from, but at the same time we will also limit Bob's cheating ability by replacing his y by an outcome of yet another run of the protocol played with Alice's and Bob's roles reversed. By iterating this several times we obtain our main protocol.

Let $\log^* n$ stand for the number of times we can apply the function $\lceil \log x \rceil$ until we get 1 from n. For instance, $\log^* 17 = \log^* 2^{16} = 4$.

Theorem 7. *There is a $(n-2, n-3)$-Shannon good protocol running in $2\log^* n + 1$ rounds and communicating $n^2 + O(n \log n)$ bits. If n is even and Bob is honest or n is odd and Alice is honest, then the min entropy of the protocol is at least $n - O(\log n)$.*

Proof. Let $k = \log^* n - 1$. Define $\ell_0 = n$ and $\ell_i = \lceil \log \ell_{i-1} \rceil$, for $i = 1, \ldots, k$, so $\ell_{k-1} \in \{3, 4\}$ and $\ell_k = 2$.

The protocol of the theorem is P_k where for $i = 1, \ldots, k$ the protocol $P_i(\text{Alice}, \text{Bob})$ is defined as follows.

[3]One can wrongly think that the concatenation of 3 round protocols P(Alice,Bob) and P(Bob,Alice) has 5 (and not 6) rounds, since the 3rd and 4th messages are on the same directions. Actually, the 3rd and 4th messages are on the opposite directions because the last message in P(Alice,Bob) is send by Alice, and the first message in P(Bob,Alice) is sent by Bob, who plays Alice's part.

1. Alice picks $x_1, x_2, \ldots, x_{\ell_i} \in \{0,1\}^n$ uniformly at random and sends them to Bob.
2. Alice and Bob now run protocol $P_{i-1}(\text{Bob}, \text{Alice})$ (note that players exchange their roles) and set y to the outcome of that protocol.
3. Alice picks an index $j \in \{1, \ldots, \ell_i\}$ uniformly at random and sends it to Bob.
4. The outcome R_i of this protocol is $x_j \oplus y$.

We claim that the protocol P_i is $(n-2, n-\log 4\ell_i)$-Shannon good. This implies the theorem since $\ell_k = 2$.

Lemma 8. *For all $i = 0, 1, \ldots, k$ the following is true.*

1. *If Alice follows the protocol $P_i(\text{Alice}, \text{Bob})$ then the outcome R_i satisfies $H(R_i) \geq n-2$.*
2. *If Bob follows the protocol $P_i(\text{Alice}, \text{Bob})$ then the outcome R_i of the protocol satisfies $H(R_i) \geq n - \log 4\ell_i$.*
3. *Furthermore, if i is even and Bob is honest or i is odd and Alice is honest, then $H_\infty(R_i) \geq n - \sum_{j=0}^{i} \log \ell_j$.*

Proof. We prove all the claims simultaneously by an induction on i. For $i = 0$ the claims follow from Lemma 5. So assume that the claims are true for $i-1$ and we will prove them for i.

1) If Alice follows the protocol $P_i(\text{Alice}, \text{Bob})$ then she picks $x_1, \ldots, x_{\ell_i}$ uniformly at random. Then the protocol $P_{i-1}(\text{Bob}, \text{Alice})$ is invoked to obtain $y = R_{i-1}$. We can reason just as in the proof of Lemma 5. However this time we have a better lower bound for $H(x_1, \ldots, x_{\ell_i}, y)$. Indeed, by induction hypothesis, since Alice follows the protocol,

$$H(y|x_1, \ldots, x_{\ell_i}) \geq n - \log 4\ell_{i-1} \geq n - 2\ell_i.$$

Here the last inequality holds for all $i < k$, as $\ell_{i-1} > 4$ in this case and hence

$$\log \ell_{i-1} > 2 \quad \Rightarrow \quad 2\ell_i \geq 2\log \ell_{i-1} > \log \ell_{i-1} + 2 = \log 4\ell_{i-1}.$$

For $i = k$ we have $\ell_{i-1} \in \{3, 4\}$ and $\ell_i = 2$ and the inequality is evident.

Thus,

$$H(x_1, \ldots, x_{\ell_i}, y) = H(x_1, \ldots, x_{\ell_i}) + H(y|x_1, \ldots, x_{\ell_i}) \geq \ell_i n + (n - 2\ell_i).$$

Just as in Lemma 5, this implies

$$\begin{aligned} H(x_j \oplus y) \geq H(x_j \oplus y|j) &= \sum_{s=1}^{\ell_i} H(x_s \oplus y)/\ell_i \\ &\geq (H(x_1, \ldots, x_{\ell_i}, y) - H(y))/\ell_i \geq (\ell_i n + (n - 2\ell_i) - n)/\ell_i = n - 2. \end{aligned}$$

2) Assume that Bob follows the protocol $P_i(\text{Alice}, \text{Bob})$ but Alice deviates from it by carefully choosing $x_1, \ldots, x_{\ell_i}$ and j. Then the protocol $P_{i-1}(\text{Bob}, \text{Alice})$ is invoked to obtain

$y = R_{i-1}$. By induction hypothesis $H(y|x_1, \ldots, x_{\ell_i}) \geq n - 2$. Now Alice chooses $j \in \{1, \ldots, \ell_i\}$ and we have

$$\begin{aligned} H(x_j \oplus y) &\geq H(x_j \oplus y | \langle x_1, \ldots, x_{\ell_i} \rangle) \geq H(y|\langle x_1, \ldots, x_{\ell_i}\rangle) - H(j|\langle x_1, \ldots, x_{\ell_i}\rangle) \\ &\geq H(y|\langle x_1, \ldots, x_{\ell_i}\rangle) - H(j) \geq n - 2 - \log \ell_i. \end{aligned}$$

The claim about min entropy follows by induction. The base of induction $i = 0$ holds by Lemma 5. The induction step: assume that $i > 0$ is even and Bob is honest. Then $i - 1$ is odd and on stage 2 Bob plays Alice's part, thus we may apply the induction hypothesis about min entropy of the outcome of $P_{i-1}(\text{Bob}, \text{Alice})$. By induction hypothesis for the random string y selected on stage 2, we have $H_\infty(y) \geq n - \sum_{k=0}^{i-1} \log \ell_k$. That is, for every n bit string s,

$$\Pr[y = s] \leq 2^{-(n - \sum_{k=0}^{i-1} \log \ell_k)}.$$

By union bound for every s the probability that the outcome equals s is

$$\Pr[x_j \oplus y = s] \leq \sum_j \Pr[y = x_j \oplus s] \leq \ell_i 2^{-(n - \sum_{k=0}^{i-1} \log \ell_k)} = 2^{-(n - \sum_{k=0}^{i} \log \ell_k)}.$$

If i is odd and Alice is honest, then the arguments are even simpler:

$$\Pr[x_j \oplus y = s] = (1/\ell_j) \sum_j \Pr[y = x_j \oplus s] \leq 2^{-(n - \sum_{k=0}^{i-1} \log \ell_k)} < 2^{-(n - \sum_{k=0}^{i} \log \ell_k)}.$$

□

By the lemma, the protocol P_k is $(n-2, n-3)$ good. It runs in $2k+3 = 2(\log^* n - 1) + 3$ rounds.

The number of communicated bits is equal to

$$n^2 + n + \log n + \sum_{i=1}^{k} (n\ell_i + \log \ell_i)$$

All ℓ_i's in the sum are at most $\log n$ and decrease faster than a geometric progression. Hence the sum is at most its largest term ($n \log n$) times a constant. □

Remark 2. Our protocol $P_0(\text{Alice}, \text{Bob})$ is similar to the Random Shift Protocol of [25] (call it RS_n), and our protocol $P_k(\text{Alice}, \text{Bob})$ is obtained from $P_0(\text{Alice}, \text{Bob})$ by a recursion similar to that used to obtain the Iterated Random Shift Protocol of [25] (call it IRS_n) from the Random Shift Protocol RS_n. The difference is the following. The Random Shift Protocol RS_n is almost symmetric: both players choose a bunch of n-bit strings, $a_1, \ldots, a_n$ and $b_1, \ldots, b_n$, respectively, and the protocol outputs the list of strings $\{a_i \oplus b_j \mid i, j \leq n\}$ rather than a single string (the asymmetry is caused by the fact that they choose their strings in turn). In the Iterated Random Shift Protocol IRS_n, we first run RS_n and obtain a list of n^2 strings, then we apply $IRS_{\lfloor \log n^2 \rfloor}$ to choose a single string from the list. If n is smaller than a certain constant, then IRS_n invokes the protocol of [15].

From Theorem 7 and Lemma 4 we obtain the following

Corollary 9. *There is a $(n-5)$-Shannon good protocol running in $4\log^* n + 2$ rounds and communicating $O(n^2)$ bits. If either Alice or Bob is honest then the min entropy of the protocol is at least $n/2 - O(\log n)$.*

In [25], Sanghvi and Vadhan establish that any protocol for random selection that guarantees a constant statistical distance of the output from the uniform distribution (for some constant less than 1) requires at least $\Omega(\log^* n)$ rounds. The following lemma implies that this lower bound translates to protocols guaranteeing Shannon entropy $n - O(1)$.

Lemma 10. *For every integer c the following holds. If X is a random variable with range $\{0,1\}^n$ and $H(X) \geq n - c$, then the statistical distance of X and U_n is at most $1 - 2^{-2c-7}$.*

We prove this lemma in Appendix.

Corollary 11. *If P is a protocol that is $(n - O(1))$-Shannon good then P has at least $\Omega(\log^* n)$ rounds.*

5 Random selection protocols in constantly many rounds with linear communication length

In the previous section we have constructed protocols P_i, for $i = 0, \ldots, \log^* n - 1$, that guarantee Shannon entropy close to n and communicate $O(n^2)$ bits. In this section we will address the possibility of reducing the amount of communication in the protocols.

Let us focus on the basic protocol $P_0(\text{Alice}, \text{Bob})$, as that protocol contributes to the communication the most. The protocol can be viewed as follows.

1. Alice picks $x \in \{0,1\}^{m_A}$ uniformly at random and sends it to Bob.
2. Bob picks $y \in \{0,1\}^{m_B}$ uniformly at random and sends it to Alice.
3. Alice picks $l \in \{0,1\}^{m'_A}$ uniformly at random and sends it to Bob.
4. A fixed function $f : \{0,1\}^{m_A} \times \{0,1\}^{m_B} \times \{0,1\}^{m'_A} \to \{0,1\}^n$ is applied to x, y and l to obtain the outcome $f(x, y, l)$.

We will denote such a protocol by $P_0(\text{Alice}, \text{Bob}, f)$. In the basic protocol the parameters are: $m_A = n^2$, $m_B = n$ and $m'_A = \log n$. We would like to find another suitable function f with a smaller domain keeping the guarantee $n - o(n)$ for the entropy of the outcome.

Remark 3. Three rounds in the protocol are necessary in order to obtain the required guarantees on the output of the protocol. Indeed, by a result of [26], in any two round protocol at least one of the parties can force the output to have Shannon entropy close to $n/2$. The idea of the proof is the following. In a two round protocol, if for some x, the range of $f(x, \cdot)$ is smaller than $n2^{n/2}$, then Alice can enforce entropy $n/2 + \log n$ by picking this x. On the other hand, if the range of $f(x, \cdot)$ is larger than $n2^{n/2}$ for all x, then there is a set S of cardinality at most $2^{n/2}$ that intersects images of all functions $f(x, \cdot)$, which can be proven by a probabilistic argument. Bob can cheat by enforcing the output to lie in S.

5.1 Non-explicit protocol

The following claim indicates that finding a suitable function f should be possible.

Theorem 12. *If $f : \{0,1\}^n \times \{0,1\}^n \times \{0,1\}^{5\log n} \to \{0,1\}^n$ is taken uniformly at random among all functions, then with probability at least $1/2$, the protocol $P_0(Alice, Bob, f)$ satisfies:*

1. *If Alice follows the protocol $P_0(Alice, Bob, f)$, then the outcome R satisfies $H(R) \geq n - O(1)$.*
2. *If Bob follows the protocol $P_0(Alice, Bob, f)$, then the outcome R of the protocol satisfies $H(R) \geq H_\infty(R) \geq n - O(\log n)$.*

Proof. We will define certain properties of a function f and we will show that most functions have such properties, and that any such function satisfies the lemma. The latter will be done using the second item of Lemma 2.

The properties of a function f will ensure that the outcome of the protocol is resilient. More specifically, let $K = \{0,1\}^n$ and $L = \{0,1\}^{5\log n}$. The properties of $f : K \times K \times L \to K$ are as follows:

1. For any $S \subseteq \{0,1\}^n$, and for any function $x \mapsto y(x)$ from K to K,

$$\Pr_{x\in K, l\in L}[f(x, y(x), l) \in S] \leq |S|/2^n + 1/n^2,$$

2. For every $s \in \{0,1\}^n$ and for every $x \in K$,

$$\Pr_{y\in K}[\exists l \in L\ f(x, y, l) = s] \leq 2n^5/2^n,$$

that is,

$$|\{y \in K : \exists l \in L\ f(x, y, l) = s\}| \leq 2n^5.$$

The first condition means that the outcome of the protocol is $(\mu, \mu+1/n^2)$-resilient when Alice follows the protocol and Bob uses a deterministic strategy, specified by a function $x \mapsto y(x)$. By Lemma 2 it implies that the entropy of the outcome is $n - O(1)$ in that case. Indeed, the outcome of the protocol is $(2^{-i}, \varepsilon_i)$-resilient, where $\varepsilon_i = 2^{-i} + 1/n^2$. By Lemma 2 the entropy of the outcome is at least $n - \sum_{0\leq i<n} \varepsilon_i(i+1)$. We can split the sum $\sum_{0\leq i<n} \varepsilon_i(i+1)$ into two sums: the sum of $2^{-i}(i+1)$ and the sum of $(i+1)/n^2$. The first sum is smaller than the of infinite series $\sum_{i=0}^{\infty} 2^{-i+1}(i+1) = O(1)$. The second sum is equal to $n(n+1)/2n^2 \leq 1$.

The second property of the function f immediately implies that the min entropy (and hence entropy) of the outcome is at least $-\log(2n^5/2^n) = n - O(\log n)$.

It remains to prove that there is a function with properties 1 and 2. We show that for n large enough the probability that a random function satisfies each of the properties is at least 99/100, hence a random function satisfies both properties with probability at least 98/100. To this end we will use the Chernoff bound in the following two forms:

Lemma 13. *Assume that we are given independent random variables $Z_1, \ldots, Z_N$ with values in $\{0,1\}$. Then (a) the probability that their sum $Z = Z_1 + \cdots + Z_N$ exceeds twice the expectation $\mathrm{E}Z$ of Z is less than $e^{-\mathrm{E}Z/4}$ [2, Cor A.1.14] and (b) the probability that Z exceeds $\mathrm{E}Z + \alpha N$ is less than $e^{-2\alpha^2 N}$ [2, Thm A.1.4].*

1. Let $S \subseteq \{0,1\}^n$ and $x \mapsto y(x)$ be any mapping from K to K. The size of the set $\{(x,l) \mid f(x,y(x),l) \in S\}$ is the sum of $N = n^5 2^n$ independent random variables $Z_{x,l}$, $x \in K$, $l \in L$, where $Z_{x,l}$ is 1, if $f(x,y(x),l) \in S$, and is 0 otherwise. The expected size of this set is N times the probability $|S|/2^n$ of the event $Z_{x,l} = 1$. The first property of the function f claims that the size of this set exceeds the expected size by at most αN where $\alpha = 1/n^2$. Thus by Lemma 13(b) the property does not hold for a specific pair $S, (x \mapsto y(x))$ with probability at most $e^{-2\alpha^2 N} = e^{-2(1/n^2)^2 n^5 2^n} = e^{-2n2^n}$. By union bound over S and mappings $x \mapsto y(x)$ Property 1 does not hold with probability at most $2^{2^n} \cdot 2^{n2^n} \cdot e^{-2n2^n}$, which is negligible.

2. For the second condition, for any pair $(s,x) \in K \times K$, consider $N = |K||L| = n^5 2^n$ independent random variables $Z_{y,l}$, $y \in K$, $l \in L$, where $Z_{y,l} = 1$, if $f(x,y,l) = s$, and $Z_{y,l} = 0$ otherwise. We claim that

$$|\{y \in K : \exists l \in L \ f(x,y,l) = s\}| \leq \sum_{y \in K, l \in L} Z_{y,l}.$$

Indeed, the size of the set

$$\{y \in K : \exists l \in L \ f(x,y,l) = s\}$$

can be expressed as $\sum_{y \in K} \max\{Z_{y,l} \mid l \in L\}$. Since $\max\{Z_{y,l} \mid l \in L\} \leq \sum_{l \in L} Z_{y,l}$, the claim follows.

Hence, if for a specific pair (s,x) the inequality in the second property is false, then the sum $Z = \sum_{y \in K, l \in L} Z_{y,l}$ exceeds $2n^5$. The expectation of this sum is n^5. Thus by Lemma 13(a) we have that $\Pr_f[Z > 2n^5] \leq e^{-n^5/4}$. Hence for every pair s, x, the probability of event $|\{y \in K : \exists l \in L \ f(x,y,l) = s\}| > 2n^5$ for a random f is at most $e^{-n^5/4}$. By union bound over x and s, with a probability at least $1 - 2^{2n} \cdot e^{-n^5/4}$ the second property holds. □

From Theorem 12 and Lemma 4 we obtain the following

Corollary 14. *There is a 6-round protocol communicating $O(n)$ bits that is $(n - O(\log n))$-Shannon good and $(n/2 - O(\log n))$-ME good.*

5.2 Geometric protocol and the problem of Kakeya

The question is how to find an explicit function f with similar properties. We exhibit here an explicit function f such that the protocol $P_0(\text{Alice}, \text{Bob}, f)$ guarantees Shannon entropy at least $n - o(n)$ (if at least one party is honest).

Fix a finite field F and a natural $m \geq 2$. Let $q = |F|$. Consider the following protocol:

1. Alice picks at random a vector $d = (1, d_2, \ldots, d_m) \in F^m$ and sends it to Bob.
2. Bob picks at random $x = (x_1, \ldots, x_m) \in F^m$ and sends it to Alice.
3. Alice picks at random $t \in F$ and sends it to Bob.
4. The output of the protocol is

$$y = x + td = (x_1 + t, x_2 + td_2, \ldots, x_m + td_m).$$

The geometric meaning of the protocol is as follows. Alice picks at random a direction of an affine line in the m-dimensional space F^m over F. Bob chooses a random affine line going in that direction. Alice outputs a random point lying on the line.

Assume that Bob is honest. Then it is easy to lower bound the entropy of the output y of this protocol.

Lemma 15. *If Bob is honest, then the outcome y of the protocol satisfies*

$$H(y) \geq H_\infty(y) \geq (m-1)\log q.$$

Proof. Fix an outcome $s \in F^m$ and fix a deterministic strategy of Alice. That is, Alice chooses some d in the first round and some $t = t(x)$ in the third round. Then by union bound

$$\Pr_x[x + t(x)d = s] \leq \sum_{t \in F} \Pr_x[x + td = s] = q \cdot q^{-m}.$$

□

Note that Alice can cheat this much. For example, Alice can force $y_1 = 0$ by choosing always $t = -x_1$.

In the case when Alice is honest, we are able to prove the bound $H(y) \geq m \log q - O(m^3)$. This is related to the following problem that is known as Kakeya problem for finite fields.

Kakeya problem. *Let L be a collection of affine lines in F^m such that for each direction there is at least one line in L going in that direction. Let P_L denote the set of all points from lines from L. How small can be $|P_L|$?*

Call any set of lines L satisfying the conditions of Kakeya problem a *Kakeya family*. For every Kakeya family L consider the following deterministic Bob's strategy: choose any point from the line in L going in direction d specified by Alice. Using this strategy, Bob can force the outcome to be in P_L, and hence its entropy be at most $\log |P_L|$. Thus to prove that the entropy of the outcome is at least α (provided Alice is honest), we need the lower bound $|P_L| \geq 2^\alpha$ for Kakeya problem. Dvir [12] has shown that $|P_L| \geq \binom{q+m-1}{m}$. It turns out that the technique of [12, 13] used to show the lower bound for $|P_L|$ is suitable also to prove that the entropy of the outcome of the protocol (provided Alice is honest) is at least $m \log q - O(m^3)$.

Theorem 16. *If Alice is honest then the outcome Y of the geometric protocol is $(\mu, 2m\mu^{1/m})$-resilient (for all μ) and $H(Y) \geq m \log q - O(m^3)$.*

Proof. We start by proving that Y is $(\mu, 2m\mu^{1/m})$-resilient for all μ. By Lemma 3 we may assume that Bob uses a deterministic strategy: for every d he chooses a point $x(d) \in F^m$. We have to show that the random variable $Y = x(d) + dt$ is (μ, δ)-resilient for

$$\delta = 2m\mu^{1/m}. \tag{8}$$

For the sake of contradiction assume that there is $S \subset F^m$ with

$$|S| = \mu q^m \text{ and } \Pr[x(d) + dt \in S] > \delta. \tag{9}$$

Let us find a non-zero low-degree polynomial $P \in F[x_1, \ldots, x_m]$ that vanishes on S, that is, $P(x_1, \ldots, x_m) = 0$ for all $(x_1, \ldots, x_m) \in S$. Such polynomial can be found by solving a

system of $|S|$ linear homogeneous equations. Indeed, for every $x \in S$ the condition $P(x) = 0$ is a homogeneous linear equation in the coefficients of P. We need to choose the degree of P so that the number of coefficients of P be greater than the number of equations. Assuming that the degree of P in each variable is at most k, we are fine if $(k+1)^m > |S| = \mu q^m$. Thus we can let

$$k = \mu^{1/m} q. \tag{10}$$

Definition 3. Call a direction d *good* if $x(d) + dt \in S$ with probability more than $\delta/2$ (for a random t chosen with uniform distribution).

Claim 1. *More than $\delta/2$ fraction of directions are good.*

Proof. Recall that we assume that with probability more than δ (when a pair d, t is chosen with uniform distribution) it happens that $x(d) + dt \in S$ (Equation (9)). Hence more than $\delta - \delta/2 = \delta/2$ fraction of directions are good. □

For all good d consider the univariate polynomial $P(x(d) + dt)$ of t. We claim that this is a zero polynomial. Indeed, its degree is at most km and it vanishes in more than $q\delta/2$ points, as $P(x(d) + dt) = 0$ whenever $x(d) + dt \in S$. We have chosen k and δ so that $km = q\delta/2$ (Equations (8) and (10)). Thus $P(x(d) + dt)$ has more roots than its degree and hence is a zero polynomial.

Represent $P(x)$ as a sum of homogeneous polynomials: $P(x) = P_0 + P_1(x) + \cdots + P_l(x)$, where P_j is a homogeneous polynomial of degree j and P_l is non-zero. Since $x(d) + dt$ is a linear function of t and P_j is a homogeneous polynomial of degree j, the univariate polynomial $P_j(x(d) + dt)$ equals $P_j(d)t^j$ plus some polynomial of t of degree less than j. (Indeed, for every monomial $Q \in F[y_1, \ldots, y_m]$ and for every $a \in F^m$, the degree of the polynomial $Q(a+y) - Q(y)$ is less than that of Q and $Q(dt) = Q(d)t^i$, where i is the degree of Q.) As

$$P(x(d) + dt) = P_0 + P_1(x(d) + dt) + \cdots + P_l(x(d) + dt),$$

the polynomial $P(x(d) + dt)$ equals $P_l(d)t^l$ plus some polynomial of degree less than l.

By Definition 3, $P(x(d) + dt)$ is a zero polynomial of t for all good d. Hence $P_l(d) = 0$ for all good d. By Claim 1 there at least $\delta q^{m-1}/2$ good d's. The Schwartz-Zippel lemma states that a non-zero polynomial in $F[y_1, \ldots, y_n]$ of degree l cannot have more than $l|F|^{n-1}$ zeros. On the other hand, the non-zero degree-l polynomial $P_l(1, d_1, \ldots, d_{m-1})$ has $m-1$ variables and more than

$$\delta q^{m-1}/2 = (2m\mu^{1/m} q^{m-1})/2 = m\mu^{1/m} q^{m-1} = kmq^{m-2} \geq lq^{m-2}$$

roots (the first equality holds by (8) and the last equality holds by (10)). This contradiction shows that the outcome Y is $(\mu, 2m\mu^{1/m})$-resilient.

Let us show now that the outcome Y of the protocol has large Shannon entropy (provided Alice is honest). By Lemma 2,

$$H(Y) \geq m \log q - \sum_{i=0}^{m \log q} (i+1)\varepsilon_i,$$

where $\varepsilon_i = 2m2^{-i/m}$. Therefore

$$m \log q - H(Y) \leq \sum_{i=0}^{\infty} (i+1) 2m 2^{-i/m} = 2m \sum_{i=0}^{\infty} (i+1) 2^{-i/m}.$$

We claim that the last sum is $O(m^2)$. Indeed, we can rewrite it as

$$\sum_{i=0}^{\infty} \sum_{j=0}^{i} 2^{-i/m} = \sum_{j=0}^{\infty} \sum_{i=j}^{\infty} 2^{-i/m}.$$

The inner sum $\sum_{i=j}^{\infty} 2^{-i/m}$ is a sum of geometric series with the quotient $2^{-1/m}$ and the first term $2^{-j/m}$ and thus equals

$$\frac{2^{-j/m}}{1 - 2^{-1/m}} = \frac{2^{-j/m}}{1 - (1 - \Omega(1/m))} = O(m 2^{-j/m}).$$

Thus the outer sum is

$$\sum_{j=0}^{\infty} O(m 2^{-j/m}) = O(m) \sum_{j=0}^{\infty} 2^{-j/m} = O(m) \cdot O(m) = O(m^2).$$

Hence $H(Y) \geq m \log q - O(m^3)$. □

If we choose $m = n^{1/4}$ and $\log q = n^{3/4}$, then the lower bounds for $H(Y)$ in the cases when Alice cheats and Bob cheats coincide and are equal to $n - O(n^{3/4})$. Thus we get an explicit 3 round protocol with linear communication and entropy $n - o(n)$:

Theorem 17. *There is an explicit $(n - O(n^{3/4}))$-Shannon good 3-round protocol that communicates $2n$ bits.*

Using Lemma 4 we obtain the following corollary:

Corollary 18. *There is a 6-round explicit protocol that communicates $O(n)$ bits and that is $(n - O(n^{3/4}))$-Shannon good and $(n/2 - O(1))$-ME good.*

5.3 Explicit 3-round linear communication candidate protocols

The above results leave open the following question: Is there an explicit $(n - O(\log n))$-Shannon good protocol running in constantly many rounds with communication length $O(n)$?

We propose the following three protocols that we believe have the required properties. Consider the protocol $P_0(\text{Alice}, \text{Bob}, f)$ where f is one of the following three functions.

1. $f_{\text{rot}} : \{0,1\}^n \times \{0,1\}^n \times \{0, \ldots, n-1\} \to \{0,1\}^n$ defined by $f(x, y, j) = x^j \oplus y$, where x^j is the j-th rotation of x, $x^j = x_{j+1} \cdots x_n x_1 \cdots x_j$.

2. $f_{\text{lin}} : F^{m-1} \times F^m \times F \to F^m$, where $F = GF[2^k]$, $m = n/\log n$, $k = \log n$ and $f(x, y, j) = (1, x_1, \ldots, x_{m-1}) * j + (y_0, \ldots, y_{m-1})$. This function is similar to that used in the geometric protocol, we have just changed the values of m and q.

3. $f_{\mathrm{mul}} : F \times F \times \{0, \dots, n-1\} \to F$, where $F = GF[2^n]$, $h_0, \dots, h_{n-1}$ are some distinct elements of F, and $f(x, y, j) = x * h_j + y$ (this function depends on the choice of $h_0, \dots, h_{n-1}$).

In particular the function f_{rot} is interesting as it would allow very efficient implementation. We conjecture that for $f \in \{f_{\mathrm{rot}}, f_{\mathrm{lin}}, f_{\mathrm{mul}}\}$ the protocol $P_0(\mathrm{Alice}, \mathrm{Bob}, f)$ is $(n - O(\log n))$-Shannon good. However we are able to prove only the following

Theorem 19. *For all $f \in \{f_{\mathrm{rot}}, f_{\mathrm{lin}}, f_{\mathrm{mul}}\}$ the protocol $P_0(Alice, Bob, f)$ is $(n/2 - n^{o(1)}, n - \log n)$-Shannon good and the min entropy of the outcome is at least $n - \log n$ when Bob follows the protocol.*

Proof. 1. All the three functions have the following feature: for all fixed x, j and uniformly distributed y the random variable $f(x, y, j)$ has uniform distribution. This implies that the outcome of the protocol has min entropy at least $n - \log n$ provided Bob follows the protocol. This is proved by an analysis similar to that in the proof of Lemma 5: Fix a deterministic strategy of Alice, which picks a particular x in the first round and then sends a $j = j(y)$ in the third round. For every s the probability of event $f(x, y, j(y)) = s$ does not exceed the probability of event $\exists j,\ f(x, y, j) = s$, which is at most $n2^{-n}$ by union bound over j's.

2. Assume that Alice follows the protocol. WLOG Bob is deterministic. Let x be chosen uniformly at random and y be set depending on x. As we have seen in the proof of Lemma 5, the entropy of the outcome is at least $\sum_{j=0}^{n-1} H(f(x, y, j))/n$. Obviously, the arithmetic mean of any $n \geq 2$ numbers $a_0, \dots, a_{n-1}$ is equal to the arithmetic mean of $n(n-1)/2$ numbers $(a_k + a_l)/2$ for $k \neq l \in \{0, \dots, n-1\}$.

Fix $k \neq l \in \{0, \dots, n-1\}$. By inequalities (7) and (1), we have

$$H(f(x, y, k)) + H(f(x, y, l)) \geq H(f(x, y, k), f(x, y, l)) \geq H(f(x, y, k) - f(x, y, l)).$$

All the three functions have the following feature: $f(x, y, k) - f(x, y, l)$ does not depend on y and equals

$$\begin{cases} x^k \oplus x^l & \text{if } f = f_{\mathrm{rot}}, \\ (1, x_1, \dots, x_{k-1}) * (k - l) & \text{if } f = f_{\mathrm{lin}}, \\ x * (h_k - h_l) & \text{if } f = f_{\mathrm{mul}}. \end{cases}$$

In the case $f = f_{\mathrm{mul}}$ the difference $f(x, y, k) - f(x, y, l)$ has uniform distribution and hence $H(f(x, y, k) - f(x, y, l)) = n$, which implies that the entropy of the outcome is at least $n/2$ provided Alice is honest.

For $f = f_{\mathrm{lin}}$ the difference $f(x, y, k) - f(x, y, l)$, is uniformly distributed in the set of all k-dimensional vectors whose first coordinate is $k - l$. Hence $H(f(x, y, k) - f(x, y, l)) = (k - 1) \log n = n - \log n$. This implies that the entropy of the outcome is at least $n/2 - (\log n)/2$ provided Alice is honest.

The case $f = f_{\mathrm{rot}}$ is the hardest one. Note that in this case it is not true that $H(f(x, y, k) - f(x, y, l))$, that is, $H(x^k \oplus x^l)$ is close to n for *all* $k \neq l \in \{0, \dots, n-1\}$. For example, if n is even, $k = 0$, $l = n/2$, then ith bit and $(i + n/2)$th bit of $x^k \oplus x^l$ coincide for all $i = 0, \dots, n/2 - 1$, hence $H(x^k \oplus x^l) \leq n/2$. However, *on average* $H(x^k \oplus x^l)$ is close to n.

Lemma 20. *Assume that $n \geq 2$ and x is chosen with uniform probability distribution in $\{0, 1\}^n$. Then the average of $H(x^k \oplus x^l)$ for $k \neq l \in \{0, \dots, n-1\}$ is at least $n - n^{o(1)}$.*

Proof. Fix $k \neq l \in \{0, \ldots, n-1\}$. The mapping $x \mapsto x^k \oplus x^l$ is a homomorphism hence $x^k \oplus x^l$ has uniform distribution over its range.

The range of the mapping $x \mapsto x^k \oplus x^l$ is equal to the set of all $z \in \{0,1\}^n$ such that the system of equations

$$x_{(i+k) \bmod n} \oplus x_{(i+l) \bmod n} = z_i, \quad i = 0, 1, \ldots, n-1 \tag{11}$$

is consistent. In this system every equation has two variables. Thus it is convenient to represent this system by a directed graph, call it $G_{n,l,k}$. The nodes of the graph are variables $x_0, x_1, \ldots, x_{n-1}$ and arcs of the graph correspond to equations: ith arc starts in the node $x_{(i+k) \bmod n}$ and ends in $x_{(i+l) \bmod n}$. It is easy to see that each vertex of $G_{n,l,k}$ has one incoming edge and one outgoing edge. Hence $G_{n,l,k}$ is a union of vertex-disjoint oriented cycles $C_1, \ldots, C_m$.

We claim that the dimension of the image of the mapping $x \mapsto x^k \oplus x^l$ is equal to $n - m$. Indeed, the system (11) can be split in m sub-systems, associated with the cycles $C_1, \ldots, C_m$. Those sub-systems do not have common variables. Let C_t be one of the cycles, and let $z_0, \ldots, z_{n-1}$ be obtained from some $x \in \{0,1\}^n$ by equations (11). Then

$$\sum_{z_i \in C_t} z_i = \sum_{z_i \in C_t} (x_{(i+k) \bmod n} \oplus x_{(i+l) \bmod n})$$

is even, since every variable x_j has either 0 or 2 occurrences in this sum. Thus for each cycle C_t we obtain a constraint $\sum_{z_i \in C_t} z_i \equiv 0 \pmod 2$. On the other hand, assume that $z_0, \ldots, z_{n-1}$ satisfy all such constraints. Then for each equation the value of any variable from that equation determines uniquely the value of the other variable from that equation. Thus for each cycle C_t, we can pick one variable x_i from that cycle and set $x_i = 0$, say. The values of the remaining variables from the cycle can be determined uniquely (the last equation is fulfilled because of the constraint associated to the cycle).

Thus we need to count the number of cycles in $G_{n,l,k}$. A variable x_i is connected in $G_{n,l,k}$ to all variables of the form $x_{(i+(l-k)j) \bmod n}$, $j = 0, \ldots, n-1$. If n is prime, then the range of the mapping $j \mapsto (i + (l-k)j) \bmod n$ consists of all residues modulo n. In this case $G_{n,l,k}$ consists of one cycle and hence $H(x^k \oplus x^l) = n - 1$.

In general case, the cardinality of the range of the mapping $j \mapsto (i + (l-k)j) \bmod n$ is equal to $n/\gcd(l-k, n)$. Hence each cycle has $n/\gcd(l-k, n)$ nodes and the number of cycles is $\gcd(l-k, n)$. That is, $H(x^k \oplus x^l) = n - \gcd(l-k, n)$.

Thus we need to estimate the average of $\gcd(l-k, n)$ where $k \neq l \in \{0, \ldots, n-1\}$ are chosen with uniform distribution. Obviously it is equal to

$$\frac{\sum_{i=1}^{n-1} \gcd(i, n)}{n-1}.$$

The sum $\sum_{i=1}^{n-1} \gcd(i, n)$ is a well studied subject in the Number Theory, and it is called the *gcd-sum*. By a result of Broughan [6] we have $\sum_{i=1}^{n-1} \gcd(i, n) \leq n^{1+o(1)}$. Hence the average of $H(x^k \oplus x^l)$ is at least $n - n^{o(1)}$. □

The theorem is proved. □

6 Open questions

1. Is there an explicit $(n - O(\log n))$-Shannon good protocol running in constantly many rounds with communication length $O(n)$?

2. Is there a $(n - O(1))$-Shannon good protocol with communication length $o(n^2)$?

3. Given r, n, what is maximal $h = h_r(n)$ such that there is a h-Shannon good protocol running in r rounds? Lemma 8 provides a lower bound $h_r(n) \geq n - \log^{(\lfloor (r-1)/2 \rfloor)} n$. Building on results of [25], Stepanov [26] showed that $h_r(n) \leq n - \frac{1}{8} \log^{(\log^* \log^* n + r)} n$.

4. Is it true that that for all $f \in \{f_{\mathrm{rot}}, f_{\mathrm{lin}}, f_{\mathrm{mul}}\}$ the protocol $P_0(\mathrm{Alice}, \mathrm{Bob}, f)$ is $(n - O(\log n))$-Shannon good?

Acknowledgments

We would like to thank to Troy Lee and John Tromp for useful discussions and Navin Goyal for pointing us to the problem of Kakeya. We also thank anonymous referees for valuable comments on the paper. Part of the work was done while the second, third, fourth, and sixth author were visiting CWI, Amsterdam. H. Buhrman was supported by EU project QAP and BRICKS project AFM1. H. Buhrman and M. Koucký were supported in part by an NWO VICI grant (639.023.302). M. Koucký was supported in part by grant GA ČR 201/07/P276, 201/05/0124, project No. 1M0021620808 of MŠMT ČR and Institutional Research Plan No. AV0Z10190503. The work of N. Vereshchagin was partially supported by the Russian Academic Excellence Project '5-100' and by the RFBR grant 19-01-00563.

References

[1] N. Alon and M. Naor. Coin-flipping games immune against linear-sized coalitions. In Proc. 31st FOCS, 1990.

[2] Noga Alon, Joel Spencer, *The probabilistic method*. John Wiley & sons, 2nd edition, 2000.

[3] Andris Ambainis, Harry Buhrman, Yevgeniy Dodis, and Hein Röhrig, Multiparty Quantum Coin Flipping. IEEE Conference on Computational Complexity 2004, pages 250–259, 2004.

[4] M. Blum. Coin flipping by telephone. In IEEE Spring COMPCOM, 1982.

[5] M. Ben-Or and N. Linial. Collective coin-flipping.Randomness and Computation, S. Micali ed., Academic Press, New York, 1989.

[6] Broughan, K.A. The gcd-sum function. Journal of Integer Sequences, 4, Article 01.2.2 (2001).

[7] Harry Buhrman, Matthias Christandl, Michal Koucký, Zvi Lotker, Boaz Patt-Shamir, Nikolai K. Vereshchagin. High Entropy Random Selection Protocols. Proceedings of 10th International Workshop, APPROX 2007, and 11th International Workshop, RANDOM 2007, Princeton, NJ, USA, August 20-22, 2007. Proceedings. Lecture Notes in Computer Science, volume 4627/2007 pages 366-379.

[8] C. Cachin, C. Crepeau, and J. Marcil. Oblivious transfer with a memory-bounded receiver. In Proc. 39th FOCS, 1998.

[9] I. Damgard. Interactive hashing can simplify zero-knowledge protocol design. In Proc. CRYPTO '95, Springer LNCS 403, 1994.

[10] I. Damgard, O. Goldreich, and A. Wigderson. Hashing functions can simplify zero-knowledge protocol design (too). TR RS-94-39. BRICS, 1994.

[11] Y. Ding, D. Harnik, A. Rosen, and R. Shaltiel. Constant-round oblivious transfer in the bounded storage model. In Proc. 1st TCC, Springer LNCS 2951, 2004.

[12] Z. Dvir. On the size of Kakeya sets in finite fields. J. Amer. Math. Soc., 22:1093-1097, 2009.

[13] Z. Dvir and A. Wigderson. Kakeya sets, new mergers and old extractors. In FOCS '08: Proceedings of the 49th Annual IEEE Symposium on Foundations of Computer Science, pages 625-633. IEEE Computer Society, 2008.

[14] U. Feige. Noncryptographic selection protocols. In Proc. 40th FOCS, 1999.

[15] O. Goldreich, S. Goldwasser, and N. Linial, Fault-tolerant computation in the full information model. SIAM Journ. on Computing 27 (2), pages 506–544, 1998.

[16] O. Goldreich, A. Sahai, and S. Vadhan. Honest-verifier statistical zero-knowledge equals general statistical zero-knowledge. In Proc. 30th STOC, 1998.

[17] R. Gradwohl, S. Vadhan, D. Zuckerman, Random selection with an Adversarial Majority In C. Dwork, editor, Advances in Cryptology—CRYPTO '06, number 4117 in Lecture Notes in Computer Science, pages 409–426, 2006. Electronic Colloquium on Computational Complexity, Technical Report TR06-026, 2006.

[18] Gerd Mockenhaupt, Terence Tao, Restriction and Kakeya phenomena for finite fields. Duke Math. J. 121 (2004), 35–74.

[19] M. Naor, R. Ostrovsky, R. Venkatesan, and M. Yung. Perfect zero-knowledge arguments for NP can be based on general complexity assumptions. J. Cryptology 11, 1998.

[20] A. Nguyen, J. Frison, K. P. Huy, S. Massar. New Journal of Physics

[21] An. Muchnik and N. Vereshchagin. "Shannon Entropy vs. Kolmogorov Complexity". Computer Science — Theory and Applications: First International Computer Science Symposium in Russia, CSR 2006. Proceedings. Editors: Dima Grigoriev, John Harrison, Edward A. Hirsch Lecture Notes in Computer Science, vol. 3967, 2006, pp. 281–291.

[22] R. Ostrovsky, S. Rajagopalan, U. Vazirani. Simple and efficient leader election in the full information model. In Proc. 26th STOC, 1994.

[23] A. Russell and D. Zuckerman. Perfect information leader election in $\log^* n + O(1)$ rounds. In Proc. 39th FOCS, 1998.

[24] M. Saks. A robust noncryptographic protocol for collective coin-flipping. SIAM J. Discrete Math, 1989.

[25] S. Sanghvi, S. Vadhan. The Round Complexity of Two-Party Random Selection. Thirty-seventh Annual ACM Symposium on Theory of Computing. Baltimore, MD, USA. Proceedings, Pages: 338–347.

[26] T. Stepanov. Random Selection in Few Rounds. Proceedings of 8th International Computer Science Symposium in Russia, CSR 2013. Lecture Notes in Computer Science v. 7913, pages 354–365.

[27] Thomas Wolff, Recent work connected with the Kakeya problem, in Prospects In Mathematics, H. Rossi, ed., AMS 1999.

[28] Ernst Zermelo, Über eine Anwendung der Mengenlehre auf die Theorie des Schachspiels. In Proceedings of the Fifth International Congress Mathematics (1913), pp. 501-504.

A Appendix: deferred proofs

The proof of Lemma 2. For $x \in \{0,1\}^n$, let $p_x = \Pr[X = x]$. For any non-negative integer i let

$$S_i = \{x \in \{0,1\}^n \mid p_x \leq 2^{-i}\}.$$

Since the total probability sums to one, we have $|\{0,1\}^n \setminus S_i| < 2^i$.

1. In order to prove the first claim note that

$$H(X) = \sum_x p_x(-\log p_x) \geq \sum_{x \in S_{n-j}} p_x(-\log p_x) \geq (n-j) \sum_{x \in S_{n-j}} p_x.$$

Since $|\{0,1\}^n \setminus S_{n-j}| < 2^{n-j}$ and X is $(2^{-j}, \varepsilon)$-resilient, it follows that $\Pr[X \notin S_{n-j}] \leq \varepsilon$. Hence $\sum_{x \in S_{n-j}} p_x \geq 1 - \varepsilon$ and

$$H(X) \geq (n-j)(1-\varepsilon).$$

2. To prove the second claim, we partition $\{0,1\}^n$ into slices $S_i \setminus S_{i+1}$:

$$H(X) = \sum_x p_x(-\log p_x) = \sum_{i=0}^{\infty} \sum_{x \in S_i \setminus S_{i+1}} p_x(-\log p_x) \geq \sum_{i=0}^{\infty} \sum_{x \in S_i \setminus S_{i+1}} p_x i.$$

Hence

$$n - H(X) \leq \sum_{i=0}^{\infty} \sum_{x \in S_i \setminus S_{i+1}} (n-i)p_x \leq \sum_{i=0}^{n-1} \sum_{x \in S_i \setminus S_{i+1}} (n-i)p_x \leq \sum_{i=0}^{n-1} \sum_{x \notin S_{i+1}} (n-i)p_x$$

Since X is $(2^{-j}, \varepsilon_j)$-resilient for all $j = 0, 1, \ldots, n-1$ and $|\{0,1\}^n \setminus S_{i+1}| < 2^{i+1}$, we conclude that

$$\sum_{x \notin S_{i+1}} p_x \leq \varepsilon_{n-i-1},$$

hence

$$n - H(X) \le \sum_{i=0}^{n-1} (n-i)\varepsilon_{n-i-1} = \sum_{j=0}^{n-1} (j+1)\varepsilon_j$$

□

The proof of Lemma 3. We first prove the min entropy part. Assume that Alice's strategy A guarantees that for all deterministic strategies B of Bob, the min entropy of the outcome is at least k. Let X_B denote the outcome random variable provided Bob uses a deterministic strategy B. Then for every x the probability $\Pr[X_B = x]$ is at most 2^{-k}.

Assume that Bob uses a randomized strategy $\mathbf{B}$. This strategy can be viewed as a probability distribution over his deterministic strategies. Let X denote the output random variable. Then $\Pr[X = x]$ is equal to the average value of $\Pr[X_B = x]$ with respect to that distribution. Hence the min entropy part follows from the fact that the average value of any random variable cannot exceed its maximal value, which is at most 2^{-k} in our case.

Similar arguments prove the resilience part.

The Shannon entropy part follows from the inequality $H(X) \ge H(X|\mathbf{B})$. Indeed, $H(X|\mathbf{B})$ is the average value of $H(X_B)$ over a randomly chosen B. □

Proof of Lemma 4. Assume that Alice is honest and hence follows the strategy A prescribed by the protocol $Q_{n/2}(\text{Alice}, \text{Bob})$ to select the first half of the output and the strategy B prescribed by the protocol $Q_{n/2}(\text{Bob}, \text{Alice})$ to select the second half of the output. To prove the first statement, we have to show that whatever strategy S follows Bob, Shannon entropy of the outcome X is at least $k(n/2) + l(n/2)$. By Lemma 3 we may assume that S is deterministic.

Let X_1, X_2 denote the first and the second part of the output, respectively. Then

$$H(X) = H(X_1) + H(X_2|X_1).$$

As the protocol $Q_{n/2}(\text{Alice}, \text{Bob})$ is $(k(n/2), l(n/2))$-Shannon good we have $H(X_1) \ge k(n/2)$ and it remains to show that $H(X_2|X_1) \ge l(n/2)$. As X_1 is a function of messages M_1 sent while selecting X_1, by inequality (2) the conditional entropy $H(X_2|X_1)$ is at least $H(X_2|M_1)$. As the protocol $Q_{n/2}(\text{Bob}, \text{Alice})$ is $(l(n/2), k(n/2))$-Shannon good, for every m_1 we have $H(X_2|M_1 = m_1) \ge l(n/2)$. Indeed, once we fix m_1, the action of Bob's strategy S while selecting the second half of the output becomes deterministic.

The bound on min entropy is proven in a similar way: for all x_1, x_2 we have

$$\Pr[X = (x_1, x_2)] = \Pr[X_1 = x_1] \cdot \Pr[X_2 = x_2 | X_1 = x_1].$$

The first factor here is at most $2^{-k(n/2)}$, as $Q_{n/2}(\text{Alice}, \text{Bob})$ guarantees min entropy at least $k(n/2)$ provided Alice is honest. The second factor is at most $2^{-l(n/2)}$, as for all messages m_1 we have $\Pr[X_2 = x_2 | M_1 = m_1] \le 2^{-l(n/2)}$. Since X_1 is a function of M_1, this implies that $\Pr[X_2 = x_2 | X_1 = x_1] \le 2^{-l(n/2)}$ as well. □

The proof of Lemma 10. Fix an integer c. For $x \in \{0,1\}^*$ let $p_x = \Pr[X = x]$. The statistical distance between U_n and X is equal to $\sum_{x: p_x > 2^{-n}} (p_x - 2^{-n})$. For all integer $i \le n$ let N_i stand for the cardinality of the set

$$T_i = \{x \mid 2^{-n+i-1} < p_x \le 2^{-n+i}\}. \tag{12}$$

And let w_i denote the cumulative probability of T_i. In terms of w_i, N_i the statistical distance between U_n and X can be rewritten as

$$\sum_{i=1}^{n} w_i - \sum_{i=1}^{n} N_i 2^{-n} \le \sum_{i=1}^{n} w_i - \sum_{i=1}^{n} 2^{-i} w_i.$$

Here the last inequality holds, as $w_i \le N_i 2^{-n+i}$ by (12).

Thus it suffices to prove that

$$\sum_{i=1}^{n} (1 - 2^{-i}) w_i \le 1 - 2^{-2c-7}$$

provided $H(X) \ge n - c$. This can be done similar to the proof of Lemma 2. Indeed,

$$H(X) \le \sum_{i \le n} \sum_{x \in T_i} (-p_x \log p_x) < \sum_{i \le n} \sum_{x \in T_i} p_x (n - i + 1) = \sum_{i \le n} w_i (n + 1 - i) = n + 1 - \sum_{i \le n} i w_i$$

hence

$$\sum_{i \le n} i w_i \le c + 1 \tag{13}$$

Here i ranges over all integers $i \le n$, including negative ones. However, the contribution of negative i's is bounded by a constant. Indeed, as $2^{n-i} w_i \le N_i \le 2^n$ we can conclude that $w_i \le 2^i$ hence

$$0 \ge \sum_{i<0} i w_i \ge \sum_{i<0} i 2^i = -2.$$

Thus, inequality (13) implies that the sum of $i w_i$ over positive i's is bounded by a constant:

$$\sum_{i=1}^{n} i w_i \le c + 3. \tag{14}$$

Split the sum $\sum_{i=1}^{n} (1 - 2^{-i}) w_i$ into two sums: the sum over all $i \ge 2(c+3)$ and the rest. Let $p = \sum_{1 \le i < 2(c+3)} w_i$ and $q = \sum_{n \ge i \ge 2(c+3)} w_i$. Then

$$\sum_{i=1}^{n} (1 - 2^{-i}) w_i \le p + q(1 - 2^{-2(c+3)}) \le p + (1-p)(1 - 2^{-2(c+3)}) = 1 - (1-p) 2^{-2(c+3)}.$$

It remains to show that $p \le 1/2$. This follows from (14). Indeed,

$$\sum_{i=1}^{n} i w_i \ge \sum_{i=2(c+3)}^{n} i w_i \ge 2(c+3) p.$$

Thus (14) implies that $2(c+3)p \le c + 3 \Rightarrow p \le 1/2$. Lemma 10 is proved. □